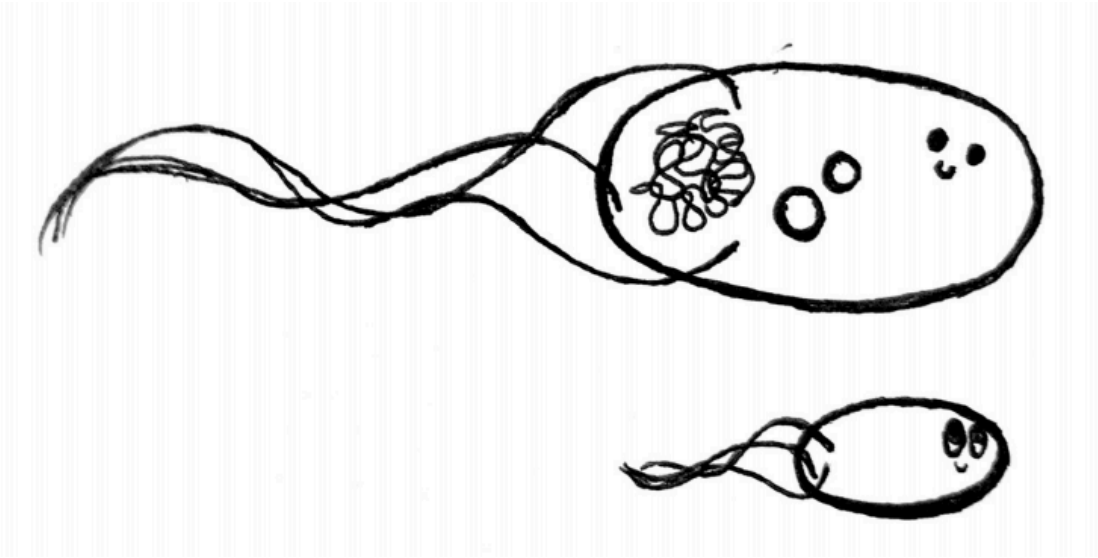
MICROBIAL GENETICS

Bi 116

Professors Dianne K. Newman & Sarkis K. Mazmanian
TAs Ana Moiseyenko & James Ousey

Winter Term 2019/20

Tu/Th, 1PM-2:30PM, 101 Kerckhoff



COURSE READER

Syllabus for BI 116 – MICROBIAL GENETICS

Winter Term 2019/2020; Tu,Th 1pm-2:30pm, 101 Kerckhoff

		OFFICE HOURS
Professors:	Sarkis K. Mazmanian; sarkis@caltech.edu	by appointment in 275 Church
	Dianne K. Newman; dkn@caltech.edu	by appointment in 301a Braun
TAs:	Ana Moiseyenko; anamois@caltech.edu	11am-12pm, Wednesdays, 271 Church
	James Ousey; jousey@caltech.edu	10am-11am, Mondays, 271 Church

Week	Tuesday	Thursday
1	Jan. 7 (DKN) *[1-1A]* *Chromosomal organization*	Jan. 9 (SKM) *[1-2A,B]* *Nature of the Gene*
2	Jan. 14 (DKN) *[2-1A,B,C]* *Mutations I*	Jan. 16 (EKP/LAM) *[2-2A,B]* *Mutations II*
3	Jan. 21 (DKN) *[3-1A,B]* *Suppressors*	Jan. 23 (DKN) *[3-2A,B]* *Gene Expression*
4	Jan. 28 (SKM) *[4-1A,B]* *Site-Specific Recombin., Transposition*	Jan. 30 (SKM) *[4-2A,B]* *Horizontal Gene Transfer*
5	Feb. 4 (DKN) *[5-1A,B]* *Stress Response*	Feb. 6 (DKN) *[5-2A,B]* *Bacterial Growth* **MIDTERM OUT**
6	Feb. 11 (SKM) *[6-1A,B]* *Sporulation* **MIDTERM DUE**	Feb. 13 (SKM) *[6-2A,B]* *Secretion*
7	Feb. 18 (DKN) *[7-1A,B]* *Motility, Chemotaxis*	Feb. 20 (DKN) *[7-2A,B]* *Biofilms*
8	Feb. 25 (SKM) *[8-1A,B]* *Yeast Genetics*	Feb. 27 (DKN) *[8-2A,B]* *Archaeal Genetics*
9	Mar. 3 (DKN) *[9-1A,B]* *Genetics and Geobiology*	Mar. 5 (SKM) *[9-2A,B,C]* *Metagenomics*
10	Mar. 10 (SKM) *[10-1A,B]* *Host-Microbe Symbiosis*	Mar. 12 (SKM) *[10-2A,B]* *Biotechnology*
11	**FINAL RESEARCH PROPOSAL DUE Feb. 17 at 5pm**	

Format. For each class, 1-2 original research papers will be assigned as required reading. During class, these papers will be discussed critically and in detail. The emphasis here is on active learning through discussion—this is not a lecture course. We expect each student to participate actively and often. To ensure this participation, each student will be responsible for leading the discussion for one or two topics (starting in Week 3). All students will need to come prepared to summarize the major points in a paper, and to comment specifically on a particular experiment or conclusion. Our goal is to ensure that everyone participates equally and that students become comfortable asking questions and engaging in constructive discussions. It thus is essential that you read and think about the assigned papers. If you have not done so, then don't come to class.

Level of Work

This is not an introductory course. It presupposes some familiarity with the techniques and

intellectual vantage points of molecular biology, as well as a comfortable acquaintance with general aspects of microbiology and cell biology (bacterial growth dynamics, composition, the cell cycle, transcription, protein synthesis, protein structure & function). You will probably find the workload overwhelming if you are trying to learn these things as you go along the term.

Grades. One third of your grade will be based on the quality of your classroom participation. Another third will come from an open-materials but time-limited midterm examination. The final third will come from a hypothesis-oriented research proposal. Besides a documented medical excuse, no other reason will be considered for submission of late exams. Any copying from published materials, the Internet, from another member of the class or any other source is plagiarism and a violation of the Honor Code.

Midterm Exam: Handed out February 6, 2020 in class; Due February 11, 2020 in class.
Final Proposal: See separate description with due dates.

We grade this class on a curve. Although we can't anticipate what percentage range will be assigned to each grade, it is generally the case that you will need to have a 60% or higher in order to pass the class.

Readings

The course reader contains the papers we will discuss in class. In addition, we suggest review articles as background reading to complement the week's theme—these suggestions can be found on the course website.

Week 1 [Introduction, Chromosome, Nature of the Gene]

Tuesday 1-1A: F. Jacob and E.L. Wollman, 1958, Genetic and physical determinations of chromosomal segments in *Escherichia coli*.

Thursday 1-2A: F.H.C. Crick, L. Barnett, S. Brenner, and R.J. Watts-Tobin, 1961, General nature of the genetic code for proteins.

Thursday 1-2B: S. Benzer, 1961, On the topography of the genetic fine structure.

Week 2 [Mutations]

Tuesday 2-1A: S. Brenner, L. Barnett, F.H.C. Crick, and A. Orgel, 1961, The theory of mutagenesis.

Tuesday 2-1B: J.A. Shapiro, 1969, Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*.

Thursday 2-2A: S.E. Luria and M. Delbrück, 1943, Mutations of bacteria from virus sensitivity to virus resistant.

Thursday 2-2B: J. Cairns, J. Overbaugh and S. Miller, 1988, The origin of mutants.

Week 3 [Suppressors, Gene Expression]

Tuesday 3-1A: J. Jarvik and D. Botstein, 1975, Conditional-lethal mutations that suppress genetic defects in morphogenesis by altering structural proteins.

Tuesday 3-1B: K.L. Bieker and T.J. Silhavy, 1989, PrlA is important for the translocation of exported proteins across the cytoplasmic membrane of *Escherichia coli.*

Thursday 3-2A: A.B. Pardee, F. Jacob and J. Monod (PaJaMo paper), 1959, The genetic control and cytoplasmic expression of "inducibility" in the synthesis of β-galactosidase by *E. coli.*

Thursday 3-2B: E. Englesberg et al., 1969, An analysis of "revertants" of a deletion mutant in the *C* gene of the L-arabinose gene complex in *Escherichia coli* B/r: isolation of initiator constitutive mutants (I^c)

Week 4 [Recombination, Transposition, HGT]

Tuesday 4-1A: M. Gottesman and M. Yarmolinksy, 1968, Integration-negative mutants of bacteriophage lambda

Tuesday 4-1B: M. Silverman and M. Simon, 1980, Phase variation: genetic analysis of switching mutants

Thursday 4-2A: N. Kleckner, R. Chan, BK Tye, and D. Botstein, 1975, Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition

Thursday 4-2B: V. Freeman, 1951, Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*

Week 5 [Stress Response, Growth in Stationary Phase]

Tuesday 5-1A: R. Lange and R. Hengge-Aronis, 1991, Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*.

Tuesday 5-1B: B.M. TJ. Mescas et al., 1993, The activity of an *Escherichia coli* heat-inducible σ-factor, is modulated by expression of outer membrane proteins.

Thursday 5-2A: M.M. Zambrano et al., 1993, Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures.

Thursday 5-2B: S. Finkel and R. Kolter, 1999, Evolution of microbial diversity during prolonged starvation

Thursday 5-2C: E.R. Zinser and R. Kolter, 1999, Mutations enhancing amino acid catabolism confer a growth advantage in stationary phase

Week 6 [Sporulation, Secretion]

Tuesday 6-1A: J. Dworkin and R. Losick, 2005, Developmental commitment in a bacterium.

Tuesday 6-1B: C.D. Webb et al., 1997, Bipolar Localization of the Replication Origin Regions of Chromosomes in Vegetative and Sporulating Cells of *B. subtilis*.

Thursday 6-2A: D. Oliver and J. Beckwith, 1982, Identification of a New Gene (*secA*) and Gene Product Involved in the Secretion of Envelope Proteins in *Escherichia coli.*

Thursday 6-2B: S.K. Mazmanian et al., 1999, *Staphylococcus aureus* Sortase, an Enzyme that Anchors Surface Proteins to the Cell Wall.

Week 7 [Motility, Chemotaxis, Biofilms]

Tuesday 7-1A: M. Silverman and M. Simon, 1974, Flagellar rotation and the mechanism of bacterial motility.

Tuesday 7-1B: J.S. Parkinson and S.R. Parker, 1979, Interaction of the *cheC* and *cheZ* gene products is required for chemotactic behavior in *Escherichia coli.*

Thursday 7-2A: G.A. O'Toole and R. Kolter, 1998, Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development.

Thursday 7-2B: R.N. Esquivel and M. Pohlschroder, 2014, A conserved type IV pilin signal peptide H-domain is critical for the post-translational regulation of flagella-dependent motility

Week 8 [Yeast, Archaeal Genetics]

Tuesday 8-1A: P. Novick, C. Field, R. Schekman, 1980, Identification of 23 Complementation Groups Required for Post-translational Events in the Yeast Secretory Pathway.

Tuesday 8-1B: R.J. Deshaies and R. Schekman, 1987, A Yeast Mutant Defective at an Early Stage in Import of Secretory Protein Precursors into the Endoplasmic Reticulum.

Thursday 8-2A: Abdul-Halim et al., 2019. Preprint. Lipid Anchoring of Archaeosort Substrates and Mid-Cell Growth in Haloarchaea.

Thursday 8-2B: W.W. Metcalf et al., 1997, A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors.

Week 9 [Genetics and Geobiology, Metagenomes]

Tuesday 9-1A: C.W. Saltikov *et al.*, 2003, The *ars* detoxification system is advantageous but not required for As(V) respiration by the genetically tractable *Shewanella* species strain ANA-3.

Tuesday 9-1B: C.W. Saltikov and D.K. Newman, 2003, Genetic identification of a respiratory arsenate reductase

Tuesday 9-1C: D. Malasarn et al., 2004, *arrA* is a reliable marker for As(V) respiration.

Thursday 9-2A: C. Lozupone and R. Knight, 2005, UniFrac: A New Phylogenetic Method for Comparing Microbial Communities.

Thursday 9-2B: Caporaso et al., 2010. QIIME allows analysis of high-throughput community sequencing data.

Thursday 9-2C: J. Qin et al., 2010. A Human Gut Microbial Gene Catalog Established by Metagenomic Sequencing.

Week 10 [Host-Microbe Symbiosis, Biotechnology]

Tuesday 10-1A: S.M. Lee et al., 2013, Bacterial Colonization Factors Control Specificity and Stability of the Gut Microbiota.

Tuesday 10-1B: A.L. Goodman, 2009. Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat.

Thursday 10-2A: S. Moriano-Gutierrez et al., 2019, Critical Symbiont Signals Drive Both Local and Systemic Changes in Diel and Developmental Host Gene Expression.

Thursday 10-2B: L.A. Marraffini and E.J. Sontheimer, 2010, Self Versus Non-self Discrimination During CRISPR RNA-directed Immunity.

Week 1

Introduction

Chromosomal Organization

Nature of the Gene

GENETIC AND PHYSICAL DETERMINATIONS OF CHROMOSOMAL SEGMENTS IN *ESCHERICHIA COLI*

By FRANÇOIS JACOB and ELIE L. WOLLMAN
Service de Physiologie microbienne, Institut Pasteur, Paris

In *Escherichia coli* K-12, conjugation involves the *oriented transfer* of a chromosomal segment from a donor to a recipient bacterium. This process is oriented in the sense that the different genetic determinants located on the chromosomal segment transferred from the donor penetrate into the recipient in a predetermined order, and according to a rather precise time schedule (Wollman & Jacob, 1955). This characteristic of the mating process makes conjugation in bacteria a suitable material for relating genetic analysis to genetic structures, and for comparing genetic evaluation of these structures with physical measurements.

After summarizing our present knowledge of the process of conjugation in *E. coli* K-12, it is intended, in the first part of this paper, to report the available information on the organization of the genetic material in this organism. The second part will be mainly concerned with the effects of ultra-violet light and of disintegration of radioactive phosphorus on the processes of conjugation and recombination. These last experiments allow, as will be discussed, a comparison between genetic and physical determinations of chromosomal segments in bacterial crosses.

I. THE GENETIC SYSTEM OF *E. COLI* K-12

The process of conjugation

When two strains of *E. coli* K-12, which differ in such properties as synthesis of essential metabolites, fermentation of sugars, resistance or sensitivity to bacteriophages or to drugs, are mixed, genetic recombination may be demonstrated between characters of the parental types. In the recombinants thus formed the characters of the parents are reassorted and certain types of recombinants may be easily scored by plating on suitable selective media (Tatum & Lederberg, 1947; Lederberg, 1947). Sexual differentiation in *E. coli* K-12 has been demonstrated, and genetic recombination involves the transfer of genetic material from *donor* to *recipient* bacteria (Hayes, 1953*a*; Cavalli, Lederberg & Lederberg, 1953). Whereas no essential difference has been recognized between different

strains of recipient (or, F^-) bacteria, two main types of donors may be distinguished. Most of the K-12 strains, including the wild type, are F^+: upon mixing with F^- cells they exhibit a low frequency of recombination (10^{-5} or less of any type of recombinant). Some strains, however, exhibit a high frequency of recombination (from 10^{-1} to 10^{-3}) and are thus called *Hfr* (Cavalli, 1950; Hayes, 1953*b*). Closer analysis shows that only certain characters of an *Hfr* strain are transmitted at high frequency to recombinants, whereas others are transmitted at the same low frequency as is observed in $F^+ \times F^-$ crosses (Hayes, 1953*b*).

The high frequency of recombination that may be observed in $Hfr \times F^-$ crosses allows an analysis of the process of conjugation at the cell level (Wollman, Jacob & Hayes, 1956). One may schematically distinguish several successive steps in this process (Jacob & Wollman, 1955).

(i) The first step consists in the establishment of an *effective contact* between cells of opposite mating types. Under suitable conditions of environment and cell density this step is rapidly completed and practically all possible matings occur within 30 min. after mixing of bacteria of opposite mating types. Electron micrographs demonstrate the existence of a bridge which unites conjugating bacteria (Anderson, Wollman & Jacob, 1957).

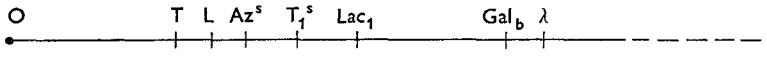
(ii) The second step consists in the oriented *transfer* of a chromosomal segment of the *Hfr* donor into the F^- recipient. The mechanism of transfer has been analysed by interrupting the process of conjugation at different times after its onset, by means of a Waring blendor (Wollman & Jacob, 1955). It was thus found that the genetic characters linearly arranged on the *Hfr* chromosome segment penetrate into the recipient in a predetermined order and always the same extremity, O (for origin), first. This process is slow enough to be interrupted at various times by mechanical treatment. Interruption of the mating process does not prevent any genetic character which has already penetrated the recipient from being later integrated into a recombinant. Even when the process is not artificially interrupted, spontaneous breaks do occur during transfer of the *Hfr* chromosome, with the result that the length of the segment transferred varies from one mating pair to the other. As a result of transfer a partial zygote or *merozygote* is formed, which comprises the whole F^- recipient and a chromosomal segment of *Hfr* origin.

(iii) The third step involves genetic recombination proper, that is the series of events which lead to the *integration* of *Hfr* markers to form a recombinant chromosome.

(iv) The fourth step, or *expression*, comprises the events which, through segregation and phenotypic expression, extend from the formation of a

recombinant chromosome to that of a fully expressed recombinant bacterium.

Of these four main steps, the last two, integration and expression, are common to the different known processes of genetic transfer in bacteria, i.e. conjugation, transformation and transduction. Analysis of the results obtained in the study of bacterial conjugation involves the comparison, in each case, of the nature and extent of genetic transfer from donor bacteria to the zygotes with the subsequent integration of these characters to recombinants. The characters transmitted with high frequency by strain *HfrH*, the strain originally described by Hayes (1953*b*) are represented in Fig. 1. The order and relative distances of the characters located on this segment may be determined, as indicated in the legend, both by the genetic analysis of recombinants and by the time at which these characters penetrate, during transfer, into the F^- recipient.



A	—	—	90	70	40		25	15
B	8	8½	9	11	18		25	26

Fig. 1. Genetic map of the segment injected with high frequency by the *Hfr* isolated by Hayes. The location of the different characters as measured in crosses *Hfr* × F^- by *A*: the percentage of $T^+L^+S^r$ recombinants which have inherited the different *Hfr* alleles; *B*: the time at which individual *Hfr* characters start penetrating into the F^- recipient in a Waring blendor experiment.

The mating systems of E. coli *K-12*

When an attempt is made to compare conjugation between *Hfr* and F^- bacteria on the one hand, and between F^+ and F^- bacteria on the other, the question arises as to whether the low frequency of recombination observed in the latter crosses is a consequence of a low frequency of conjugation, of transfer, or of integration. In $F^+ \times F^-$ crosses, the F^+ character itself is transmitted at high frequency to F^- cells (Cavalli *et al.* 1953) and this is also the case for the ability to produce certain colicins (Fredericq & Betz-Bareau, 1953). This indicates that the frequency of effective contacts, and hence of conjugation, is as high in $F^+ \times F^-$ crosses as in *Hfr* × F^- crosses (Jacob & Wollman, 1955), a prediction which is verified by microscopic studies (Anderson *et al.* 1957). Other lines of evidence support the hypothesis that low frequency of recombination is a consequence of a low frequency of transfer.

The question therefore arises of the origin of those recombinants which are formed at low frequency in $F^+ \times F^-$ crosses. They could either result from a low but constant probability for each F^+ cell to transfer the character under consideration—or from the presence, in F^+ cultures, of a small proportion of *Hfr* mutants of variable nature. Evidence for the validity of

the latter hypothesis comes from both the quantitative and qualitative results of 'fluctuation tests' as well as from the possibility of isolating, with a high yield, those *Hfr* mutants which are responsible for the formation of recombinants of any type in $F^+ \times F^-$ crosses (Jacob & Wollman, 1956*a*).

It may be pointed out that both mechanisms could play a part in the formation of recombinants in $F^+ \times F^-$ crosses. The demonstration that most of these recombinants, if not all of them, are formed by *Hfr* mutants present in the F^+ population, however, indicates that this mechanism is the prevailing one. If any recombinant were formed by F^+ donors, the detection of this latter mechanism would be extremely difficult, and there is, at the present time, no experimental evidence for its existence.

The patterns of chromosome transfer in Hfr *mutants*

The finding that most, if not all, recombinants formed in an $F^+ \times F^-$ cross are due to *Hfr* mutants has two main implications. On the one hand, the fact that any known marker of an F^+ may be transmitted to a recombinant, implies that there must exist *Hfr* mutants able to transfer such markers at high frequency. On the other hand, the fact that any *Hfr* strain, such as the strain of Hayes, may transfer at high frequency only a group of characters suggests that the different *Hfr* mutants present in an F^+ culture must differ as to the nature of the chromosomal segment they are able to transfer.

This is indeed what is found. Any known genetic character of *E. coli* K-12 may be transferred at high frequency by a given type of *Hfr* mutant. These mutants differ from each other in the nature of the chromosomal segment they are able to transfer at high frequency. No *Hfr* mutant has been isolated so far which can transfer at high frequency all the genetic markers known (Jacob & Wollman, 1956*a*).

When studying any *Hfr* strain, one may obtain, as shown in previous sections, two types of information. On the one hand, one may determine, by genetic analysis, which characters are transmitted at high frequency to recombinants as well as their relative frequency of transmission. On the other hand, one may determine, by a blendor experiment, the sequence in which these characters are transferred and the time at which any given character enters the recipient.

Such an analysis has been carried out with a variety of *Hfr* mutants isolated from different strains of *E. coli* K-12 (Jacob & Wollman, 1957). Preliminary results are summarized in Table 1. It is seen that different *Hfr* strains differ from each other not only in those *characters* that they are able to transfer at high frequency, but also in the *order* in which any group of characters may be transferred. The simplest representation of these results

consists in assuming that, for any given *Hfr* strain, a chromosomal segment is transferred through a given extremity O (the origin) and that it is the position of O which determines which characters are transferred at high frequency, as well as the order of their transfer. A remarkable feature of the genetic system of *E. coli* K-12 appears to be the existence of a predetermined pattern of arrangement of the genetic characters that an *Hfr* mutation will affect by determining the position of O and hence the orientation of transfer. When considering a given genetic character B, it is found to be linked to a character A on one side and to a character C on the other side, these linkage relationships being retained as long as O is not

Table 1. *Patterns of Hfr mutants*

Schematic representation of the characters injected with high frequency by some of the isolated *Hfr* mutants. Each line corresponds to an *Hfr* strain and the order of injection corresponds to the characters from left to right.

Type	O←
H	TL Az T_1 Lac T_6 Gal λ
1	L T B_1 M Mtol Xyl Mal S^r
2	T_1 Az L T B_1 M Mtol Xyl Mal S^r
3	T_6 Lac T_1 Az L T B_1 M Mtol Xyl Mal S^r
4	B_1 M Mtol Xyl Mal S^r λ Gal
5	M B_1 TL Az T_1 Lac T_6 Gal λ

located closer to B than either A or C. When B becomes linked to O on one side, either it remains linked to A, and then becomes completely unlinked to C, or it remains linked to C and then becomes completely unlinked to A. For instance, in an *Hfr* of type 5 (Table 1), the characters TL appear to be linked to M (methionine) on the one side and to Gal on the other side, the order of transfer being O-M-TL-Gal. In the classical *HfrH*, characters TL are now linked to O on one side and to Gal on the other side, the order of transfer being O-TL-Gal. On the contrary, in *Hfr* of type 2, TL are linked on one side to O, but on the other side they are linked to M and the sequence of transfer has now become O-L-T-M.

Such a relationship is valid for all the known characters of K-12 and for all the *Hfr* strains which have been isolated up to now, whatever the previous history of the F^+ strains of K-12 from which they originate.

When trying to draw a genetic map of K-12, one is thus faced with a paradoxical situation. On the one hand, we do not possess any direct information on the genetic system of either F^+ or F^- bacteria. On the other hand, when analysing any given $Hfr \times F^-$ cross, we can only determine a genetic map of the chromosomal segment transferred at high frequency by that *Hfr* strain. We are therefore compelled to compare the genetic segments thus determined. From such a comparison, the conclusion is reached that all the known genetic characters of *E. coli* K-12 are

linked, a conclusion which was already attained by the analysis of $F^+ \times F^-$ crosses (Clowes & Rowley, 1954; Cavalli & Jinks, 1956), but that they cannot be arranged along a straight line as is the case in most chromosomes. There is no reason in effect for interrupting the determined linkage group at any place preferentially to any other one, except when describing the properties of a particular *Hfr* strain. One is thus led to dispose all the

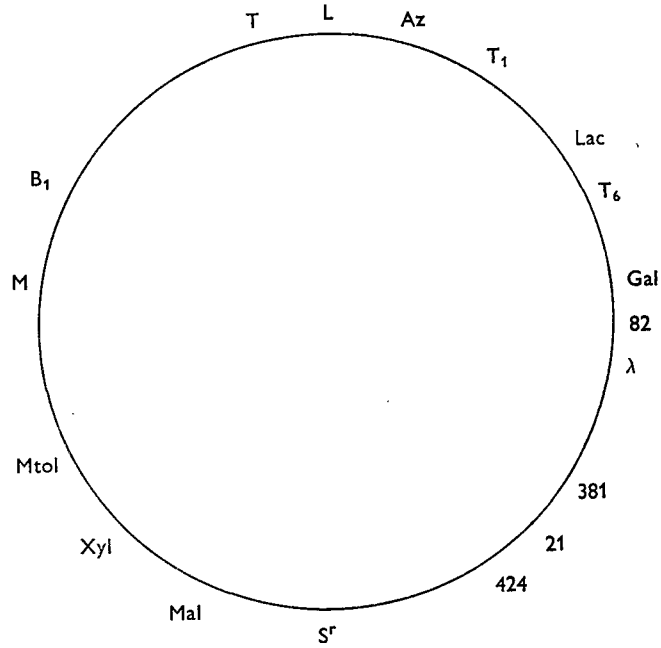


Fig. 2. Diagrammatic representation of the K-12 chromosome, as it results from a comparative study of the segments injected with high frequency by different *Hfr* strains. This diagram represents only the sequence of characters, not the distances between them. Symbols refer to threonine (T), leucine (L), methionine (M) and thiamine (B_1) synthesis; resistance to sodium azide (Az), bacteriophages T_1 and T_6, streptomycin (S); fermentation of lactose (Lac), galactose (Gal), maltose (Mal), xylose (Xyl) and mannitol (Mtol); ultra-violet inducible prophages 82, λ, 381, 21 and 424. (Possible mechanism of the $F^+ \rightarrow$ *Hfr* mutation: insertion of a specific factor at a given place of the circular linkage-group would result in the interruption of the circle. One extremity of the linkage group would behave as the origin. The other would carry the *Hfr* character.)

known genetic characters on a circle (Fig. 2). Since all *Hfr* mutants derive from F^+ strains, the underlying pattern of all *Hfr* would be that of the F^+ genetic system. It seems unnecessary to emphasize that this diagrammatic representation, which is the simplest one that will account for the observed results at the present time, is not meant to imply that the bacterial chromosome is actually circular.

Thus we have, on the one hand, in every individual *Hfr* a linear representation, one extremity of which is perfectly determined by the origin O,

and, on the other hand, a circular representation for the whole sequence. Under this scheme, the pattern of any existing *Hfr* may be obtained by interruption of the circle at the proper place, one of the extremities thus formed becoming the origin O.

Once the complete sequence of the K-12 linkage group has been determined by comparing the segments transmitted with high frequency by the various *Hfr* strains, one may analyse the capacity of a given *Hfr* type to transmit to recombinants those characters which appear only with low frequency. It seems that any *Hfr* type is indeed able to transmit all the known characters, but the frequency with which a given character is transmitted decreases the further from O this character is located in the particular *Hfr* strain. Such a genetic polarity may equally be interpreted as resulting from a gradient of pairing (Cavalli & Jinks, 1956) or from a gradient of transfer (Wollman & Jacob, 1957). The available evidence seems to indicate that the genetic polarity does reflect a polarity of transfer. The further from O a character is located, the greater the chance that a break will prevent its transfer during conjugation. When characters are located at a given distance from O, the probability for their transfer becomes so small that they appear very rarely among recombinants. These are the characters which are said to be transmitted to recombinants at low frequency in an $Hfr \times F^-$ cross.

Among characters transmitted with low frequency seems to be the *Hfr* character itself. Several of the isolated *Hfr* strains have been analysed for their capacity to transmit their *Hfr* character. With these strains, the recombinants having inherited from the *Hfr* parent the characters located close to O and transmitted with high frequency (proximal characters) are all F^-. On the contrary, among the few recombinants which have inherited from the *Hfr* parent characters located close to the extremity opposite to O (terminal characters), many are *Hfr*. When crossed with F^- bacteria, these *Hfr* recombinants appear to inject their markers in the same order as the one found for the *Hfr* parent from which they were derived. These properties of *Hfr* strains are similar to those already described for the two *Hfr* strains which were first isolated (Hayes, 1953*b*; Cavalli & Jinks, 1956). It appears, therefore, that the *Hfr* character does segregate among recombinants but that its linkage to other markers depends upon the strain considered. In any given *Hfr* strain, the *Hfr* character seems to be located at, or close to, the terminal extremity of the linkage group.

If confirmed by further investigation, this could be interpreted by the assumption that the event, which, in the $F^+ \rightarrow Hfr$ mutation, is supposed to result in the rupture of the circle at a given point, not only would determine the position of O at one of the two ends of the linkage group, but also the

location of the *Hfr* character at the other end. The properties of the different *Hfr* strains could thus be accounted for by the single hypothesis that the insertion of a specific factor at the proper place in the circular linkage group would determine the rupture of the circle. One extremity of the linkage group would behave as the origin O and would be injected during the process of mating. The other extremity would be terminal and would carry the *Hfr* character.

Again it must be clearly stated that this hypothesis provides only a formal model and that other schemes might well explain the experimental results. The proposed model, however, appears to be the simplest one that accounts for all the data. Although the nature of the postulated factor, whose insertion would be responsible for the breaking of the circle, remains unknown, it might well be conceived as being similar to the 'controlling elements' described in maize by McClintock (1956).

One must conclude that any *Hfr* may, upon conjugation, inject into the F^- recipient, a particular sequence of genetic characters in an oriented way. The distance between these characters may be determined by genetic analysis, when these characters are not very far apart. For distant characters, genetic analysis is grossly deformed by the increasing probability of breaks occurring, during transfer, between these characters. A more precise determination of the distances may be obtained by measurement of the time at which different characters penetrate the F^- recipient during conjugation.

II. EFFECT OF ULTRA-VIOLET LIGHT AND ^{32}P DECAY ON BACTERIAL RECOMBINATION

Radiation and disintegration of radioactive phosphorus are known to induce various types of cellular lesions and more particularly local alterations in the genetic material. In bacterial conjugation, whereas the F^- recipient contributes to the zygote both its genetic material and its cytoplasm, the *Hfr* donor appears to contribute genetic material but very little, if any, cytoplasm. Exposure of the *Hfr* donor to physical agents is therefore likely to affect only one of the genetic elements taking part in recombination. Lesions induced in the *Hfr* may affect either the transfer of genetic material or its integration in the zygote after transfer.

As shown in previous sections, only a segment of the *Hfr* chromosome may be investigated in a cross between a given *Hfr* strain and a recipient F^-. The following discussion will only be concerned with the segment injected at high frequency by the strain *HfrH* isolated by Hayes (Fig. 1).

When crosses are performed between *HfrH* $T^+L^+Az^sT_1^s\ Lac_1^+\ Gal_b^+\ S^s$

and *P678* $F^-T^-L^-Az^rT_1^r\ Lac_1^-\ Gal_b^-\ S^r$,* a chromosomal segment O-TL-Gal (Fig. 1) of the *Hfr* is transmitted at high frequency. The S^s marker not being transferred at high frequency, streptomycin may be used for eliminating the *Hfr* parent after mating. By plating the zygotes on suitable selective media, one may select, in the same experiment, recombinants which have inherited certain characters of the *Hfr* parent, either *proximal* to O, such as T^+L^+ which are close enough to be used together as selective markers (recombinants $T^+L^+S^r$) or distal to O such as Gal^+ (recombinants Gal^+S^r). One may also select those recombinants which have received the T^+L^+ as well as the Gal^+ characters of the *Hfr* (recombinants $T^+L^+Gal^+S^r$). In such an experiment, one may compare the frequencies with which these different types of recombinants are formed as well as their genetic constitution, that is the distribution among these recombinants of the genetic markers contributed by the *Hfr* parent.

The effect of ultra-violet irradiation

If *Hfr* donors are first exposed to ultra-violet light and then mated with non-irradiated F^- recipients, the irradiated *Hfr* lose their ability to form recombinants exponentially as a function of the dose of ultra-violet (Fig. 3). The capacity to form either $T^+L^+S^r$ recombinants or Gal^+S^r recombinants is lost at the same rate. This indicates that the transfer of the distal markers is not more reduced than that of the proximal ones. It is not known whether ultra-violet light may or may not prevent transfer, but if it does, it is not by reducing the size of the transferred piece.

On the contrary, the capacity to form recombinants which have received *both* extremities of the TL-Gal segment ($T^+L^+Gal^+S^r$ recombinants) is lost at a faster rate. One must, therefore, conclude that lesions produced in the *Hfr* parent before mating alter the process of integration occurring in the zygote after transfer. They decrease the probability of simultaneous integration of distant markers.

The difference observed between the slopes of the $T^+L^+S^r$ and $T^+L^+Gal^+S^r$ recombinants indicates that the fraction of the $T^+L^+S^r$ recombinants which have inherited the Gal^+ marker from the *Hfr* parent decreases as a function of the dose of ultra-violet. One may therefore expect to observe an alteration of the genetic constitution of the recombinants formed by irradiated *Hfr* parents. As shown in Table 2, the frequency with which unselected markers of the *Hfr* are present in both $T^+L^+S^r$ and Gal^+S^r recombinants is strikingly reduced by irradiation. This means that irradiation of the *Hfr* parent results in a loosening of the linkage observed

* Synthesis of threonine (T), leucine (L), fermentation of lactose (Lac), galactose (Gal), sensitivity (s) or resistance (r) to sodium azide (Az), to phage T1 and to streptomycin (S).

between the markers of the TL-Gal segment. In other words, the probability of a crossover occurring between two markers is increased by irradiating the donor.

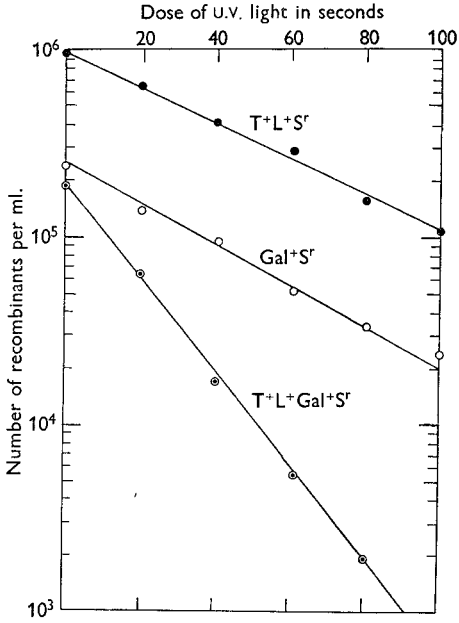


Fig. 3. Recombination between ultra-violet irradiated *Hfr* donor and non-irradiated F^- recipient. A suspension of *HfrH* is exposed to various doses of ultra-violet light. Samples of irradiated suspensions are mixed in standard conditions with non-irradiated *P678* F^- recipients and aliquots are plated on different selective media. The number of recombinants $T^+L^+S^r$, Gal^+S^r and $T^+L^+Gal^+S^r$/ml. of mating mixture are plotted on a logarithmic scale versus the dose of ultra-violet light in seconds.

Table 2. *Effect of ultra-violet irradiation on recombination*

Genetic composition of the recombinants formed in different crosses between *HfrH* and *P678* F^-. In the two last crosses, one of the parents was submitted to a standard dose of ultra-violet leaving about 30–50% survivors. The figures represent the percentage of recombinants $T^+L^+S^r$ or Gal^+S^r having inherited the characters Az, T_1, Lac and Gal of the *Hfr* parent.

	Genetic constitution of recombinants							
	$T^+L^+S^r$				Gal^+S^r			
Crosses	Az	T_1	Lac	Gal	TL	Az	T_1	Lac
Hfr × F^- control	91	72	52	29	74	78	73	74
Hfr U.V. × F^-	52	34	11	3	16	21	29	36
Hfr × F^- U.V.	88	60	35	16	41	49	57	62

Irradiation of the recipient also exerts some effect on the genetic constitution of recombinants, but to a much smaller extent than irradiation of the donor (Table 2).

This action of ultra-violet light on bacterial recombination appears to be similar to the effect which the same doses of ultra-violet exert on phage

recombination (Jacob & Wollman, 1955). Both effects can be best interpreted according to the hypothesis that, in bacteria as well as in phages, genetic recombination does not occur by breakage and reunion of complete strands but rather by a mechanism connected with replication, which, having commenced on one chromosome, would shift and be finished on the other (Levinthal, 1954).

Lesions produced randomly on the irradiated genetic material would interfere with the process of replication and therefore increase the frequency of the replication shifting from one parental chromosome to the other. This would result in a decrease of the size of the piece, or pieces, contributed by the irradiated parent to recombinants. Hence an apparent stretching of the linkage group.

The effect of ^{32}P *disintegration*

It is known that, like bacteriophages (Hershey, Kamen, Kennedy & Gest, 1951), bacteria containing ^{32}P of high specific radioactivity lose their viability as a function of ^{32}P decay (Fuerst & Stent, 1956). This method is of special interest since the lethal effects of ^{32}P decay cannot be accounted for by the ionizations produced, but appear rather to result mainly from 'short range' consequence of radioactive disintegrations, such as the transmutation $^{32}P \rightarrow {}^{32}S$ or the recoil energy sustained by the decaying P nucleus (Hershey *et al.* 1951; Stent, 1953). The favoured hypothesis is that the lethal effect of ^{32}P decay is due to those disintegrations which produce a rupture in the DNA chain.

This method may be applied to the study of bacterial recombination by mating ^{32}P-labelled *Hfr* donors with non-radioactive F^- recipients. By comparing the results of two types of experiments, it is possible to distinguish between the effects of ^{32}P decay on integration and its effects on transfer.

In one type of experiment radioactive *Hfr* donors are mated with non-radioactive recipients and ^{32}P decay allowed to occur *in the zygotes just after mating*, by storage in the cold. The capacity of the zygotes to give rise to recombinants is then measured as a function of ^{32}P decay. This type of experiment makes it possible to measure the effect of ^{32}P decay on *integration* independently of any effect on transfer.

The second type of experiment consists in allowing ^{32}P decay to occur *in the* Hfr *donors before mating*. Radioactive *Hfr* bacteria are stored in the cold and their *mating capacity*, that is their ability to transmit genetic characters to recombinants upon mating with F^- non-radioactive recipients, is measured as a function of ^{32}P decay. In this type of experiment radioactive disintegration may affect both the transfer of the genetic material and its integration (Fuerst, Jacob & Wollman, 1956, 1958).

(1) *The effects of* ^{32}P *decay occurring in the zygotes after mating.* When radioactive *Hfr* donors are mated with non-radioactive F^- recipients in non-radioactive medium and the early zygotes formed are stored in the cold, it is found that the capacity of these zygotes to form recombinants of any type decreases as a function of the time of storage, that is of ^{32}P decay (Fig. 4). This instability of the capacity of the early zygotes formed indicates that the genetic segment contributed by the *Hfr* parent remains susceptible to radioactive disintegration even after its transfer to the zygote. This is proof that the effects of ^{32}P decay are indeed a consequence of events occurring in the genetic material.

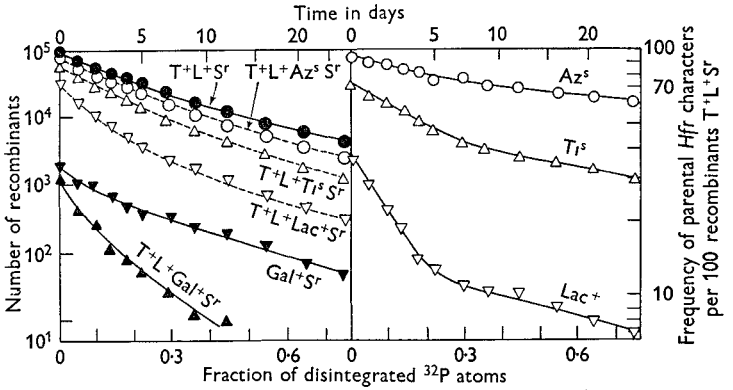


Fig. 4. Effect of ^{32}P decay occurring in the zygotes. Bacteria *HfrH* $T6S^s$ grown in a medium containing 110 mc./mg. of radio-phosphorus are washed, resuspended in buffer and mixed with an excess of non-radioactive *P678* F^- $T6^sS^r$, in a medium containing streptomycin, to prevent the *Hfr* parent from synthesizing nucleic acids. After 40 min. at 37°, KCN M/100 and an excess of phage T_6 are added to stop conjugation (Hayes, 1957). After 10 min. at 37°, the mixture is diluted in protective medium and samples are frozen in liquid nitrogen. Every day a sample is thawed and aliquots are plated on selective media.

The number of recombinants $T^+L^+S^r$, Gal^+S^r and $T^+L^+Gal^+S^r$ (left, solid lines) and the *proportion* of $T^+L^+S^r$ recombinants having one of the *Hfr* characters Az^s, $T1^s$, Lac^+ and Gal^+ (right) are plotted on a logarithmic scale versus the time in days and the fraction of disintegrated ^{32}P atoms. On the left figure are also plotted in dotted lines the *numbers* of recombinants $T^+L^+S^r$ having one of the *Hfr* characters Az^s, $T1^s$, Lac^+ and Gal^+, as calculated from the curves on the right.

When, in the same experiment, zygotes are sampled at different time intervals after mating, it is found that the capacity of the zygotes to form recombinants is the less sensitive to ^{32}P decay the later the time of sampling and that it becomes practically insensitive for the zygotes sampled after 120 min. This indicates that the genetic information carried by the injected chromosomal segment of the radioactive *Hfr* has been transferred to non-radioactive material.

Genetic analysis of the recombinants formed by early zygotes in the course of ^{32}P decay gives the following information. First of all, as may be

seen on Fig. 4, the number of recombinants which inherit either the T^+L^+ characters or the Gal^+ character of the *Hfr* parent decreases at about the same rate, whereas the number of those recombinants which inherit *both* the T^+L^+ and the Gal^+ characters ($T^+L^+Gal^+S^r$ recombinants) decreases at a much faster rate. These results, which are comparable to those obtained after ultra-violet treatment, indicate that the effect of ^{32}P decay in the zygote is to decrease the integration of the transferred genetic characters. Analysis of the $T^+L^+S^r$ recombinants shows that the linkage of the T^+L^+ characters to any of the unselected markers located on the TL-Gal segment of the *Hfr* donor such as Az^s, T_1^s, Lac^+ or Gal^+ is the more sensitive to ^{32}P decay the farther the marker is from T^+L^+. The initial slopes of the curves thus obtained depend on the order and relative distances of the characters located on the TL-Gal segment. These results suggest that ^{32}P disintegration destroys the integrity of genetic segments and that the sensitivity of any genetic segment is roughly proportional to its length.

It may be seen in Fig. 4 that the sensitivities of the linkages to T^+L^+ of the different markers decrease as a function of the time of storage and finally tend to a common slope. The explanation for this fact appears to be that the recombinants scored after a long time of storage correspond to multiple crossovers which simulate a reduction in size of the contribution of the *Hfr* chromosome.

(2) *The effects of* ^{32}P *decay occurring in the* Hfr *before mating.* When radioactive *Hfr* donors are stored and, after various time intervals, are mated with non-radioactive F^- recipients in non-radioactive medium, it is found that the capacity of these *Hfr* to form recombinants of any type decreases exponentially as a function of the time of storage (Fig. 5). However, their capacity for transmitting to recombinants the characters T^+L^+, *proximal* to O, decreases at a rate which is about one-third of the rate at which their capacity to transmit the *distal* Gal^+ character decreases. Moreover the capacity for transmitting *both* the T^+L^+ and the Gal^+ characters ($T^+L^+Gal^+S^r$ recombinants) is lost at the same rate as the capacity to transmit the Gal^+ character alone (Gal^+S^r recombinants). This suggests that when ^{32}P disintegration takes place in the *Hfr*, it affects not only the process of integration but also the transfer of genetic material from the donor to the recipient. Everything happens as though breaks had occurred on the *Hfr* genetic segment, certain *Hfr* cells still being able to transfer a proximal piece but not a distal one. Such an interpretation of the results is in agreement with the hypothesis according to which the lethal effects of ^{32}P decay result from interruptions on the DNA chain (Stent & Fuerst, 1955).

Analysis of the $T^+L^+S^r$ recombinants as to the presence of the genetic markers of the TL-Gal segment of the *Hfr* reveals that linkages $T^+L^+Az^s$, $T^+L^+T_1^s$, $T^+L^+Lac^+$, etc. are the more sensitive, the farther away the marker is, considered from the T^+L^+ selected characters (Fig. 5). By comparing the rates of survival to ^{32}P decay of the linkages considered, one may determine the order and relative distances of the markers. The fact, however, that the radioactive *Hfr* lose their capacity for transmitting both the T^+L^+ and Gal^+ characters together ($T^+L^+Gal^+S^r$ recombinants) at the same rate as they lose their capacity of transmitting the Gal^+ character alone (Gal^+S^r

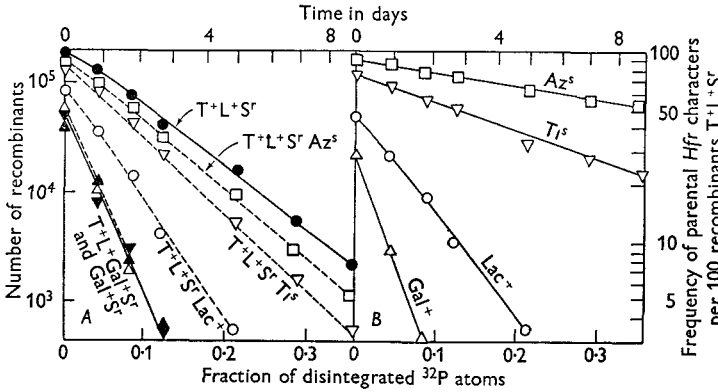


Fig. 5. Effect of ^{32}P decay occurring in the *Hfr* before mating. Bacteria *HfrH* grown in a medium containing 70 mc./mg. of phosphorus are centrifuged, washed and resuspended in protective medium. Samples are frozen in liquid nitrogen. Every day a sample is thawed, resuspended in buffer and mixed in standard conditions with an excess of non-radioactive recipient *P678* F^- ($2 \cdot 10^6$ *Hfr* and 10^8 F^-/ml.). After one hour at 37°, samples are diluted and aliquots are plated on selective media.

The number of recombinants $T^+L^+S^r$, Gal^+S^r and $T^+L^+Gal^+S^r$ (left solid lines) and the *proportion* of $T^+L^+S^r$ recombinants having one of the *Hfr* characters Az^s, T_1^s, Lac^+ (right) are plotted on a logarithmic scale versus the time in days and the fraction of disintegrated ^{32}P atoms. On the left figure are also plotted in dotted lines the *numbers* of $T^+L^+S^r$ recombinants having one of the *Hfr* characters Az^s, T_1^s, Lac^+ as calculated from the curves on the right.

recombinants), indicates that the slopes of the curves of Fig. 5 represent the rate of survival of the capacity of the *Hfr* to transmit any given character to a recombinant. Since these survival curves remain exponential over a long period, the mechanism of this loss in transmissibility of genetic characters appears to be mainly, if not exclusively, an exponential decrease of their transfer. The slopes of these survival curves would then be proportional to the distance of the various characters from the extremity O of the chromosomal segment. When the relative rates of decay of the individual markers are compared (T^+L^+ 0·35, Az^s 0·39, T_1^s 0·43, Lac^+ 0·67, Gal^+ 1), they are found indeed to be very similar to the relative times of entry of the same markers in a blendor experiment (T^+L^+ 0·34, Az^s 0·36, T_1^s 0·44, Lac^+ 0·72, Gal^+ 1).

Disintegration of radiophosphorus therefore allows the determination of the distances between genetic markers and the genetic maps which may thus be drawn of chromosomal segments transferred at high frequency by *Hfr* strains, are in every respect comparable with the maps, which may be determined on the one hand by genetic analysis and on the other hand by the time at which genetic characters are transferred in the course of conjugation.

III. DISCUSSION

Recombination of genetic characters in bacteria may be achieved by different processes which all involve the transfer of genetic material from a donor into a recipient bacterium. In the three known processes in which this genetic transfer occurs there is little doubt that the information is carried by DNA. In transformation, pieces of naked DNA are directly taken in from the culture medium into the recipient bacteria. In transduction, it is apparently a piece of bacterial DNA which has become included within the protein coat of a phage and which, together with the DNA of the phage, is injected into the recipient bacterium upon infection. In bacterial conjugation, the process of genetic transfer is accomplished by a more elaborate mechanism, which involves contact between bacteria of opposite mating types and the oriented injection of a chromosomal segment from the donor to the recipient. Non-chromosomal material, such as an enzyme or vegetative phage, has been found not to be transferred to any measurable extent during conjugation. Although such evidence is not definite, it suggests that, if there is any transfer of cytoplasmic constituents, it must be limited as compared with that of the genetic material. That this genetic material is nucleic acid is indicated by the effects on the formation of recombinants of ^{32}P disintegration occurring in the zygotes when ^{32}P-labelled *Hfr* is mated with non-radioactive recipients.

The most characteristic feature of bacterial conjugation is the progressive transfer of an oriented chromosomal segment of the donor which always injects the same extremity first into the recipient. The polarity of the transfer offers an opportunity of determining the position and distances of the genetic characters located on the chromosomal segment transferred by different independent methods. We will briefly compare the results obtained therewith.

Genetic analysis would allow an evaluation of the distances between characters in terms of frequency of recombination between such characters. However, the frequencies thus determined may vary according to the class of recombinants selected. This results both from the incomplete nature of the zygotes which contain the complete genome of the recipient and only

a fragment of the donor's, and from the variability in length of this fragment among different zygotes. By selecting recombinants for characters proximal to O (such as the $T^+L^+S^r$ recombinants obtained with *HfrH*), the recombination frequencies which may be measured (Fig. 1) would reflect both the real probability of recombination occurring between two characters and the heterogeneity in size of the segments transferred. Correction for this latter effect, which is formally comparable to a defect in pairing of homologous chromosomes (Cavalli & Jinks, 1956), may be obtained by different methods. One of them is to select those zygotes which have received chromosomal segments of equal length. This can be achieved by comparing normal crosses with crosses between lysogenic *Hfr* and non-lysogenic F^- where zygotic induction of the prophage selects those zygotes which have not received the prophage (Jacob & Wollman, 1956*b*; Wollman & Jacob, 1957). Another method consists in assuming a constant probability of breakage per length unit of the chromosome and determining the correction factor by trial and error. By combination of these two methods corrected values of the distances between genetic characters have been obtained which seem in good agreement with the results obtained with other methods (Jacob & Wollman, unpublished).

Blendor experiments permit a representation of the distances between genetic characters in time units by comparing their relative time of penetration into the recipient. The validity of this representation depends upon whether or not the transferred segment proceeds at a constant rate. Comparison between experiments done at different temperatures (Fisher, 1957), and even more the comparison between blendor and ^{32}P disintegration experiments indicate that this is the case for the characters located on the TL-Gal segment, that is along a segment which represents about one-third of the total chromosome length. The rate of penetration appears to decrease for characters located farther. It may be calculated that penetration of the whole chromosome at a constant rate would take about a hundred minutes.

Disintegration of radiophosphorus ^{32}P in the *Hfr* prior to mating appears to prevent the genetic markers from being transferred at a rate proportional to their distance from the origin O, that is to the number of phosphorus atoms existing between O and the marker considered. The genetic lesions provoked by ^{32}P decay appear, therefore, to be the direct consequence of the $^{32}P \rightarrow {}^{32}S$ transmutations which occur in the phosphodiester bonds which form the backbone of the nucleotide chain of DNA. On assuming that the efficiency of any radioactive disintegration is identical all along the chromosomal segment, a direct relationship may be established between the genetic and the physical structures of this segment. However, in order to evaluate

the distances between markers in terms of number of phosphorus atoms, one must know what is the efficiency with which ^{32}P disintegration prevents transfer. In the absence of any direct information on this efficiency, the assumption may be made, that it is the same as for the lethal effects in bacteria and in phage.

It becomes thus possible to compare the different measures of the chromosome of *E. coli* K-12. It is thus found that one minute in penetration corresponds roughly to about 20 corrected recombination units and to about 10^5 nucleotide pairs. This last figure is in agreement with an independent estimate of the DNA content of an *E. coli* nucleus, which would amount to about 10^7 nucleotide pairs, and of the time of transfer of the whole chromosome at a constant rate, which would take about 100 min. The three independent methods of measuring genetic segments in *E. coli* K-12 are therefore in reasonable agreement.

IV. SUMMARY

(1) Conjugation between *Hfr* donor and F^- recipient corresponds to the oriented transfer, under a rather precise time schedule, of a chromosomal segment from the donor to the recipient. The order and the distance between the characters located on the segment may be determined both by genetic analysis and by the time at which individual *Hfr* characters penetrate into the recipient.

(2) Various types of *Hfr* mutants may be isolated which differ by the characters they are able to transfer with high frequency and by the order in which the characters are transferred.

(3) By comparing the segments transferred with high frequency by various *Hfr* mutants, it is possible to determine a complete sequence of the known markers. Since no interruption can be observed, no extremity can be determined in the linkage group. One is thus led to dispose all the known genetic characters on a circle. In this model *Hfr* mutants would result from the insertion of a factor and the consequent interruption of the circle.

(4) Exposure of the *Hfr* donors to ultra-violet light before mating impairs integration of the characters located on the transferred segment. The number of genetic exchanges occurring between two given characters is increased by irradiation. This suggests that recombination in bacteria does not occur by breakage and reunion but rather by some 'copying choice' mechanism.

(5) If *Hfr* donors containing ^{32}P are mated with non-radioactive recipients, and if the mating process is interrupted early in order to allow

^{32}P decay to occur, the zygotes lose, as a function of time, their capacity to produce recombinants. Radiophosphorus disintegration appears to act by altering integration of the transferred *Hfr* markers.

(6) If radioactive *Hfr* are stored, in order to allow ^{32}P decay to occur before they are mated with non-radioactive recipients, they lose, as a function of time, their capacity to produce recombinants. Radiophosphorus disintegration appears to act by altering both transfer and integration of genetic material.

(7) It is possible to determine, by this method, the order and the distances between the markers located on the *Hfr* segment transferred with high frequency. The genetic map thus obtained does not differ from the map determined by other methods. It appears, therefore, possible to relate information on K-12 genetic material gained by genetic and by physical methods.

REFERENCES

ANDERSON, T. F., WOLLMAN, E. L. & JACOB, F. (1957). *Ann. Inst. Pasteur*, **93**, 450.

CAVALLI, L. L. (1950). *Boll. Ist. Sierot. Milano*, **29**, 1.

CAVALLI, L. L. & JINKS, J. L. (1956). *J. Genet.* **54**, 87.

CAVALLI, L. L., LEDERBERG, J. & LEDERBERG, E. M. (1953). *J. Gen. Microbiol.* 8, 89.

CLOWES, R. & ROWLEY, D. (1954). *J. Gen. Microbiol.* **11**, 250.

FISHER, K. W. (1957). *J. Gen. Microbiol.* **16**, 136.

FREDERICQ, P., & BETZ-BAREAU, M. (1953). *C.R. Soc. Biol., Paris*, **147**, 1653.

FUERST, C. R., JACOB, F. & WOLLMAN, E. L. (1956). *C.R. Acad. Sci., Paris*, **243**, 2162.

FUERST, C. R., JACOB, F. & WOLLMAN, E. L. (1958). *Ann. Inst. Pasteur* (in Preparation).

FUERST, C. R. & STENT, G. S. (1956). *J. Gen. Physiol.* **40**, 73.

HAYES, W. (1953*a*). *J. Gen. Microbiol.* **8**, 72.

HAYES, W. (1953*b*). *Cold Spr. Harb. Symp. Quant. Biol.* **18**, 75.

HAYES, W. (1957). *J. Gen. Microbiol.* **16**, 97.

HERSHEY, A. D., KAMEN, M. D., KENNEDY, J. W. & GEST, H. (1951). *J. Gen. Physiol.* **34**, 305.

JACOB, F. & WOLLMAN, E. L. (1955). *C.R. Acad. Sci., Paris*, **240**, 2566.

JACOB, F. & WOLLMAN, E. L. (1956*a*). *C.R. Acad. Sci., Paris*, **242**, 303.

JACOB, F. & WOLLMAN, E. L. (1956*b*). *Ann. Inst. Pasteur*, **91**, 486.

JACOB, F. & WOLLMAN, E. L. (1957). *C.R. Acad. Sci., Paris*, **245**, 1840.

LEDERBERG, J. (1947). *Genetics*, **32**, 505.

LEVINTHAL, C. (1954). *Genetics*, **39**, 169.

MCCLINTOCK, B. (1956). *Cold Spr. Harb. Symp. Quant. Biol.* **21**, 197.

STENT, G. S. (1953). *Cold Spr. Harb. Symp. Quant. Biol.* **18**, 255.

STENT, G. S. & FUERST, C. R. (1955). *J. Gen. Physiol.* **38**, 441.

TATUM, E. L. & LEDERBERG, J. (1947). *J. Bact.* **53**, 673.

WOLLMAN, E. L. & JACOB, F. (1955). *C.R. Acad. Sci., Paris*, **240**, 2449.

WOLLMAN, E. L. & JACOB, F. (1957). *Ann. Inst. Pasteur*, **93**, 323.

WOLLMAN, E. L., JACOB, F. & HAYES, W. (1956). *Cold Spr. Harb. Symp. Quant. Biol.* **21**, 141.

GENERAL NATURE OF THE GENETIC CODE FOR PROTEINS

By Dr. F. H. C. CRICK, F.R.S., LESLIE BARNETT, Dr. S. BRENNER and Dr. R. J. WATTS-TOBIN

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge

THERE is now a mass of indirect evidence which suggests that the amino-acid sequence along the polypeptide chain of a protein is determined by the sequence of the bases along some particular part of the nucleic acid of the genetic material. Since there are twenty common amino-acids found throughout Nature, but only four common bases, it has often been surmised that the sequence of the four bases is in some way a code for the sequence of the amino-acids. In this article we report genetic experiments which, together with the work of others, suggest that the genetic code is of the following general type:

(*a*) A group of three bases (or, less likely, a multiple of three bases) codes one amino-acid.

(*b*) The code is not of the overlapping type (see Fig. 1).

(*c*) The sequence of the bases is read from a fixed starting point. This determines how the long sequences of bases are to be correctly read off as triplets. There are no special 'commas' to show how to select the right triplets. If the starting point is displaced by one base, then the reading into triplets is displaced, and thus becomes incorrect.

(*d*) The code is probably 'degenerate'; that is, in general, one particular amino-acid can be coded by one of several triplets of bases.

The Reading of the Code

The evidence that the genetic code is not overlapping (see Fig. 1) does not come from our work, but from that of Wittmann[1] and of Tsugita and Fraenkel-Conrat[2] on the mutants of tobacco mosaic virus produced by nitrous acid. In an overlapping triplet code, an alteration to one base will in general change three adjacent amino-acids in the polypeptide chain. Their work on the alterations produced in the protein of the virus show that usually only one amino-acid at a time is changed as a result of treating the ribonucleic acid (RNA) of the virus with nitrous acid. In the rarer cases where two amino-acids are altered (owing presumably to two separate deaminations by the nitrous acid on one piece of RNA), the altered amino-acids are not in adjacent positions in the polypeptide chain.

Brenner[3] had previously shown that, if the code were universal (that is, the same throughout Nature), then all overlapping triplet codes were impossible. Moreover, all the abnormal human hæmoglobins studied in detail[4] show only single amino-acid changes. The newer experimental results essentially rule out all simple codes of the overlapping type.

If the code is not overlapping, then there must be some arrangement to show how to select the correct triplets (or quadruplets, or whatever it may be) along the continuous sequence of bases. One obvious suggestion is that, say, every fourth base is a 'comma'. Another idea is that certain triplets make 'sense', whereas others make 'nonsense', as in the comma-free codes of Crick, Griffith and Orgel[5]. Alternatively, the correct choice may be made by starting at a fixed point and working along the sequence of bases three (or four, or whatever) at a time. It is this possibility which we now favour.

Experimental Results

Our genetic experiments have been carried out on the *B* cistron of the r_{II} region of the bacteriophage *T4*, which attacks strains of *Escherichia coli*. This is the system so brilliantly exploited by Benzer[6,7]. The r_{II} region consists of two adjacent genes, or 'cistrons', called cistron *A* and cistron *B*. The wild-type phage will grow on both *E. coli B* (here called *B*) and on *E. coli K*12 (λ) (here called *K*), but a phage which has lost the function of either gene will not grow on *K*. Such a phage produces an *r* plaque on *B*. Many point mutations of the genes are known which behave in this way. Deletions of part of the region are also found. Other mutations, known as 'leaky', show partial function; that is, they will grow on *K* but their plaque-type on *B* is not truly wild. We report here our work on the mutant *P* 13 (now re-named *FC* 0) in the *B*1 segment of the *B* cistron. This mutant was originally produced by the action of proflavin[8].

We[9] have previously argued that acridines such as proflavin act as mutagens because they add or delete a base or bases. The most striking evidence in favour of this is that mutants produced by acridines are seldom 'leaky'; they are almost always completely lacking in the function of the gene. Since our note was published, experimental data from two sources have been added to our previous evidence: (1) we have examined a set of 126 r_{II} mutants made with acridine yellow; of these only 6 are leaky (typically about half the mutants made with base analogues are leaky); (2) Streisinger[10] has found that whereas mutants of the lysozyme of phage *T4* produced by base-analogues are usually leaky, all lysozyme mutants produced by proflavin are negative, that is, the function is completely lacking.

If an acridine mutant is produced by, say, adding a base, it should revert to 'wild-type' by deleting a base. Our work on revertants of *FC* 0 shows that it usually

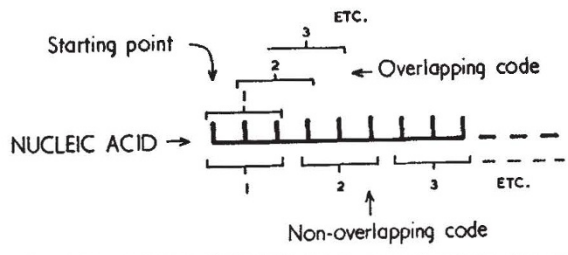


Fig. 1. To show the difference between an overlapping code and a non-overlapping code. The short vertical lines represent the bases of the nucleic acid. The case illustrated is for a triplet code

reverts not by reversing the original mutation but by producing a second mutation at a nearby point on the genetic map. That is, by a 'suppressor' in the same gene. In one case (or possibly two cases) it may have reverted back to true wild, but in at least 18 other cases the 'wild type' produced was really a double mutant with a 'wild' phenotype. Other workers[11] have found a similar phenomenon with r_{II} mutants, and Jinks[12] has made a detailed analysis of suppressors in the h_{III} gene.

The genetic map of these 18 suppressors of *FC* 0 is shown in Fig. 2, line *a*. It will be seen that they all fall in the *B*1 segment of the gene, though not all of them are very close to *FC* 0. They scatter over a region about, say, one-tenth the size of the *B* cistron. Not all are at different sites. We have found eight sites in all, but most of them fall into or near two close clusters of sites.

In all cases the suppressor was a non-leaky *r*. That is, it gave an *r* plaque on *B* and would not grow on *K*. This is the phenotype shown by a complete deletion of the gene, and shows that the function is lacking. The only possible exception was one case where the suppressor appeared to back-mutate so fast that we could not study it.

Each suppressor, as we have said, fails to grow on *K*. Reversion of each can therefore be studied by the same procedure used for *FC* 0. In a few cases these mutants apparently revert to the original wild-type, but usually they revert by forming a double mutant. Fig. 2, lines *b*–*g*, shows the mutants produced as suppressors of these suppressors. Again all these new suppressors are non-leaky *r* mutants, and all map within the *B*1 segment for one site in the *B*2 segment.

Once again we have repeated the process on two of the new suppressors, with the same general results, as shown in Fig. 2, lines *i* and *j*.

All these mutants, except the original *FC* 0, occurred spontaneously. We have, however, produced one set (as suppressors of *FC* 7) using acridine yellow as a mutagen. The spectrum of suppressors we get (see Fig. 2, line *h*) is crudely similar to the spontaneous spectrum, and all the mutants are non-leaky *r*'s. We have also tested a (small) selection of all our mutants and shown that their reversion-rates are increased by acridine yellow.

Thus in all we have about eighty independent *r* mutants, all suppressors of *FC* 0, or suppressors of suppressors, or suppressors of suppressors of suppressors. They all fall within a limited region of the gene and they are all non-leaky *r* mutants.

The double mutants (which contain a mutation plus its suppressor) which plate on *K* have a variety of plaque types on *B*. Some are indistinguishable from wild, some can be distinguished from wild with difficulty, while others are easily distinguishable and produce plaques rather like *r*.

We have checked in a few cases that the phenomenon is quite distinct from 'complementation', since the two mutants which separately are phenotypically *r*, and together are wild or pseudo-wild,

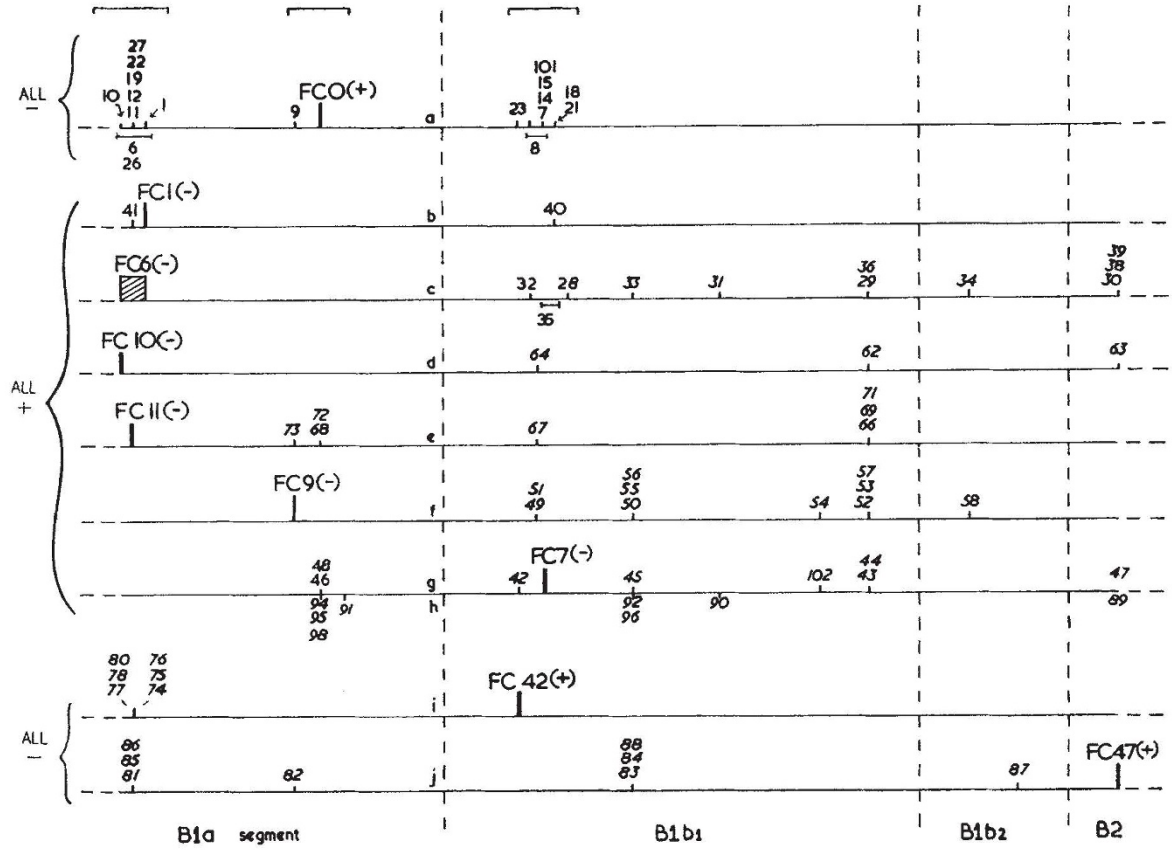


Fig. 2. A tentative map—only very roughly to scale—of the left-hand end of the *B* cistron, showing the position of the *FC* family of mutants. The order of sites within the regions covered by brackets (at the top of the figure) is not known. Mutants in italics have only been located approximately. Each line represents the suppressors picked up from one mutant, namely, that marked on the line in bold figures

must be put together in the same piece of genetic material. A simultaneous infection of *K* by the two mutants in separate viruses will not do.

The Explanation in Outline

Our explanation of all these facts is based on the theory set out at the beginning of this article. Although we have no direct evidence that the *B* cistron produces a polypeptide chain (probably through an RNA intermediate), in what follows we shall assume this to be so. To fix ideas, we imagine that the string of nucleotide bases is read, triplet by triplet, from a starting point on the left of the *B* cistron. We now suppose that, for example, the mutant *FC* 0 was produced by the insertion of an additional base in the wild-type sequence. Then this addition of a base at the *FC* 0 site will mean that the reading of all the triplets to the right of *FC* 0 will be shifted along one base, and will therefore be incorrect. Thus the amino-acid sequence of the protein which the *B* cistron is presumed to produce will be completely altered from that point onwards. This explains why the function of the gene is lacking. To simplify the explanation, we now postulate that a suppressor of *FC* 0 (for example, *FC* 1) is formed by deleting a base. Thus when the *FC* 1 mutation is present by itself, all triplets to the right of *FC* 1 will be read incorrectly and thus the function will be absent. However, when both mutations are present in the same piece of DNA, as in the pseudo-wild double mutant *FC* (0 + 1), then although the reading of triplets between *FC* 0 and *FC* 1 will be altered, the original reading will be restored to the rest of the gene. This could explain why such double mutants do not always have a true wild phenotype but are often pseudo-wild, since on our theory a small length of their amino-acid sequence is different from that of the wild-type.

For convenience we have designated our original mutant *FC* 0 by the symbol + (this choice is a pure convention at this stage) which we have so far considered as the addition of a single base. The suppressors of *FC* 0 have therefore been designated −. The suppressors of these suppressors have in the same way been labelled as +, and the suppressors of these last sets have again been labelled − (see Fig. 2).

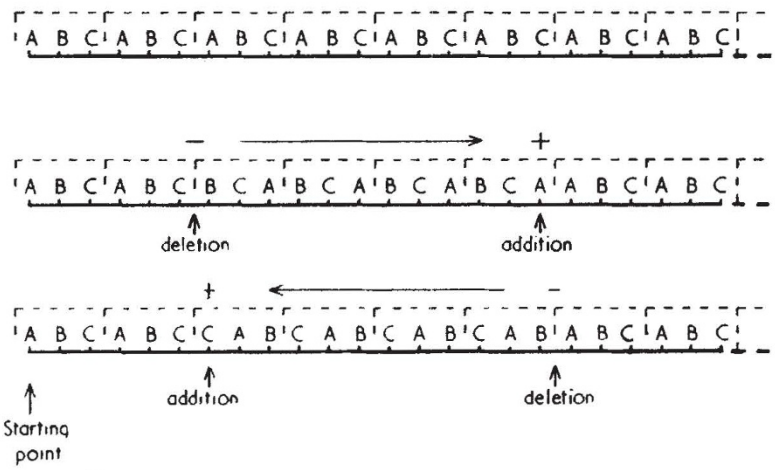


Fig. 3. To show that our convention for arrows is consistent. The letters *A*, *B* and *C* each represent a different base of the nucleic acid. For simplicity a repeating sequence of bases, *ABC*, is shown. (This would code for a polypeptide for which every amino-acid was the same.) A triplet code is assumed. The dotted lines represent the imaginary 'reading frame' implying that the sequence is read in sets of three starting on the left

Double Mutants

We can now ask: What is the character of any double mutant we like to form by putting together in the same gene any pair of mutants from our set of about eighty? Obviously, in some cases we already know the answer, since some combinations of a + with a − were formed in order to isolate the mutants. But, by definition, no pair consisting of one + with another + has been obtained in this way, and there are many combinations of + with − not so far tested.

Now our theory clearly predicts that all combinations of the type + with + (or − with −) should give an *r* phenotype and not plate on *K*. We have put together 14 such pairs of mutants in the cases listed in Table 1 and found this prediction confirmed.

Table 1. DOUBLE MUTANTS HAVING THE *r* PHENOTYPE

− With −	+ With +	
FC (1 + 21)	*FC* (0 + 58)	*FC* (40 + 57)
FC (23 + 21)	*FC* (0 + 38)	*FC* (40 + 58)
FC (1 + 23)	*FC* (0 + 40)	*FC* (40 + 55)
FC (1 + 9)	*FC* (0 + 55)	*FC* (40 + 54)
	FC (0 + 54)	*FC* (40 + 38)

At first sight one would expect that all combinations of the type (+ with −) would be wild or pseudo-wild, but the situation is a little more intricate than that, and must be considered more closely. This springs from the obvious fact that if the code is made of triplets, any long sequence of bases can be read correctly in one way, but incorrectly (by starting at the wrong point) in two different ways, depending whether the 'reading frame' is shifted one place to the right or one place to the left.

If we symbolize a shift, by one place, of the reading frame in one direction by → and in the opposite direction by ←, then we can establish the convention that our + is always at the head of the arrow, and our − at the tail. This is illustrated in Fig. 3.

We must now ask: Why do our suppressors not extend over the whole of the gene? The simplest postulate to make is that the shift of the reading frame produces some triplets the reading of which is 'unacceptable'; for example, they may be 'nonsense', or stand for 'end the chain', or be unacceptable in some other way due to the complications of protein structure. This means that a suppressor of, say, *FC* 0 must be within a region such that no 'unacceptable' triplet is produced by the shift in the reading frame between *FC* 0 and its suppressor. But, clearly, since for any sequence there are *two* possible misreadings, we might expect that the 'unacceptable' triplets produced by a → shift would occur in different places on the map from those produced by a ← shift.

Examination of the spectra of suppressors (in each case putting in the arrows → or ←) suggests that while the → shift is acceptable anywhere within our region (though not outside it) the shift ←, starting from points near *FC* 0, is acceptable over only a more limited stretch. This is shown in Fig. 4. Somewhere in the left part of our region, between *FC* 0 or *FC* 9 and the *FC* 1 group, there must be one or more unacceptable triplets when a ← shift is made; similarly for

the region to the right of the *FC* 21 cluster. Thus we predict that a combination of a + with a − will be wild or pseudo-wild if it involves a → shift, but that such pairs involving a ← shift will be phenotypically *r* if the arrow crosses one or more of the forbidden places, since then an unacceptable triplet will be produced.

Table 2. Double Mutants of the Type (+ with −)

− \ +	*FC* 41	*FC* 0	*FC* 40	*FC* 42	*FC* 58*	*FC* 63	*FC* 38
FC 1	*W*	*W*	*W*		W		W
FC 86		W	W	W	W	W	
FC 9	*r*	*W*	W	W	*W*		W
FC 82	*r*		W	W	W	W	
FC 21	*r*	*W*			W		W
FC 88	*r*	*r*			W	W	
FC 87	*r*	*r*	*r*	*r*			W

W, wild or pseudo-wild phenotype; *W*, wild or pseudo-wild combination used to isolate the suppressor; *r*, *r* phenotype.
* Double mutants formed with *FC* 58 (or with *FC* 34) give sharp plaques on *K*.

We have tested this prediction in the 28 cases shown in Table 2. We expected 19 of these to be wild, or pseudo-wild, and 9 of them to have the *r* phenotype. In all cases our prediction was correct. We regard this as a striking confirmation of our theory. It may be of interest that the theory was constructed before these particular experimental results were obtained.

Rigorous Statement of the Theory

So far we have spoken as if the evidence supported a triplet code, but this was simply for illustration. Exactly the same results would be obtained if the code operated with groups of, say, 5 bases. Moreover, our symbols + and − must not be taken to mean literally the addition or subtraction of a single base.

It is easy to see that our symbolism is more exactly as follows:

+ represents $+m$, modulo n
− represents $-m$, modulo n

where n (a positive integer) is the coding ratio (that is, the number of bases which code one amino-acid) and m is any integral number of bases, positive or negative.

It can also be seen that our choice of reading direction is arbitrary, and that the same results (to a first approximation) would be obtained in whichever direction the genetic material was read, that is, whether the starting point is on the right or the left of the gene, as conventionally drawn.

Triple Mutants and the Coding Ratio

The somewhat abstract description given above is necessary for generality, but fortunately we have convincing evidence that the coding ratio is in fact 3 or a multiple of 3.

This we have obtained by constructing triple mutants of the form (+ with + with +) or (− with − with −). One must be careful not to make shifts

Table 3. Triple Mutants having a Wild or Pseudo-wild Phenotype

FC (0 + 40 + 38)
FC (0 + 40 + 58)
FC (0 + 40 + 57)
FC (0 + 40 + 54)
FC (0 + 40 + 55)
FC (1 + 21 + 23)

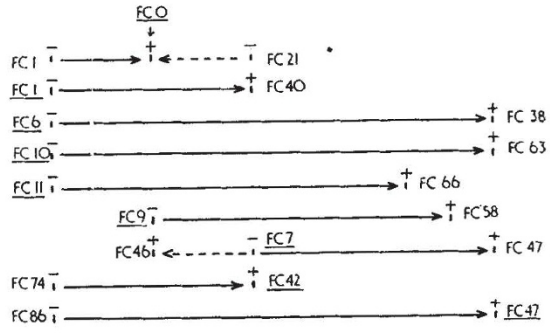


Fig. 4. A simplified version of the genetic map of Fig. 2. Each line corresponds to the suppressor from one mutant, here underlined. The arrows show the range over which suppressors have so far been found, the extreme mutants being named on the map. Arrows to the right are shown solid, arrows to the left dotted

across the 'unacceptable' regions for the ← shifts, but these we can avoid by a proper choice of mutants.

We have so far examined the six cases listed in Table 3 and in all cases the triples are wild or pseudo-wild.

The rather striking nature of this result can be seen by considering one of them, for example, the triple (*FC* 0 with *FC* 40 with *FC* 38). These three mutants are, by themselves, all of like type (+). We can say this not merely from the way in which they were obtained, but because each of them, when combined with our mutant *FC* 9 (—), gives the wild, or pseudo-wild phenotype. However, either singly or together in pairs they have an *r* phenotype, and will not grow on *K*. That is, the function of the gene is absent. Nevertheless, the combination of all three in the same gene partly restores the function and produces a pseudo-wild phage which grows on *K*.

This is exactly what one would expect, in favourable cases, if the coding ratio were 3 or a multiple of 3.

Our ability to find the coding ratio thus depends on the fact that, in at least one of our composite mutants which are 'wild', at least one amino-acid must have been added to or deleted from the polypeptide chain without disturbing the function of the gene-product too greatly.

This is a very fortunate situation. The fact that we can make these changes and can study so large a region probably comes about because this part of the protein is not essential for its function. That this is so has already been suggested by Champe and Benzer[13] in their work on complementation in the r_{II} region. By a special test (combined infection on *K*, followed by plating on *B*) it is possible to examine the function of the *A* cistron and the *B* cistron separately. A particular deletion, 1589 (see Fig. 5) covers the right-hand end of the *A* cistron and part of the left-hand end of the *B* cistron. Although 1589 abolishes the *A* function, they showed that it allows the *B* function to be expressed to a considerable extent. The region of the *B* cistron deleted by 1589 is that into which all our *FC* mutants fall.

Joining two Genes Together

We have used this deletion to re-inforce our idea that the sequence is read in groups from a fixed starting point. Normally, an alteration confined to the *A* cistron (be it a deletion, an acridine mutant, or any other mutant) does not prevent the expression of the *B* cistron. Conversely, no alteration within the *B* cistron prevents the function of the *A* cistron. This implies that there may be a region between the

two cistrons which separates them and allows their functions to be expressed individually.

We argued that the deletion 1589 will have lost this separating region and that therefore the two (partly damaged) cistrons should have been joined together. Experiments show this to be the case, for now an alteration to the left-hand end of the *A* cistron, if combined with deletion 1589, can prevent the *B* function from appearing. This is shown in Fig. 5. Either the mutant *P*43 or *X*142 (both of which revert strongly with acridines) will prevent the *B* function when the two cistrons are joined, although both of these mutants are in the *A* cistron. This is also true of *X*142 *S*1, a suppressor of *X*142 (Fig. 5, case *b*). However, the double mutant (*X*142 with *X*142 *S*1), of the type (+ with −), which by itself is pseudo-wild, still has the *B* function when combined with 1589 (Fig. 5, case *c*). We have also tested in this way the 10 deletions listed by Benzer[7], which fall wholely to the left of 1589. Of these, three (386, 168 and 221) prevent the *B* function (Fig. 5, case *f*), whereas the other seven show it (Fig. 5, case *e*). We surmise that each of these seven has lost a number of bases which is a multiple of 3. There are theoretical reasons for expecting that deletions may not be random in length, but will more often have lost a number of bases equal to an integral multiple of the coding ratio.

It would not surprise us if it were eventually shown that deletion 1589 produces a protein which consists of part of the protein from the *A* cistron and part of that from the *B* cistron, joined together in the same polypeptide chain, and having to some extent the function of the undamaged *B* protein.

Is the Coding Ratio 3 or 6 ?

It remains to show that the coding ratio is probably 3, rather than a multiple of 3. Previous rather rough extimates[10,14] of the coding ratio (which are admittedly very unreliable) might suggest that the coding ratio is not far from 6. This would imply, on our theory, that the alteration in *FC* 0 was not to one base, but to two bases (or, more correctly, to an even number of bases).

We have some additional evidence which suggests that this is unlikely. First, in our set of 126 mutants produced by acridine yellow (referred to earlier) we have four independent mutants which fall at or

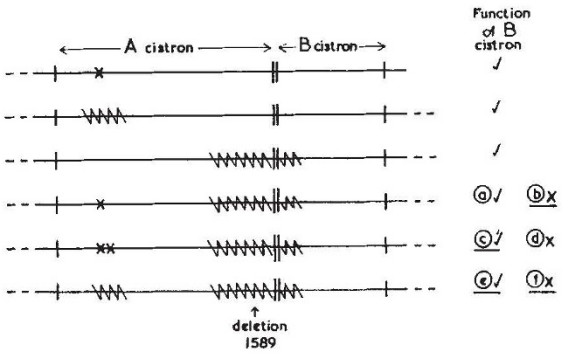


Fig. 5. Summary of the results with deletion 1589. The first two lines show that without 1589 a mutation or a deletion in the *A* cistron does not prevent the *B* cistron from functioning. Deletion 1589 (line 3) also allows the *B* cistron to function. The other cases, in some of which an alteration in the *A* cistron prevents the function of the *B* cistron (when 1589 is also present), are discussed in the text. They have been labelled (*a*), (*b*), etc., for convenience of reference, although cases (*a*) and (*d*) are not discussed in this paper. ✓ implies function; × implies no function

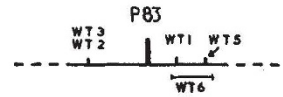


Fig. 6. Genetic map of *P* 83 and its suppressors, *WT* 1, etc. The region falls within segment *B*9*a* near the right-hand end of the *B* cistron. It is not yet known which way round the map is in relation to the other figures

close to the *FC* 9 site. By a suitable choice of partners, we have been able to show that two are + and two are −. Secondly, we have two mutants (*X*146 and *X*225), produced by hydrazine[15], which fall on or near the site *FC* 30. These we have been able to show are both of type −.

Thus unless both acridines and hydrazine usually delete (or add) an even number of bases, this evidence supports a coding ratio of 3. However, as the action of these mutagens is not understood in detail, we cannot be certain that the coding ratio is not 6, although 3 seems more likely.

We have preliminary results which show that other acridine mutants often revert by means of close suppressors, but it is too sketchy to report here. A tentative map of some suppressors of *P* 83, a mutant at the other end of the *B* cistron, in segment *B* 9*a*, is shown in Fig. 6. They occur within a shorter region than the suppressors of *FC* 0, covering a distance of about one-twentieth of the *B* cistron. The double mutant *WT* (2 + 5) has the *r* phenotype, as expected.

Is the Code Degenerate?

If the code is a triplet code, there are 64 (4 × 4 × 4) possible triplets. Our results suggest that it is unlikely that only 20 of these represent the 20 amino-acids and that the remaining 44 are nonsense. If this were the case, the region over which suppressors of the *FC* 0 family occur (perhaps a quarter of the *B* cistron) should be very much smaller than we observe, since a shift of frame should then, by chance, produce a nonsense reading at a much closer distance. This argument depends on the size of the protein which we have assumed the *B* cistron to produce. We do not know this, but the length of the cistron suggests that the protein may contain about 200 amino-acids. Thus the code is probably 'degenerate', that is, in general more than one triplet codes for each amino-acid. It is well known that if this were so, one could also account for the major dilemma of the coding problem, namely, that while the base composition of the DNA can be very different in different micro-organisms, the amino-acid composition of their proteins only changes by a moderate amount[16]. However, exactly how many triplets code amino-acids and how many have other functions we are unable to say.

Future Developments

Our theory leads to one very clear prediction. Suppose one could examine the amino-acid sequence of the 'pseudo-wild' protein produced by one of our double mutants of the (+ with −) type. Conventional theory suggests that since the gene is only altered in two places, only two amino-acids would be changed. Our theory, on the other hand, predicts that a string of amino-acids would be altered, covering the region of the polypeptide chain corresponding to the region on the gene between the two mutants. A good protein on which to test this hypothesis is

the lysozyme of the phage, at present being studied chemically by Dreyer[17] and genetically by Streisinger[10].

At the recent Biochemical Congress at Moscow, the audience of Symposium I was startled by the announcement of Nirenberg that he and Matthaei[18] had produced polyphenylalanine (that is, a polypeptide all the residues of which are phenylalanine) by adding polyuridylic acid (that is, an RNA the bases of which are all uracil) to a cell-free system which can synthesize protein. This implies that a sequence of uracils codes for phenylalanine, and our work suggests that it is probably a triplet of uracils.

It is possible by various devices, either chemical or enzymatic, to synthesize polyribonucleotides with defined or partly defined sequences. If these, too, will produce specific polypeptides, the coding problem is wide open for experimental attack, and in fact many laboratories, including our own, are already working on the problem. If the coding ratio is indeed 3, as our results suggest, and if the code is the same throughout Nature, then the genetic code may well be solved within a year.

We thank Dr. Alice Orgel for certain mutants and for the use of data from her thesis, Dr. Leslie Orgel for many useful discussions, and Dr. Seymour Benzer for supplying us with certain deletions. We are particularly grateful to Prof. C. F. A. Pantin for allowing us to use a room in the Zoological Museum, Cambridge, in which the bulk of this work was done.

[1] Wittman, H. G., Symp. 1, Fifth Intern. Cong. Biochem., 1961, for refs. (in the press).
[2] Tsugita, A., and Fraenkel-Conrat, H., *Proc. U.S. Nat. Acad. Sci.*, **46**, 636 (1960); *J. Mol. Biol.* (in the press).
[3] Brenner, S., *Proc. U.S. Nat. Acad. Sci.*, **43**, 687 (1957).
[4] For refs. see Watson, H. C., and Kendrew, J. C., *Nature*, **190**, 670 (1961).
[5] Crick, F. H. C., Griffith, J. S., and Orgel, L. E., *Proc. U.S. Nat. Acad. Sci.*, **43**, 416 (1957).
[6] Benzer, S., *Proc. U.S. Nat. Acad. Sci.*, **45**, 1607 (1959), for refs. to earlier papers.
[7] Benzer, S., *Proc. U.S. Nat. Acad. Sci.*, **47**, 403 (1961); see his Fig. 3.
[8] Brenner, S., Benzer, S., and Barnett, L., *Nature*, **182**, 983 (1958).
[9] Brenner, S., Barnett, L., Crick, F. H. C., and Orgel, A., *J. Mol. Biol.*, **3**, 121 (1961).
[10] Streisinger, G. (personal communication and in the press).
[11] Feynman, R. P.; Benzer, S.; Freese, E. (all personal communications).
[12] Jinks, J. L., *Heredity*, **16**, 153, 241 (1961).
[13] Champe, S., and Benzer, S. (personal communication and in preparation).
[14] Jacob, F., and Wollman, E. L., *Sexuality and the Genetics of Bacteria* (Academic Press, New York, 1961). Levinthal, C. (personal communication).
[15] Orgel, A., and Brenner, S. (in preparation).
[16] Sueoka, N. *Cold Spring Harb. Symp. Quant. Biol.* (in the press).
[17] Dreyer, W. J., Symp. 1, Fifth Intern. Cong. Biochem., 1961 (in the press).
[18] Nirenberg, M. W., and Matthaei, J. H., *Proc. U.S. Nat. Acad. Sci.*, **47**, 1588 (1961).

SCIENCE AND WORLD AFFAIRS

THE Seventh Pugwash Conference on Science and World Affairs was held at Stowe, Vermont, during September 5–9. Forty-one scientists from twelve countries attended the Conference*.

This Conference had as its theme "International Co-operation in Pure and Applied Science". The previous conferences have been chiefly concerned with ways of preventing the misuse of science in the wholesale destruction of mankind. In this Conference at Stowe, constructive international co-operation in science was discussed, because it is a way to create trust between nations, a trust which develops from common interests and from experience in working together.

Science misused by nations to foster their competitive interests as world powers makes possible the destruction of mankind. Science used co-operatively by all nations for the increase of human knowledge and the improvement of man's productive capacity can give all men on Earth a satisfactory and worthwhile life. Scientists bear a responsibility both to foster the constructive use of science and to help in preventing its destructive use.

The deliberations of the Conference were carried out in plenary sessions and in meetings of working groups. Similar suggestions for co-operative research activities arose independently from different working groups, and this is reflected in several places in the statement. This is a welcome indication of the essential unity in science. The discussions were carried on in a spirit of friendly co-operation, and full agreement was reached by the entire Conference on the suggestions enumerated here.

* Sir John Crawford (Australia); Prof. Hans Thirring (Austria); Prof. C. Pavan (Brazil); Prof. G. Nadjakov (Bulgaria); Prof. G. Burkhardt (Federal Republic of Germany); Sir Edward Bullard, Prof. A. Haddow, Sir Ben Lockspeiser, Prof. J. Rotblat (Great Britain); Prof. F. B. Straub (Hungary); Dr. G. Bernardini (Italy); Prof. T. Toyoda (Japan); Prof. B. V. A. Röling (Netherlands); Academician A. A. Blagonravov, Academician N. N. Bogolubov, Academician M. M. Dubinin, Prof. V. M. Khvostov, Academician N. M. Sissakian, Prof. N. A. Talensky, Academician I. E. Tamm, Academician A. V. Topchiev (U.S.S.R.); Prof. Harrison Brown, Dr. William Consolazio, Prof. Paul Doty, Prof. Bentley Glass, Prof. C. O'D. Iselin, Dr. Martin Kaplan, Prof. Chauncey Leake, Prof. Linus Pauling, Prof. Jay Orear, Prof. W. Pickering, Mr. Gerard Piel, Prof. I. Rabi, Prof. Eugene Rabinowitch, Dr. Roger Revelle, Prof. Alexander Rich, Prof. Walter Rosenblith, Dr. Eugene Staley, Dr. Alvin Weinberg, Prof. Eugene Wigner, Prof. J. R. Zacharias (United States).

(I) Co-operation in the Earth Sciences

The planet Earth is the common abode of all humans. They have a common interest, both intellectual and practical, in increasing the knowledge of the structure and dynamics of the Earth.

The following proposals were made by the Conference.

(*A*) *A survey of the entire ocean in three dimensions.* (1) *The ocean floor.* An international programme was proposed to develop a detailed map of the floor of the world ocean, including sub-bottom reflecting layers.

(2) *Waters of the ocean.* An international programme should be devised to survey and map the three-dimensional distribution of temperatures, salinity, density, dissolved oxygen, and nutrient salts, under average conditions, of the ocean and synoptic surveys to develop the broad picture of seasonal and shorter-period changes in more limited areas, as well as the study of the interactions among the major bodies of water in the ocean.

(3) *Ocean life.* An international survey and mapping showing the major biological provinces of the ocean and determination of the fertility of the waters at all levels in the food chain and the standing crop of food materials available for human use should be undertaken.

(*B*) *Earth's crust and mantle. Deep drilling programme.* The objective of drilling through the

[1] Biesele, J. J., *Mitotic Poisons and the Cancer Problem* (Amsterdam: Elsevier, 1958).
[2] Zamenhof, S., R. de Giovanni, and S. Greer, *Nature*, **181**, 827 (1958).
[3] Dunn, D. B., and J. D. Smith, *Nature*, **175**, 336 (1955).
[4] Shapiro, H. S., and E. Chargaff, *Nature*, **188**, 62 (1960).
[5] Litman, R. M., and A. B. Pardee, *Nature*, **178**, 529 (1956).
[6] Djordjevic, B., and W. Szybalski, *J. Exp. Med.*, **112**, 509 (1960).
[7] Kit, S., C. Beck, O. L. Graham, and A. Gross, *Cancer Res.*, **18**, 598 (1958).
[8] Hsu, T. C., D. Billen, and A. Levan, unpublished data.
[9] Hsu, T. C., and D. S. Kellogg, *Genetics and Cancer*, (University of Texas Press, 1959), p. 183.
[10] Hsu, T. C., *Univ. Texas Publ.*, **5914**, 129 (1959).
[11] Humphrey, R. M., unpublished data.
[12] Yerganian, G., and M. J. Leonard, *Science* (in press).
[13] McCoy, T. A., M. Maxwell, and P. F. Kruse, *Proc. Soc. Exp. Biol. & Med.*, **100**, 115 (1959).
[14] Hsu, T. C., and D. S. Kellogg, *J. Nat. Cancer Inst.*, **25**, 221 (1960).
[15] Hsu, T. C., and D. S. Kellogg, *J. Nat. Cancer Inst.*, **24**, 1067 (1960).
[16] Kit, S., unpublished data.
[17] Dewey, W. C., unpublished data.
[18] Szybalski, W., personal communication.
[19] Freese, E., *Brookhaven Symp. in Biol.*, **12**, 63 (1959).
[20] Schwartz, D., *Genetics*, **45**, 1141 (1960).
[21] Yerganian, G., and C. A. Livingston, *Proc. Am. Ass'n Cancer Res.*, **2**, 159 (1956).
[22] Livingston, C. A., and G. Yerganian, *Genetics*, **41**, 652 (1956).
[23] Szybalski, W., discussion of paper by Freese, *Brookhaven Symp. in Biol.*, **12**, 75 (1959).
[24] Zamenhof, S., and G. Griboff, *Nature*, **174**, 306 (1954).

ON THE TOPOGRAPHY OF THE GENETIC FINE STRUCTURE

BY SEYMOUR BENZER

DEPARTMENT OF BIOLOGICAL SCIENCES, PURDUE UNIVERSITY

*Read before the Academy, April 27, 1960**

In an earlier paper,[1] a detailed examination was made of the structure of a small portion of the genetic map of phage T4, the *r*II region. This region, which controls the ability of the phage to grow in *Escherichia coli* strain K, consists of two adjacent cistrons, or functional units. Various *r*II mutants, unable to grow in strain K, have mutations affecting various parts of either or both of these cistrons. The topology of the region; i.e., the manner in which its parts are interconnected, was intensively tested and it was found that the active structure can be described as a string of subelements, a mutation constituting an alteration of a point or segment of the linear array.

This paper is a sequel in which inquiry is made into the topography of the structure, i.e., local differences in the properties of its parts. Specifically, are all the subelements equally mutable? If so, mutations should occur at random throughout the structure and the topography would be trivial. On the other hand, sites or regions of unusually high or low mutability would be interesting topographic features.

The preceding investigation of topology was done by choosing mutants showing no detectable tendency to revert. This avoided any possible confusion between recombination and reverse mutation, so that a qualitative (yes-or-no) test for re-

combination was possible. The class of non-reverting mutants automatically included those marked by relatively large alterations, which will be referred to as "deletions." Such a mutant is defined for the present purposes as one which inter-

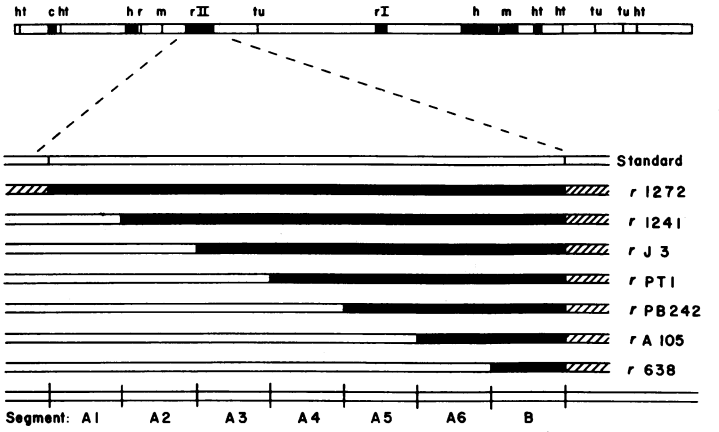


FIG. 1*a*.—At the top, the *r*II region is shown compared with the entire genetic map of the phage. This map is a composite[15] of markers mapped in T4 and the related phage T2. Seven segments of the *r*II region are defined by a set of "deletions" beginning at different points and extending to the right-hand end (and possibly beyond, as indicated by shading).

sects (fails to give recombination with) two or more mutants that do recombine with each other. Deletions provided overlaps of the sort needed to test the topology and to divide the map into segments.

The present investigation of topography, however, is concerned with differentiation of the various points in the structure. For this purpose mutants which do revert are of the greater interest, since they are most likely to contain small alterations. As a rule (there are exceptions) an *r*II mutant that reverts behaves as if its alteration were localized to a point. That is to say, mutants that intersect with the same mutant also intersect with each other. In a cross, recombination can be scored only if it is clearly detectable above the spontaneous reversion noise of the mutants involved. Therefore, the precision with which a mutation can be mapped is limited by its reversion rate. The detailed analysis of topography can best be done with mutants having low, non-zero reversion rates.

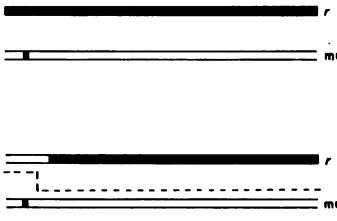


FIG. 1*b*.—Mapping a mutation by use of the reference deletions. If mutant x has a mutation in segment 1, it is overlapped by *r*1272, but not by *r*1241. Therefore, standard-type recombinants (as indicated by the dotted line) can only arise when x is crossed with *r*1241.

Some thousands of such *r*II mutants, both spontaneous and induced, have been analyzed and the resultant topographic map is presented here.

Assignment of Mutations to Segments.—To test thousands of mutants against one another for recombination in all possible pairs would require millions of crosses. This task may be greatly reduced by making use of deletions. Each mutant is first tested against a few key deletions. The recombination test gives a negative result

if a deletion overlaps the mutation in question and a positive result if it does not overlap. These results quickly locate a mutation within a particular segment of the map. It is then necessary to test against each other only the group of mutants having mutations within each segment, so that the number of tests needed is much smaller. In addition, if the order of the segments is known, the entire set of point mutations becomes ordered to a high degree, making use of only qualitative tests.

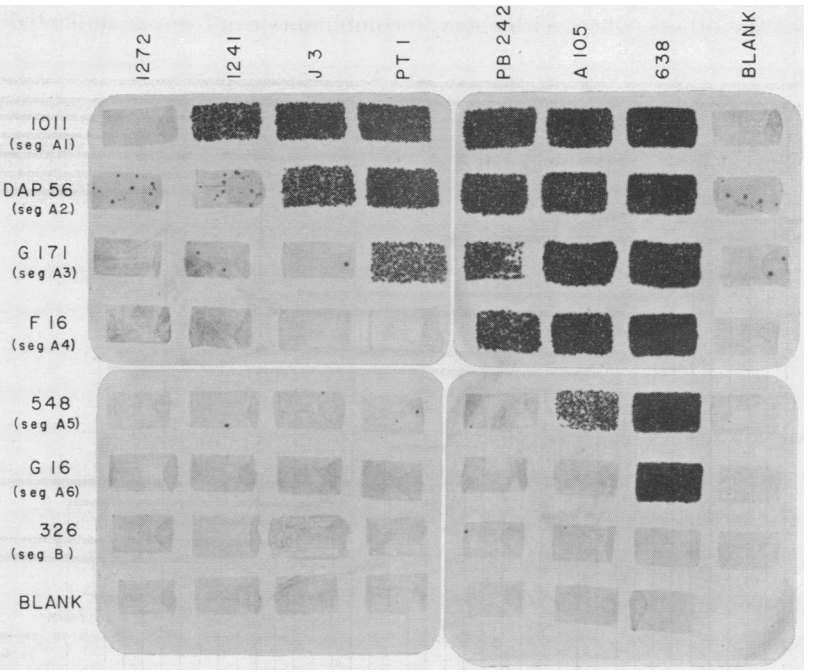


FIG. 2.—Crosses for mapping *r*II mutations. The photograph is a composite of four plates. Each row shows a given mutant tested against the reference deletions of Figure 1*a*. Plaques appearing in the blanks are due to revertants present in the mutant stock. The results show each of these mutations to be located in a different segment.

Procedure for crosses—The broth medium is 1% Difco bacto-tryptone plus 0.5% NaCl. For plating, broth is solidified with 1.2% agar for the bottom layer and 0.7% for the top layer. Stocks are grown in broth using *E. coli* BB which does not discriminate between *r*II mutants and the standard type. To cross two mutants, one drop of each at a titer of about 10^9 phage particles/ml is placed in a tube and cells of *E. coli* B are added (roughly 0.5 ml of a 1-hour broth culture containing about 2×10^8 cells/ml). The *r*II mutants are all able to grow on strain B and have an opportunity to undergo genetic recombination. After allowing a few minutes for adsorption, a droplet of the mixture is spotted (using a sterile paper strip) on a plate previously seeded with *E. coli* K. If the mutants recombine to produce standard type progeny, plaques appear on K. A negative result signifies that the proportion of recombinants is less than about 10^{-3}% of the progeny.

Within any one segment, however, the order of the various sites remains undetermined. This order can still be determined, if desired, by quantitative measurements of recombination frequencies.

In order to facilitate this project many more deletions have been mapped than were described in the previous paper. These suffice to carve up the structure into 47 distinct segments. By virtue of the proper overlaps, the order of almost all of

these segments is established. Observe first the seven large mutations in Figure 1*a*. These are of a kind which begin at a particular point and extend all the way to one end. Thus, they serve to divide the structure into the seven major segments shown.

Consider a small mutation located in the segment Al, as indicated in Figure 1*b*. It is overlapped by *r*1272 and therefore when crossed with it cannot give rise to standard type recombinants. It will, however, give a positive result with *r*1241 or any of the others, since, with them, recombinants can form as indicated by the

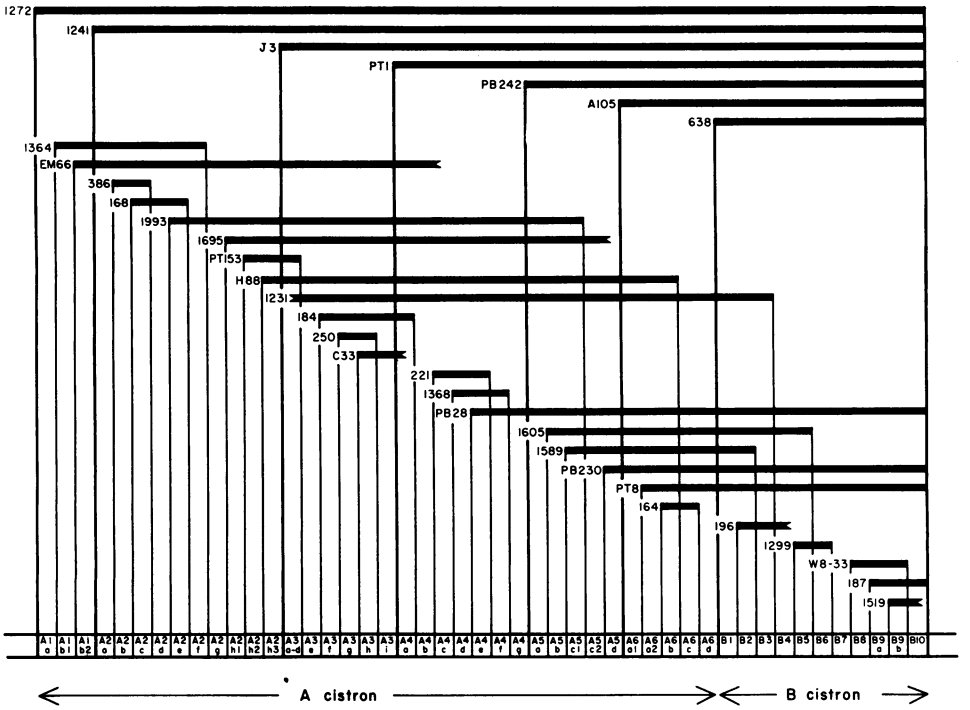


FIG. 3.—Deletions used to divide the main segments of Figure 1 into 47 smaller segments. (Some ends have not been used to define a segment, and are drawn fluted.) The A and B cistrons, which are defined by an independent functional test, coincide with the indicated portions of the recombination map. Most of the mutants are of spontaneous origin. Possible exceptions are EM66, which was found in a stock treated with ethyl methane sulfonate, and the PT and PB mutants, which were obtained from stocks treated with heat at low pH. The PT mutants were contributed by Dr. E. Freese.

dotted line. A point mutation located in the second segment will give zero with mutants *r*1272 and *r*1241 but not with the rest, and so on. Thus, if any point mutant is tested against the set of seven reference mutants in order, the segment in which its mutation belongs is established simply by counting the number of zeros. Figure 2 shows photographs of the test plates for seven mutants, each having its mutation located in a different segment.

Only these seven patterns, with an uninterrupted row of zeros beginning from the left, have ever been observed for thousands of mutants tested against these seven deletions. The complete exclusion of the other 121 possible patterns confirms the linear order of the segments.

Now a given segment can be further subdivided by means of other mutations having suitable starting or ending points. Figure 3 shows the set used in this study and the designation of each segment. Each mutant is first tested against the seven which have been chosen to define main segments. Once the main segment is known,

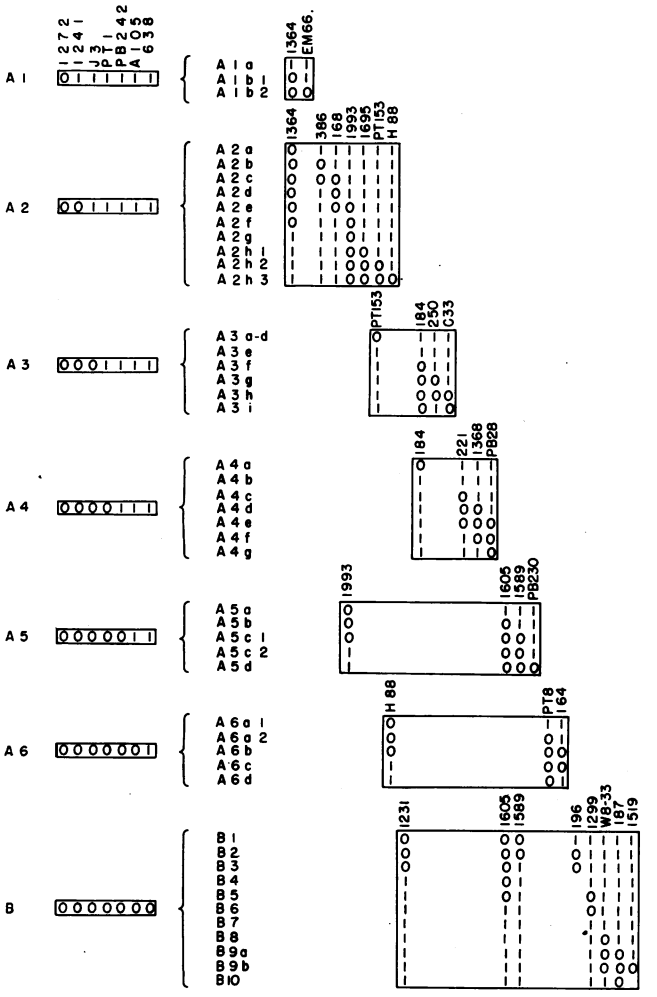


FIG. 4.—The test pattern which identifies the location of a point mutation in each of the segments of Figure 3. The test is done in two stages. An unknown mutant is first crossed with the "big seven" of Figure 1 in order. *Zero* signifies no detectable recombination and *one* signifies some, and the number of zeros defines the major segment. Once this is known, the mutant is crossed with the pertinent selected group of deletions to determine the small segment to which it belongs.

the mutant is tested against the appropriate secondary set. Figure 4 shows the pattern which identifies the location of a point mutation within each of the small segments. Thus, in two steps, a point mutation is mapped into one of the 47 segments.

The order of the first 42 segments, Ala through B6, is uniquely defined. Unfortunately, there remains a gap between $r1299$ and rW8-33. Therefore the order of segments B8 through B10, although fixed among themselves, could possibly be the reverse of that shown.[17] Also if there exists space to the right of segment B10, a mutation in that segment might map as if it were in segment B7, so that the latter segment must be tentatively regarded as a composite.

In the previous topology paper, the possibility that the structure contains branches was not eliminated. As pointed out by Delbrück, the existence of a branch would not lead to any contradiction with a linear topology if loss of a segment containing the branch point automatically led to loss of the entire branch.

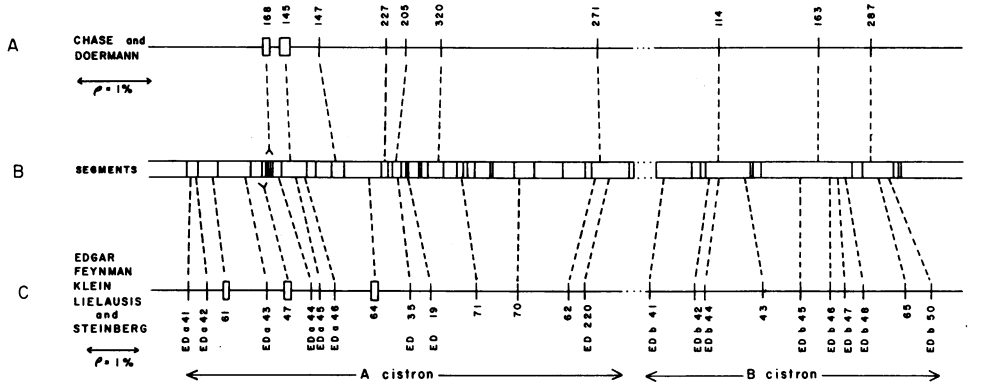


FIG. 5.—Correlation of the results of deletion mapping with the conventional method. *A:* The map constructed by Chase and Doermann[2] for ten rII mutants of phage T4B, using quantitative measurements of recombination frequency. The interval between adjacent mutations is drawn proportional to the frequency of recombination in a cross between the two. *C:* The map constructed in similar fashion by Edgar *et al.* (personal communication) with rII mutants of the very closely related phage T4D. The procedure used by Edgar *et al.* gives higher recombination frequencies. Therefore, the scales of the two maps are adjusted in the figure to produce a good over-all fit. Some of the mutations cover several sites and are drawn as having a corresponding length. A gap is left between the two cistrons because crosses between mutations in different cistrons give abnormally high frequencies due to the role of heterozygotes[16].

All of these mutations have also been mapped by the deletion method, and dotted lines indicate their locations in the various segments (*B*). The length of each segment is drawn in proportion to the number of distinct sites that have been found within it.

To show that a given segment is *not* a branch, it is required to find a mutation which penetrates it partially. From the mutations shown in Figure 3, it can be concluded that no branch exists that contains more than one of the 47 segments.

Comparison of Deletion Mapping by Recombination Frequencies.—The conventional method of genetic mapping makes use of recombination frequency as a measure of the distance between two mutations and requires careful quantitative measurements of the percentage of recombinant type progeny in each cross. By the method of overlapping deletions the order of mutations can be determined entirely by qualitative yes-or-no spot tests. Maps obtained independently by the two methods are compared in Figure 5. The upper part of the figure (*A*) shows the order obtained by Chase and Doermann[2] for a set of ten mutants, the distance between adjacent mutations being drawn proportional to the percentage of standard-type recombinants occurring among the progeny of a cross between the two. The central part of the figure (*B*) shows the rII region divided into the segments of

Figure 3, with the size of each segment drawn in proportion to the number of distinct sites which have been discovered within it (see below). As indicated by the dotted lines, there is perfect correlation in the order. In the lower part of the figure, a similar comparison is made for a set of *r*II mutations in the closely related phage strain T4D, which have been mapped, using recombination frequencies, by Edgar, Feynman, Klein, Lielausis, and Steinberg. Again the order agrees perfectly with that obtained by the use of deletions.

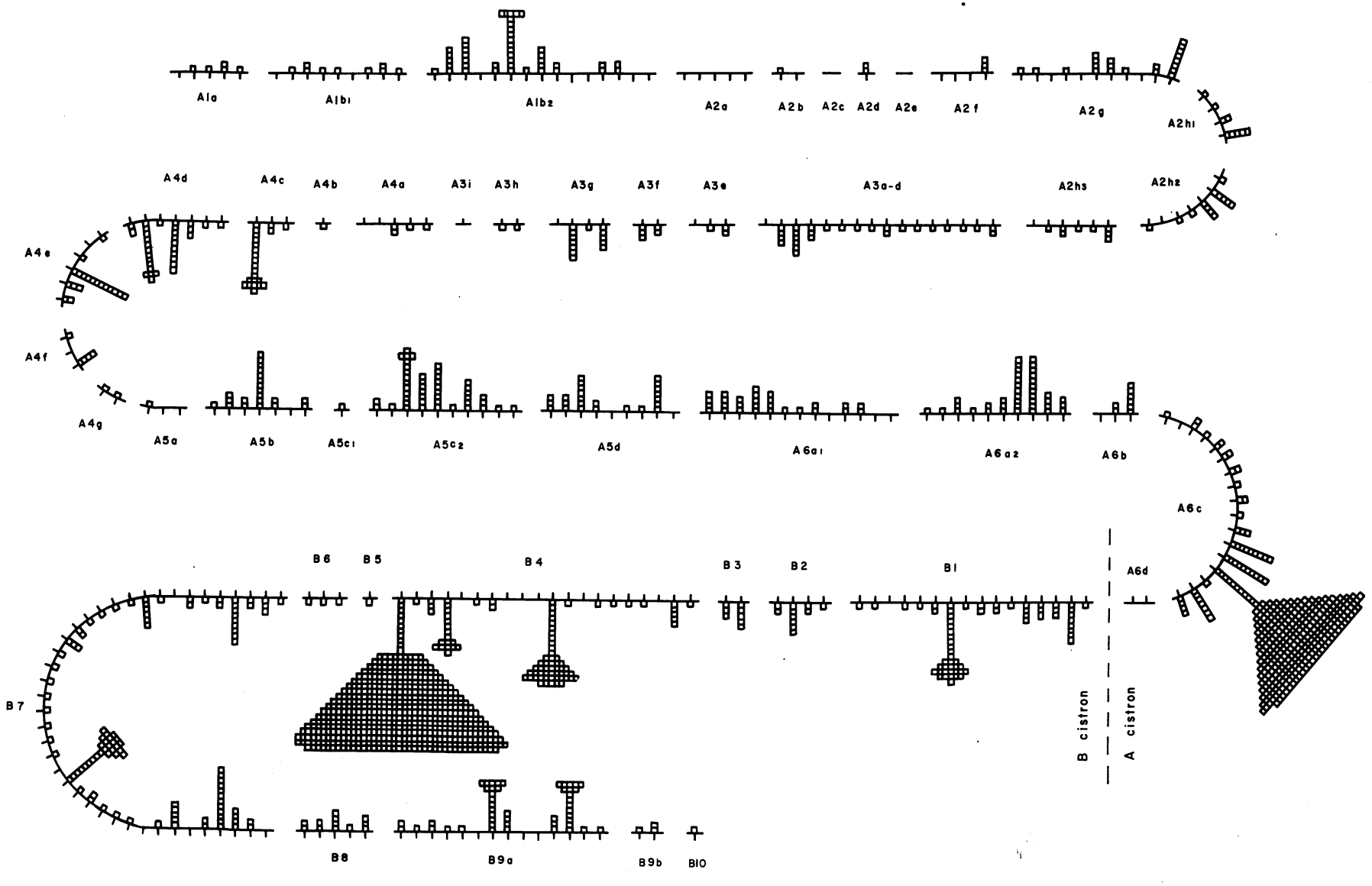
Topography for Spontaneous Mutations.—We now proceed to map reverting mutants of T4B which have arisen independently and spontaneously. The procedure is exactly as in Figure 4: first localizing into main segments, then into smaller segments. Finally mutants of the same small segment are tested against each other. Any which show recombination are said to define different sites. If two or more reverting mutants are found to show no detectable recombination with each other, they are considered to be repeats and one of them is chosen to represent the site in further tests. A set of distinct sites is thus obtained, each with its own group of repeats.

This procedure is based on the assumption that revertibility implies a point mutation. While this is a good working rule for *r*II mutants, a few exceptions have been found which appear to revert (i.e., give rise to some progeny which can produce detectable plaques on stain K) yet fail to give recombination with two or more mutants that do recombine with each other. If a mutant chosen to represent a "site" happens to be of this kind, mutations it overlaps will appear to be at the same site. Therefore, a group of "repeats" remains subject to splitting into different groups when they are tested against each other. This has not yet been done for all of the sites described here. It is, of course, in the nature of the recombination test that it is meaningful to say that two mutations are at different sites, while the converse conclusion is always tentative.

Figure 6 shows the map obtained for spontaneous mutants, with each occurrence of a mutation at a site indicated by a square. Within each segment the sites are drawn in arbitrary order. Other known sites are also indicated even though no occurrences were observed among this set of spontaneous mutants.

That the distribution is non-random leaps to the eye. More than 500 mutations have been observed at the most prominent "hotspot," while, at the other extreme, there are many sites at which only a single occurrence, or none, has so far been found.

To decide whether a given number of recurrences is significantly greater than random, the data may be compared with the expectation from a Poisson distribution. Figure 7 shows a distribution calculated to fit the least hot of the observed spontaneous sites, i.e., those at which one or two mutations have occurred, on the assumption that these sites belong to a uniform class of sites of low mutability. Comparing the observations with this curve, it would seem that if a site has four occurrences, there is a two-thirds probability that it is truly hotter than the class of sites of low mutability. Those having five or more are almost certainly hot. It can be concluded that at least sixty sites belong in a more mutable class than the coolest spots. Whether the hot sites can be divided into smaller homogeneous groups, assuming a Poisson distribution within each class, is difficult to say. Each of the two hottest sites is obviously unique.


A1a
A1b1
A1b2
A2a
A2b
A2c
A2d
A2e
A2f
A2g
A2h1
A4d
A4c
A4b
A4a
A3i
A3h
A3g
A3f
A3e
A3a-d
A2h3
A2h2
A4e
A4f
A4g
A5a
A5b
A5c1
A5c2
A5d
A6a1
A6a2
A6b
A6c
B6
B5
B4
B3
B2
B1
A6d
B7
B8
B9a
B9b
B10
B cistron
A cistron

From the distribution it can be predicted that there must exist at least 129 spontaneous sites not observed in this set of mutants. This is a minimum estimate since it is calculated on the assumption that the 2-occurrence sites are no more mutable than the 1-occurrence sites. If this is not correct, the predicted number of

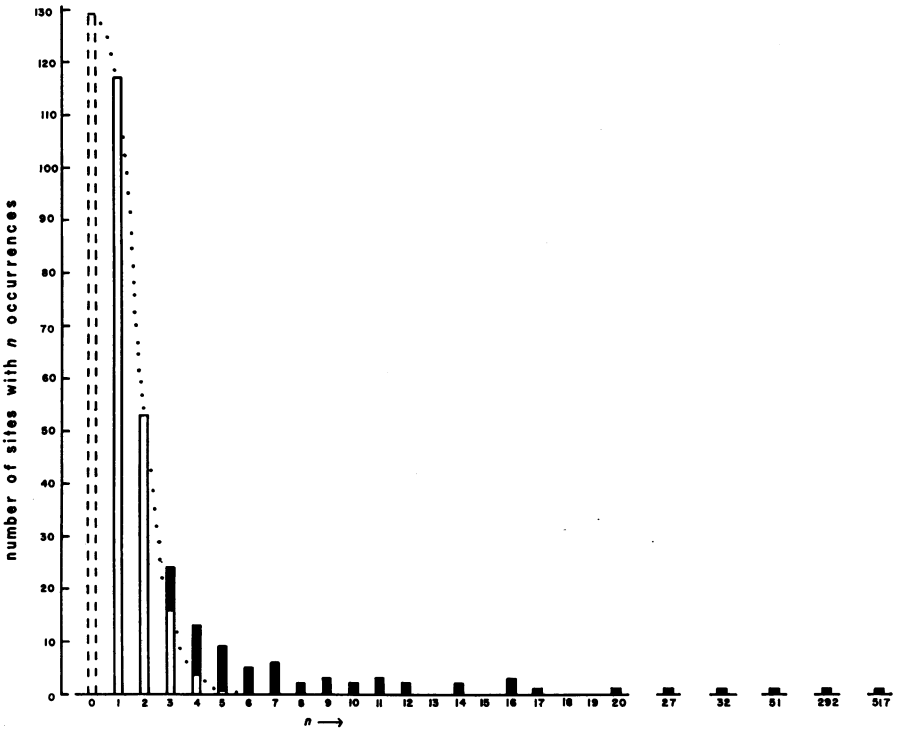


FIG. 7.—Distribution of occurrences of spontaneous mutations at various sites. The dotted line indicates a Poisson distribution fitted to the numbers of sites having one and two occurrences. This predicts a minimum estimate for the number of sites of comparable mutability that have zero occurrences due to chance (dashed column at n = 0). Solid bars indicate the minimum numbers of sites which have mutation rates significantly higher than the one- and two-occurrence class.

0-occurrence sites will be larger. Also, of course, there could exist a vast class of sites of much lower mutability. With 251 spontaneous sites identified and at least 129 more to be found, the degree of saturation of the map achieved with this set of 1,612 spontaneous mutants can be no greater than 66 per cent.

←

FIG. 6.—Topographic map of the *r*II region for spontaneous mutations. Each square represents one occurrence observed at the indicated site. Sites with no occurrences indicated are known to exist from induced mutations and from a few other selected spontaneous ones. The order of the segments is known for A1a through B7, but is only tentative for B7 through B10.[17] The arrangement of sites within each segment is arbitrary.

Each mutant arose independently in a plaque of either standard-type T4B or, in somewhat less than half of the cases, revertants of various *r*II mutants. All revertants (except F) gave results very similar to T4B. The pattern for *r*II mutants isolated from revertant F differs noticeably only in a reduced rate at the hotspot 117 (the site of its original *r*II mutation) and therefore does not significantly alter the topography. All the data for mutants isolated from standard type and from revertants are pooled in this figure.

Topography for Induced Mutations.—By the use of specific mutagens, new topographic features are revealed. This has been shown for *r*II mutants induced during reproduction of the phage inside the bacterial host cell with 5-bromouracil (Benzer and Freese[3]), proflavine (Brenner, Barnett, and Benzer[4]), and 2-aminopurine (Freese[5]). Other effective mutagens are 2,6-diaminopurine (Freese[5]) and 5-bromodeoxycytidine (Gregory, personal communication). Mutations may also be induced *in vitro*, i.e., in extracellular phase particles, by ethyl methane sulfonate (Loveless[6]) and nitrous acid (Vielmetter and Wieder;[7] Freese;[8] Tessman[9]).

*r*II mutants induced by all of these mutagens have now been mapped with respect to each other and spontaneous ones, and the results are given in Figure 8 (facing page 416) which shows the locations of over 2,400 induced and spontaneous mutations. Only *r*II mutants that have low reversion rates and are not too "leaky" on K have been included.

Each "spectrum" differs obviously from the spontaneous one. While the specificities of the various mutagens overlap in many respects, each differs significantly from the others at specific points. In making the comparison it must be borne in mind that the total number of mutants mapped is not the same for each mutagen and also that each induced set inevitably includes some proportion of spontaneous mutants. (An upper limit to this background can be set from the number of occurrences at the hottest spontaneous sites.) Also, none of the spectra are "saturated." Therefore, even if two mutagens act similarly upon a given site, it is possible, due to chance, that a few occurrences would be observed in one spectrum and not the other. Within these limitations, the map shows the comparative response at each site to each mutagen as well as the locations of various kinds of hotspots in various segments of the *r*II region.

The study of the induced mutations has added 53 new sites to the 251 identified by the spontaneous set alone, bringing the total to 304. (Four sites more are shown in Figure 8, but they come from a selected group of mutants outside this study.) Thus, a closer approach toward saturation of all the possible sites must have been made. By lumping together all the data, both spontaneous and induced, one can again make an estimate of the number of sites which must be detectable if one were to continue mapping mutants in the same proportion for the same mutagens. The result is that there must exist still a minimum of 120 sites not yet discovered. This appears discouragingly similar to the estimate based on spontaneous mutations alone. However, it need not be surprising if the use of mutagens brings into view some sites which have extremely low spontaneous mutability. With 308 sites identified and at least 120 yet to be found, the maximum degree of saturation of the map is 72 per cent.

Discussion.—One topographic feature, non-random mutability at the various sites, is obvious. Another question is whether mutable sites are distributed at random, or whether there exist portions of the map that are unusually crowded with or devoid of sites. The mapping technique used here defines only the order of sites from one segment to another (but not within a given segment). The distance between sites remains unspecified. However, all mutations in a segment more distal to a given point must be farther away than those in a more proximal segment. If the number of sites in a segment is used as a measure of its length, as in Figure 5, it can be seen that there is no major discrepancy between these distances and those

defined in terms of another measure of distance, recombination frequency. On a gross scale, therefore, there is no evidence for any large portion of the *r*II region that is unusually crowded or roomy with respect to sites. This does not necessarily mean that some other measure of distance would not reveal such regions, since it is at least conceivable that mutable sites coincide with points highly susceptible to recombination. The distribution of sites on a finer scale, within a small segment, remains to be investigated.

The number of points at which mutations can wreck the activity of a cistron is very large. This would be expected if a cistron dictates the formation of a poly peptide chain and "nonsense" mutations[10] are possible which interrupt the completion of the chain. Such mutations would be effective at any point of the structure, whereas ones which lead to "missense," i.e., the substitution of one amino acid for another, might be effective at relatively special points or regions which are crucial in affecting the active site or folding.

It would be of interest to compare the number of genetic sites to the material embodiment of the *r*II region in terms of nucleotides. Unfortunately, the size of the latter is not well known. Estimates based upon its length, in units of recombination frequency compared to the length of the entire genetic structure, are uncertain. A more direct attempt has been made using equilibrium sedimentation in a cesium chloride gradient and looking for a change in density of mutants known by genetic evidence to have portions of the *r*II region deleted (Nomura, Champe, and Benzer, unpublished). This technique has been successful in characterizing defective mutants of phage λ (Weigle, Meselson, and Paigen[11]) and is sufficiently sensitive to detect a decrease of 1 per cent in the amount of DNA per phage particle, but has so far failed with *r*II mutants. Although other explanations are possible, this result may suggest that the physical structure corresponding to the *r*II region represents less than 1 per cent of the total DNA of the phage particle, or less than 2,000 nucleotide pairs. If this is so, the number of possible sites would be of the order of at least one-fifth of the number of nucleotide pairs.

The data show that, if each site is characterized by its spontaneous mutability and response to various mutagens, the sites are of many different kinds. Some response patterns are represented only once in the entire structure. According to the Watson-Crick model[12] for DNA, the structure consists of only two types of elements, adenine-thymine (AT) pairs and guanine-hydroxymethylcytosine (GC) pairs. This does not mean, however, that there can only be two kinds of mutable sites, even if a site corresponds to a single base pair. Considering only base pair substitutions, a given AT pair can undergo three kinds of change: AT can be replaced by GC, CG, or TA. Certain of these changes may lead to a mutant phenotype, but some may not. The frequency of observable mutations at a particular AT pair will be determined by the sum of the probabilities for each type of change, each multiplied by a coefficient (either one or zero) according to whether that specific alteration at that particular pair does or does not represent a mutant type. Thus, if the probability that a base pair will be substituted is independent of its neighbors, the various AT sites may have seven different mutation rates. Similarly, there are seven rates possible for the various GC sites, so that it would be possible to account for fourteen classes by this mechanism. Some of these may have (total) spontaneous mutation rates that are similar. If a mutagen induces only certain substitu-

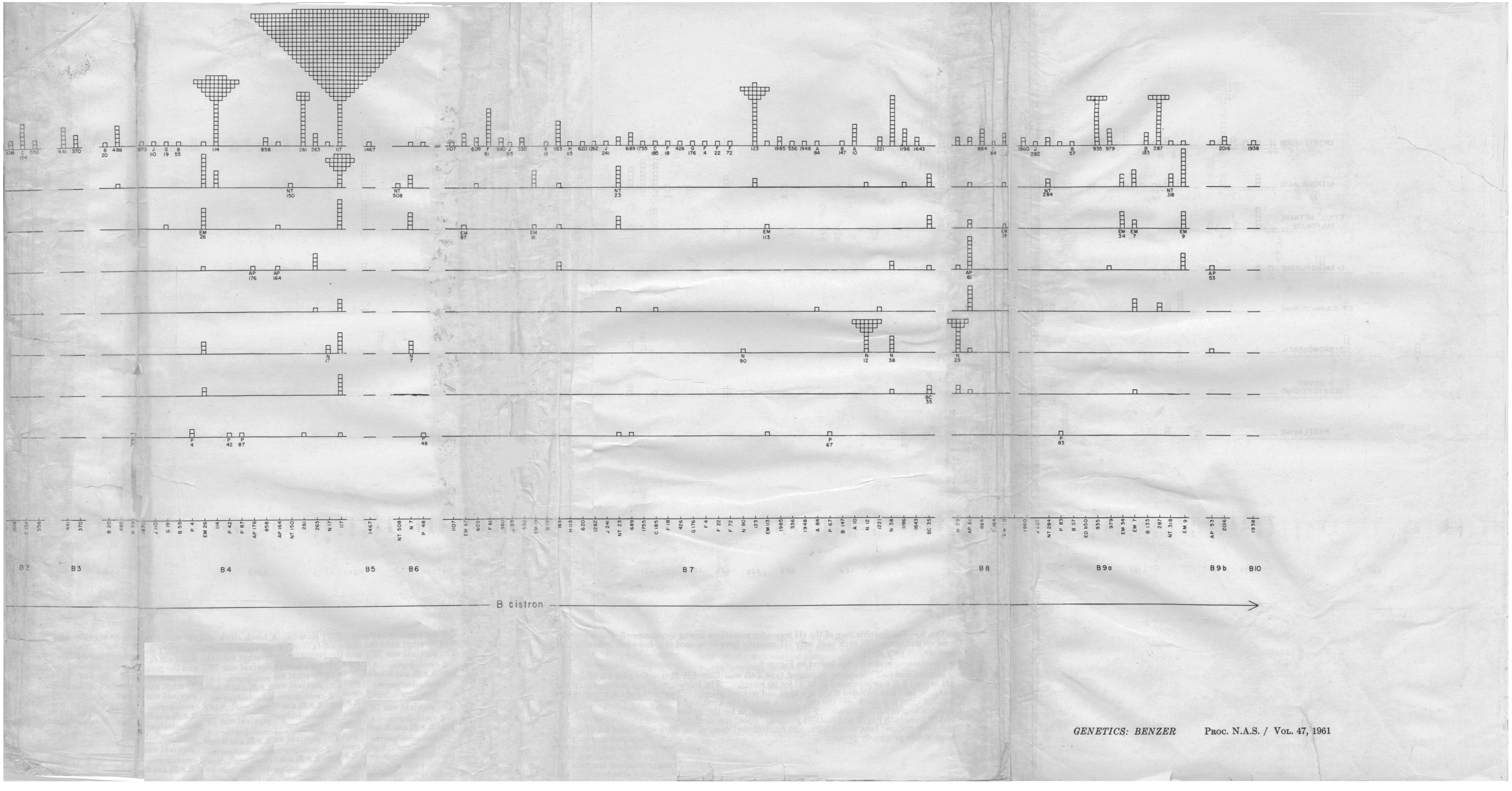
tions, it will facilitate further discriminations between sites but there should still be no more than fourteen classes.

If one allows for interactions between neighbors, the number of possible classes increases enormously. Such interactions are to be expected. As an example, consider the fact that AT pairs are held together much less strongly than are GC pairs.[13] If several AT pairs occur in succession, this segment of the DNA chain will be relatively loose, making it easier to consummate an illicit base pairing during replication. Thus, guanine and adenine, which make a very satisfactory pair of hydrogen bonds but require a larger than normal separation between the backbones, could be more readily accommodated. This would lead, in the next replication, to a replica in which one of the AT pairs has been substituted by a CG pair, with the orientation of purine and pyrimidine reversed. Thus, a region rich in AT pairs will tend to be more subject to substitution. If the same (standard-type) phenotype can be achieved by alternative sequences, the ones containing long stretches of AT pairs would tend to be lost because of their high mutability. In other words, cistrons ought to have evolved in such a way as to eliminate hotspots. The spontaneous hotspots that are observed would be remnants of an incomplete ironing-out process. In fact, a map of the *r*II region of the related phage T6 (Benzer, unpublished) also shows hotspots at locations corresponding to *r*131 and *r*117. However, while the first of these has a mutability similar to that in T4, the second is lower by a factor of four.

This point is emphasized by the data on reverse mutations. It is not uncommon for an *r*II mutant to have a reverse mutation rate that is greater than the total forward rate observed for the composite of at least 400 sites. That some of these high-rate reverse mutations represent true reversion (and not "suppressor" mutations) has been established in several cases by the most stringent criteria, including the demonstration that the revertant has exactly the same forward mutation rate at the same site as did the original standard type (Benzer, unpublished). It would therefore appear that certain kinds of highly mutable configurations are systematically excluded from the standard form of the *r*II genetic structure, and a mutation may recreate one of these banned sequences.

In the attempt to translate the genetic map into a nucleotide sequence, the detection of the various sites by forward mutation is necessarily the first step. By studies on the specificity of induction of reverse mutations,[14] one site at a time can be analyzed in the hope of identifying the specific bases involved.

Summary.—A small portion of the genetic map of phage T4, the two cistrons of the *rII* region, has been dissected by overlapping "deletions" into 47 segments. If any branch exists, it cannot be larger than one of these segments. The overlapping deletions are used to map point mutations and the map order established by this method is consistent with the order established by the conventional method that makes use of recombination frequencies. Further dissection has led to the identification of 308 distinct sites of widely varied spontaneous and induced mutability. The distributions throughout the region for spontaneous mutations and those induced by various chemical mutagens are compared. Data are included for nitrous acid and ethyl methane sulfonate acting *in vitro*, and 2-aminopurine, 2,6-diaminopurine, 5-bromouracil, 5-bromodeoxycytidine, and proflavine acting *in vivo*. The characteristic hotspots reveal a striking topography.


B2
B3
B4
B5
B6
B7
B8
B9a
B9b
B10
B cistron

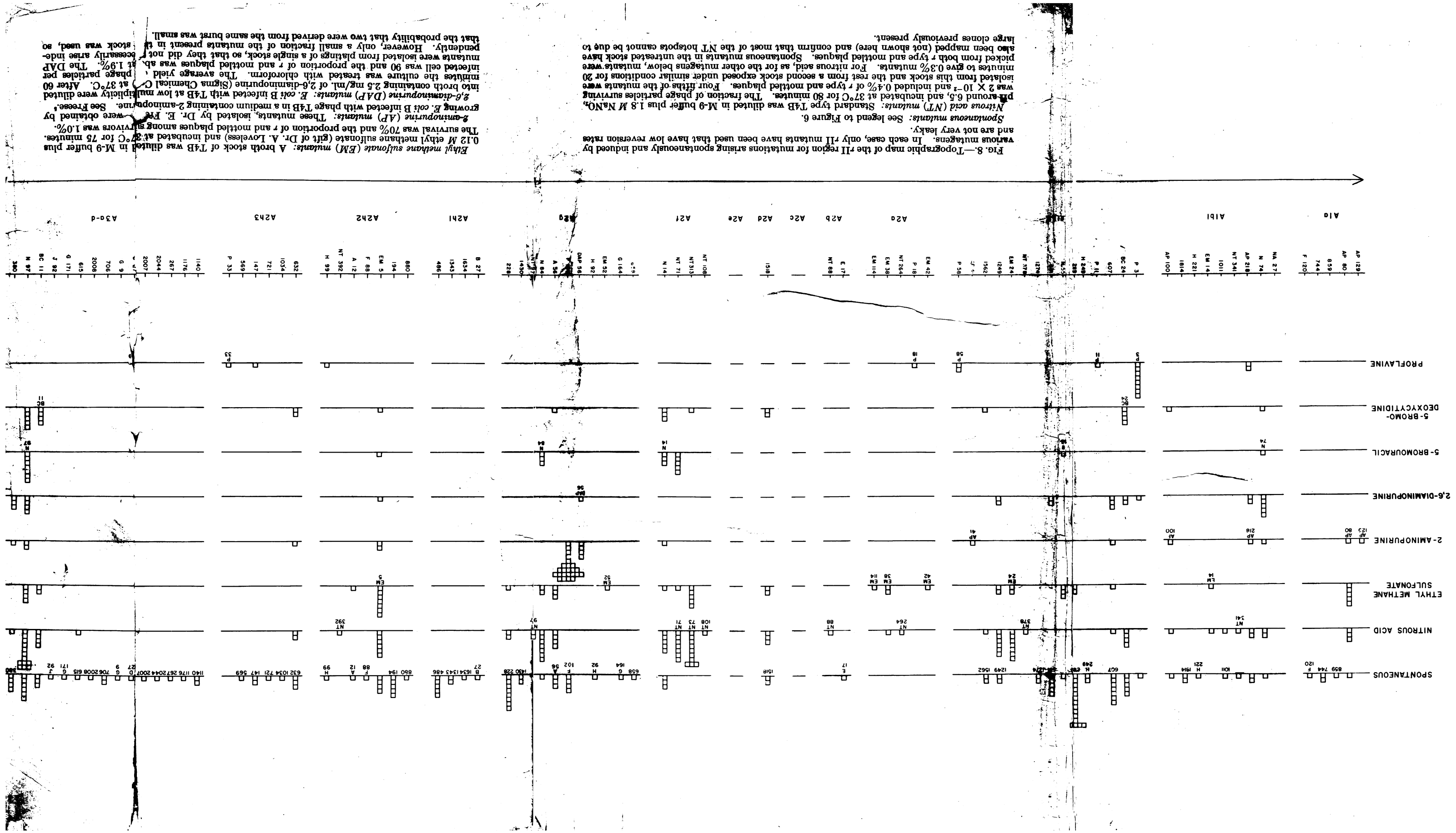


Fig. 8.—Topographic map of the *rII* region for mutations arising spontaneously and induced by various mutagens. In each case, only *rII* mutants have been used that have low reversion rates and are not very leaky.

Spontaneous mutants: See legend to Figure 6.

Nitrous acid (NT) mutants: Standard type T4B was diluted in M-9 buffer plus 1.8 *M* $NaNO_2$, pH around 6.5, and incubated at 37°C for 80 minutes. The fraction of phage particles surviving was 2×10^{-2} and included 0.4% of *r* type and mottled plaques. Four fifths of the mutants were isolated from this stock and the rest from a second stock exposed under similar conditions for 20 minutes to give 0.3% mutants. For nitrous acid, as for the other mutagens below, mutants were picked from both *r* type and mottled plaques. Spontaneous mutants in the untreated stock have also been mapped (not shown here) and confirm that most of the NT hotspots cannot be due to large clones previously present.

Ethyl methane sulfonate (EM) mutants: A broth stock of T4B was diluted in M-9 buffer plus 0.12 *M* ethyl methane sulfonate (gift of Dr. A. Loveless) and incubated at 37°C for 75 minutes. The survival was 70% and the proportion of *r* and mottled plaques among survivors was 1.0%.

2-aminopurine (AP) mutants: These mutants, isolated by Dr. E. Freese,[4] were obtained by growing *E. coli* B infected with phage T4B in a medium containing 2-aminopurine. See Freese.[4]

2,6-diaminopurine (DAP) mutants: *E. coli* B infected with T4B at low multiplicity were diluted into broth containing 2.5 mg/ml. of 2,6-diaminopurine (Sigma Chemical Co.) at 37°C. After 60 minutes the culture was treated with chloroform. The average yield of phage particles per infected cell was 90 and the proportion of *r* and mottled plaques was about 1.9%. The DAP mutants were isolated from plating of a single stock, so that they did not necessarily arise independently. However, only a small fraction of the mutants present in the stock was used, so that the probability that two were derived from the same burst was small.

It is a pleasure to thank Mrs. Karen Sue Supple, Mrs. Joan Reynolds, and Mrs. Lynne Bryant for their indefatigable assistance in mapping mutants and bookkeeping. I am indebted to Dr. Robert S. Edgar and his associates for permission to make use of their unpublished data in Figure 5 and to Dr. Ernst Freese for several deletions as well as mutants induced with 2-aminopurine and 5-bromodeoxyuridine. This research was supported by grants from the National Science Foundation and the National Institutes of Health.

* Given by invitation of the Committee on Arrangements for the Annual Meeting as part of a Symposium on Genetic Determination of Protein Structure, Robley C. Williams, Chairman.

[1] Benzer, S., these PROCEEDINGS, **45**, 1607 (1959).

[2] Chase, M., and A. H. Doermann, *Genetics*, **43**, 332 (1958).

[3] Benzer, S., and E. Freese, these PROCEEDINGS, **44**, 112 (1958).

[4] Brenner, S., L. Barnett, and S. Benzer, *Nature*, **182**, 983 (1958).

[5] Freese, E., *J. Molec. Biol.*, **1**, 87 (1959).

[6] Loveless, A., *Nature*, **181**, 1212 (1958).

[7] Vielmetter, W., and C. M. Wieder, *Z. Naturforsch*, **14b**, 312 (1959).

[8] Freese, E., *Brookhaven Symposia in Biol.*, **12**, 63 (1959).

[9] Tessman, I., *Virology*, **9**, 375 (1959).

[10] Crick, F. H. C., J. S. Griffith, and L. E. Orgel, these PROCEEDINGS, **43**, 416 (1957).

[11] Weigle, J., M. Meselson, and K. Paigen, *J. Molec. Biol.*, **1**, 379 (1959).

[12] Watson, J. D., and F. Crick, *Cold Spring Harbor Symposia Quant. Biol.*, **18**, 123 (1953).

[13] Doty, P., J. Marmur, and N. Sueoka, *Brookhaven Symposia in Biol.*, **12**, 1 (1959).

[14] Freese, E., these PROCEEDINGS, **45**, 622 (1959).

[15] Brenner, S., in *Advances in Virus Research* (New York: Academic Press, 1959), pp. 137–158.

[16] Edgar, R. S., *Genetics*, **43**, 235 (1958).

[17] The terms topology and topography are used here in the following senses (Webster's New Collegiate Dictionary, 1959)—*topology:* the doctrine of those properties of a figure unaffected by any deformation without tearing or joining; *topography:* the art or practice of graphic and exact delineation in minute detail, usually on maps or charts, of the physical features of any place or region.

[18] *Note added in proof.* Recent data have established that the orientation shown for segments B8 through B10 is the correct one.

Week 2

Mutations

J. Mol. Biol. (1961) **3**, 121–124

The Theory of Mutagenesis

In this preliminary note we wish to express our doubts about the detailed theory of mutagenesis put forward by Freese (1959*b*), and to suggest an alternative.

Freese (1959*b*) has produced evidence that shows that for the r_{II} locus of phage T4 there are two mutually exclusive classes of mutation and we have confirmed and extended his work (Orgel & Brenner, in manuscript). The technique used is to start with a standard wild type and make a series of mutants from it with a particular mutagen. Each mutant is then tested with various mutagens to see which of them will back-mutate it to wild type.

It is found that the mutations fall into two classes. The first, which we shall call the base analogue class, is typically produced by 5-bromodeoxyuridine (BD) and the second, which we shall call the acridine class, is typically produced by proflavin (PF). In general a mutant made with BD can be reverted by BD, and a mutant made with PF can be reverted by PF. A few of the PF mutants do not appear to revert with either mutagen, but the strong result is that no mutant has been found which reverts identically with both classes of mutagens, and that (with a few possible exceptions) mutants produced by one class cannot be reverted by the other.

Freese also showed that 2-aminopurine falls into the base analogue class, and that most (85%) spontaneous mutants at the r_{II} locus were not of the base analogue type. We have confirmed this and shown that they are in fact revertible by acridines. We have also shown that a number of other acridines, and in particular 5-aminoacridine, act like proflavin (Orgel & Brenner, in manuscript).

Freese has produced an ingenious explanation of these results, which should be consulted in the original for fuller details. In brief he postulated that the base analogue class of mutagens act by altering an A—T base-pair on the DNA (A = adenine, T = thymine) into a G—C pair, or *vice versa* (G = guanine, C = cytosine, or, in the T even phages, hydroxymethylcytosine). The fact that BD, which replaces thymine, could act both ways (from A—T to G—C or from G—C to A—T) was accounted for (Freese, 1959*a*) by assuming that in the latter case there was an error in pairing of the BD (such that it accidentally paired with guanine) while *entering* the DNA, and in the former case after it was already in the DNA.

Such alterations only change a purine into another purine, or a pyrimidine into another pyrimidine. Freese (1959*b*) has called these "transitions." He suggested that other conceivable changes, which he called "transversions" (such as, for example, from A—T to C—G) which change a purine into a pyrimidine and *vice versa*, occurred during mutagenesis by proflavin. This would neatly account for the two mutually exclusive classes of mutagens, since it is easy to see that a transition cannot be reversed by a transversion, and *vice versa*.

We have been led to doubt this explanation for the following reasons.

Our suspicions were first aroused by the curious fact that a comparison between the *sites* of mutation for one set of mutants made with BD and another set made with PF (Brenner, Benzer & Barnett, 1958) showed there were no sites in the r_{II} gene, among the samples studied, common to both groups.

Now this result alone need not be incompatible with Freese's theory of mutagenesis, since we have no good explanation for "hot spots" and this confuses quantitative argument. However it led us to the following hypothesis:

that acridines act as mutagens because they cause the insertion or the deletion of a base-pair.

This idea springs rather naturally from the views of Lerman (1960) and Luzzati (in preparation) that acridines are bound to DNA by sliding *between* adjacent base-pairs, thus forcing them 6·8 Å apart, rather than 3·4 Å. If this occasionally happened between the bases on *one* chain of the DNA, but not the other, during replication, it might easily lead to the addition or subtraction of a base.

Such a possible mechanism leads to a prediction. We know practically nothing about coding (Crick, 1959) but on most theories (except overlapping codes which are discredited because of criticism by Brenner (1957)) the deletion or the addition of a base-pair is likely to cause not the substitution of just one amino acid for another, but a much more substantial alteration, such as a break in the polypeptide chain, a considerable alteration of the amino acid sequence, or the production of no protein at all.

Thus one would not be surprised to find on these ideas that mutants produced by acridines were not capable of producing a slightly modified protein, but usually produced either no protein at all or a grossly altered one.

Somewhat to our surprise we find we already have data from two separate genes supporting this hypothesis.

(1) The *o* locus of phage T4 (resistance to osmotic shock) is believed to control a protein of the finished phage, possibly the head protein, because it shows phenotypic mixing (Brenner, unpublished). Using various base analogues we have produced mutants of this gene, though these map at only a small number of sites. We have failed on several occasions to produce any *o* mutants with proflavin. On another occasion two mutants were produced; one never reverted to wild type, while the other corresponded in position and spontaneous reversion rate to a base analogue site. We suspect therefore that these two mutants were not really produced by proflavin, but were the rarer sort of spontaneous mutant (Brenner & Barnett, unpublished).

(2) We have also studied mutation at the *h* locus in T2L, which controls a protein of the finished phage concerned with attachment to the host (Streisinger & Franklin, 1956).

Of the six different spontaneous h^+ mutants tested, all were easily induced to revert to *h* with 5-bromouracil (BU)†. This is especially significant when it is recalled that 85% of the spontaneous r_{II} mutants could not be reverted with base analogues (Freese, 1959*b*).

We have also shown (Brenner & Barnett, unpublished) that it is difficult to produce h^+ mutants from *h* by proflavin, though relatively easy with BU. The production of *r* mutants was used as a control.

It can be seen from Table 1 that if the production of h^+ mutants by BU and proflavin were similar to the production of *r* mutants we would expect to have obtained $\frac{57 \times 26}{108} = 13 h^+$ mutants with proflavin, whereas in fact we only found 1, and this may be spontaneous background.

† (Added in proof.) Five of these have now been tested and have been shown not to revert with proflavin.

Let us underline the difference between the *r* loci and the *o* and *h* loci. The former appear to produce proteins which are probably *not* part of the finished phage. For both the *o* and the *h* locus, however, the protein concerned forms part of the finished phage, which presumably would not be viable without it, so that a mutant can be picked up only if it forms an *altered* protein. A mutant which deleted the protein could not be studied.

TABLE 1

	r	h^+
BU	108	57
Proflavin	26	1

It is clear that further work must be done before our generalization—that acridine mutants usually give no protein, rather than a slightly modified one—can be accepted. But if it turns out to be true it would support our hypothesis of the mutagenic action of the acridines, and this may have serious consequences for the naïve theory of mutagenesis, for the following reason.

It has always been a theoretical possibility that the reversions to wild type were not true reversions but were due to the action of "suppressors" (within the gene), possibly very closely linked suppressors. The most telling evidence against this was the existence of the two mutually exclusive classes of mutagens, together with Freese's explanation.

For clearly if the forward mutation could be made at one base-pair and the reverse one at a different base-pair, we should expect, on Freese's hypothesis, exceptions to the rule about the two classes of mutagens. Since these were not found it was concluded that even close suppressors were very rare.

Unfortunately our new hypothesis for the action of acridines destroys this argument. Under this new theory an alteration of a base-pair at one place *could* be reversed by an alteration at a different base-pair, and indeed from what we know (or guess) of the structure of proteins and the dependence of structure on amino acid sequence, we should be surprised if this did not occur.

It is all too easy to conceive, for example, that at a certain point on the polypeptide chain at which there is a glutamic residue in the wild type, and at which the mutation substituted a proline, a further mutation might alter the proline to aspartic acid and that this might appear to restore the wild phenotype, at least as far as could be judged by the rather crude biological tests available. If several base-pairs are needed to code for one amino acid the reverse mutation might occur at a base-pair close to but not identical with the one originally changed.

On our hypothesis this could happen, and yet one would still obtain the two classes of mutagens. The one, typified by base analogues, would produce the substitution of one base for another, and the other, typically produced by acridines, would lead to the addition or subtraction of a base-pair. Consequently the mutants produced by one class could not be easily reversed by the mutagens of the other class.

Thus our new hypothesis reopens in an acute form the question: which back-mutations to wild type are truly to the original wild type, and which only appear to be

so? And on the answers to this question depend our interpretation of all experiments on back-mutation.

We suspect that this problem can most easily be approached by work on systems for which the amino acid sequence of the protein can be studied, such as the phage lysozyme of Dreyer, Anfinsen & Streisinger (personal communications) or the phosphatase from *E. coli* of Levinthal, Garen & Rothman (Garen, 1960). Meanwhile we are continuing our genetic studies to fill out and extend the preliminary results reported here.

Medical Research Council Unit
for Molecular Biology
Cavendish Laboratory

Pathology Laboratory
both of Cambridge University
England

S. BRENNER
LESLIE BARNETT
F. H. C. CRICK

ALICE ORGEL

Received 16 December 1960

REFERENCES

Brenner, S. (1957). *Proc. Nat. Acad. Sci., Wash.* **43**, 687.
Brenner, S., Benzer, S. & Barnett, L. (1958). *Nature*, **182**, 983.
Crick, F. H. C. (1959). In *Brookhaven Symposia in Biology*, **12**, 35.
Freese, E. (1959*a*). *J. Mol. Biol.* **1**, 87.
Freese, E. (1959*b*). *Proc. Nat. Acad. Sci., Wash.* **45**, 622.
Garen, A. (1960). 10th Symposium *Soc. Gen. Microbiol.*, London, 239.
Lerman, L. (1961). *J. Mol. Biol.* **3**, 18.
Streisinger, G. & Franklin, N. C. (1956). In *Cold Spr. Harb. Sym. Quant. Biol.* **21**, 103.

J. Mol. Biol. (1969) **40**, 93–105

Mutations caused by the Insertion of Genetic Material into the Galactose Operon of *Escherichia coli*

J. A. SHAPIRO†

Service de Génétique Cellulaire
Institut Pasteur, Paris XV, France

(*Received 1 August 1968, and in revised form 25 October 1968*)

Density-gradient analysis shows that λ transducing phages which carry spontaneous extreme polar mutations of the galactose operon have higher buoyant densities than otherwise identical phage which carry a wild-type galactose operon. Control experiments show that the density increases are not artifacts of the techniques used to isolate mutant transducing phages and that removal of the spontaneous extreme polar mutations by reversion or recombination leads to a loss of buoyant density. The simplest explanation for these results is that the mutations analysed are the consequences of the linear insertion of foreign DNA into the *E. coli* galactose operon. The insertion hypothesis provides an explanation for the strong polarity and unusual genetic properties of spontaneous extreme polar mutations.

1. Introduction

We have identified a class of spontaneous extreme polar mutations of the galactose operon which do not appear to be the results of base substitutions, frameshifts, or extended deletions (Adhya & Shapiro, 1969; Shapiro & Adhya, 1969). These mutations arose from a forward selection for *gal*$^-$ mutations based on the fact that certain mutants of *Escherichia coli* cannot grow in the presence of galactose (Adhya & Shapiro, 1969). Similar mutations of the lactose and galactose operons have been isolated by others in analogous selection experiments (Malamy, 1966; Saedler & Starlinger, 1967).

To explain the genetic peculiarities and strong polarity of these spontaneous mutations, we proposed that they are the consequences of the random insertion of large pieces of foreign DNA into a structural gene (Shapiro & Adhya, 1969). The reasoning behind this hypothesis is as follows.

Polarity. Because we postulate random insertion, the probability of the inserted fragment's being out of the proper reading frame or even inverted from its normal orientation is high. Thus, a ribosome travelling along a messenger RNA molecule transcribed from an operon containing a fragment of foreign DNA will almost certainly encounter a nonsense codon soon after passing the site of insertion. If the inserted fragment is sufficiently large and contains no chain initiation signals in the proper reading frame, a strong polar effect will result. On the simplest model, the strength of the polarity will increase with the size of the inserted fragment (cf. Newton, Beckwith, Zipser & Brenner, 1965; Newton, 1966).

† Present address: Department of Bacteriology and Immunology, Harvard Medical School, Boston, Mass. 02115, U.S.A.

Genetic properties. Clearly insertion mutations will not be suppressible by extragenic or intragenic suppressors. Those spontaneous extreme polar mutations which do revert can do so in one of two ways: (1) by deletion of the inserted fragment; or (2) by recombination ("selfing") with other regions of the bacterial chromosome homologous to the mutated locus (cf. Clark, 1964). Neither of these mechanisms should be greatly influenced by base-analogue or frameshift mutagens. Hence our negative results on mutagen-induced reversion (Adhya & Shapiro, 1969).

The fact that many presumed insertion mutations do not revert at all is to be expected if the insertion process resembles the integration of known episomes into structural genes: integration of a temperature-sensitive F-*lac* into various sites on the bacterial chromosome results in many non-reverting mutations (Beckwith, Signer & Epstein, 1966). Providing that the insertion process does not destroy large regions of the mutated gene, insertion mutations should behave as point mutations in genetic crosses (possibly displaying strong marker effects on quantitative results).

The insertion hypothesis predicts that mutant operons will have a higher DNA content than the parent wild-type operon. The observation that each *λdg* transducing phage has a characteristic buoyant density in CsCl solution depending on its DNA content made it possible to test this prediction (Weigle, Meselson & Paigen, 1959; Kayajanian & Campbell, 1966). *λdg*$^-$ phages carrying various spontaneous extreme polar mutations were isolated from a wild-type *λdg*$^+$ phage of known density by homogenote selection (Morse, Lederberg & Lederberg, 1956). The densities of these *λdg*$^-$ phages were then compared with that of the original *λdg*$^+$ phage. If a mutant operon contains added DNA, then the corresponding *λdg*$^-$ phage should be found in a denser region of the gradient than the *λdg*$^+$ phage. Four independent spontaneous extreme polar *kt*$^-$ mutations of the *galT* gene were chosen for study: S101, S104, S114 and S188 (Adhya & Shapiro, 1969). The positions of these four mutations in the *galT* gene are shown in italic characters in Figure 1. The experiments presented below

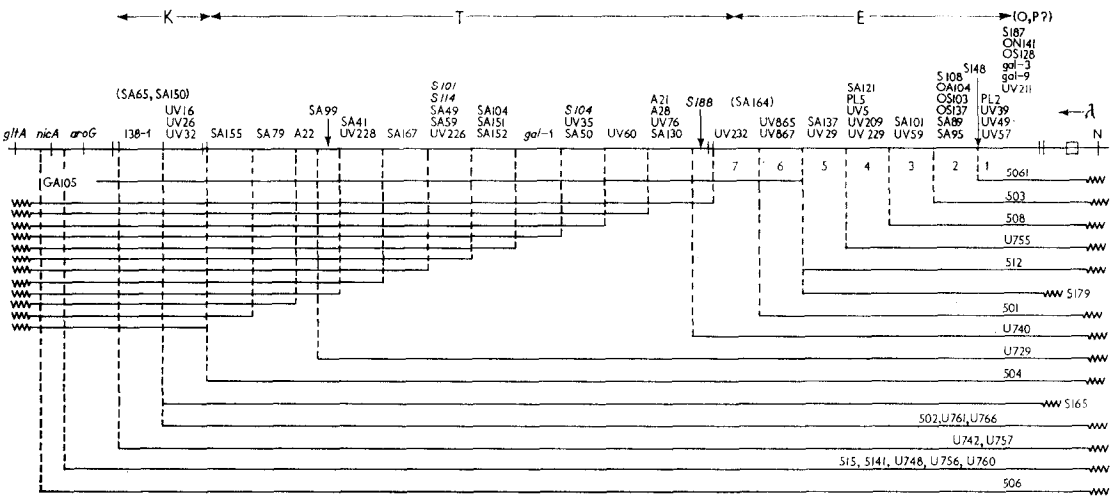


FIG. 1. A deletion map of the *E. coli* galactose operon taken from Shapiro & Adhya (1969).

show unambiguously that the presence of each of these four mutations on the DNA of a λdg phage results in an increase of that phage's buoyant density.

2. Materials and Methods

(a) *Bacteria and bacteriophages*

All of the *E. coli* K12 strains used are gal^-. The origins of all but one of the gal^- mutations are described in Adhya & Shapiro (1969). The map position of all the mutations except UV12 are shown in Fig. 1.

Strains MS71, MS72, MS73 and MS80 carry, respectively, the S101, S104, S114 and S188 kt^- spontaneous extreme polar mutations. Strains H57 and MS91 carry, respectively, the polar UV57 e^- amber mutation and the UV16 k^- point mutation. These six strains were used to prepare various λdg^- phages by homogenote selection as described below.

Strain MS2 carries the PL5 non-polar e^- mutation of Buttin (1963), strain MS8 the UV12 non-polar t^- mutation, strain MS132 the UV16 k^- mutation and the su^+_{III} amber suppressor, and strain S165 the S165 gal^- deletion which covers all the other mutations. These strains were used as transduction recipients for the differential assays of various λdg phages.

Strain MS2 was used as indicator strain for active λ^{857} phage in plaque assays.

λ^{857} (Sussman & Jacob, 1962), λc_{Isus34}, and λvir were obtained from the collection of F. Jacob. A λ^{857} lysogen was made from a derivative of Hfr H carrying a wild-type galactose operon. This lysogen was induced to provide an LFT† lysate. λ^{857} has a mutant, temperature-sensitive repressor; hence, strains lysogenic for λ^{857} can be induced by a simple temperature shift from 30°C to 42°C.

(b) *Media*

These have already been described (Adhya & Shapiro, 1969).

(c) *Isolation of* λ^{857}dg^+XV *by the "orgy" technique*

An LFT lysate of λ^{857} was adsorbed to strain MS91 at high multiplicity (m.o.i. ≥ 10), and the infected bacteria plated on M9–galactose agar. After 72 hr at 30°C, approximately 500 gal^+ transductant clones appeared on each plate. These would virtually all be double lysogens for active λ^{857} and various $\lambda^{857}dg^+$ phages. Approximately 1500 transductant clones were resuspended *en masse* in L-broth and induced to yield an HFT orgy lysate containing several hundred particles per ml. of many different λdg^+ phages. This HFT lysate was centrifuged to equilibrium in CsCl. Fractions of this gradient were assayed for transducing activity on strain MS91. A gal^+ transductant clone arising from the assay of a light fraction was purified and induced to yield an HFT lysate of $\lambda^{857}dg^+$XV, hereafter referred to as λdg^+.

(d) *Selection of* λdg *homogenotes*

Cultures of various gal^- strains were infected at low multiplicity (m.o.i. $< 10^{-2}$) with an HFT lysate of λdg^+ and plated on M9–galactose agar. Transduction at low multiplicity ensured that all transductants would carry only λdg^+ as prophage. After 72 hr at 30°C, transductant clones were repurified on galactose–tetrazolium–agar, and from each transduction a gal^+ clone which segregated red gal^- colonies was grown to saturation in L-broth at 25°C. These cultures were streaked on galactose–tetrazolium–agar previously spread with 10^{10} particles of λc_{Isus34} to counter-select non-immune segregants. Red gal^- segregants were repurified and tested for λ immunity (by cross-streaking against λc_{Isus34} and λvir) and for production of active phage. Those gal^- strains which proved to be defective lysogens were homogenotes carrying λdg^- prophages containing the appropriate gal^- mutation. The nature of the gal^- mutation on each prophage was verified in all cases by testing HFT lysates obtained after superinfection for recombination and complementation with suitable gal^- strains. We call a derivative of λdg^+ carrying the S101 mutation λdg^-_{S101}, and so forth.

† Abbreviations used: LFT, low-frequency transducing; HFT, high-frequency transducing; m.o.i., multiplicity of infection.

(e) *Production of HFT lysates by superinfection of defective lysogens*

Defective lysogens for various λ*dg* phages were grown at 30°C to approximately 10^9 cells/ml. in M9–maltose medium supplemented with 0·4% Casamino acids. The bacteria were resuspended in 0·01 M-$MgSO_4$ and a lysate of λ^{857} was adsorbed at a multiplicity of 5 for 20 min at 30°C. The adsorption mixture was diluted into twice the volume of L-broth, incubated 20 min at 42°C, and then incubated at 37°C with vigorous aeration until lysis. The HFT lysates obtained in this way never contained more than 10% transducing phages.

For the vegetative phage crosses between $\lambda dg^-{}_{S114}$ and the $\lambda dg^-{}_{UV16}$ and $\lambda dg^-{}_{UV57}$ phages, an HFT lysate of $\lambda dg^-{}_{S114}$ and λ^{857} was used to infect defective lysogens for the other λdg^- phages. In these crosses, adsorption was carried out at 42°C.

(f) *Isolation of* λdg^+ *revertants of* $\lambda dg^-{}_{S104}$ *and* $\lambda dg^-{}_{S188}$

The λdg^+ revertant phage in HFT lysates of $\lambda dg^-{}_{S104}$ and $\lambda dg^-{}_{S188}$ were isolated by transduction of strain S165 (see Table 1). Transduction was performed at m.o.i ~ 5 to ensure that practically every transductant obtained was lysogenic for both λ^{857} and a revertant λdg^+. The gal^+ transductant clones from each experiment (32 from the transduction with $\lambda dg^-{}_{S104}$, 136 from the transduction with $\lambda dg^-{}_{S188}$) were then mixed together and induced to provide two HFT lysates containing several thousand copies of each of the revertant λdg^+ phages isolated. This amplification procedure should not have altered the densities of the λdg^+ revertant phages (Weigle, 1961).

(g) *Preparative density-gradient centrifugation*

In general, preparative centrifugation was performed as described by Weigle *et al.* (1959). 3-ml. gradients were centrifuged in the SW39 or SW5OL swinging-bucket rotor at 22,000 rev./min for 48 hr at 10°C and collected in 1-drop fractions (except for the experiment presented in Fig. 2, in which 2-drop fractions were collected). 10-ml. gradients were run in the SW40 fixed-angle rotor at 25,000 rev./min for 48 hr at 10°C and collected in 3-drop fractions. Phage suspensions with a density of approximately 1·50 g cm^{-3} were prepared either by mixing phage lysates with an equal volume of a saturated solution of CsCl in 0·01 M-Tris buffer (pH 7) or by dissolving the appropriate mass of CsCl in a given volume of phage lysate. L-broth was used as the buffer for phage lysates. Various Spinco model L centrifuges were used.

(h) *Assay of transducing phages*

Each fraction to be assayed was diluted appropriately in M9 buffer. A 0·1-ml. sample of the dilution was then added to approximately 0·2 ml. of a lysate of λ^{857} helper phage grown lytically on strain MS2. (The helper phage lysate contained more than 10^{10} active phage/ml. and no transducing phage.) To this mixture was added approximately 0·2 ml. of the appropriate gal^- indicator strain resuspended in 0·01 M-$MgSO_4$ at approximately 2×10^9 cells/ml. Adsorption was carried out for 30 min at 30°C. The adsorption mixture was stored overnight at 4°C, and then all of it was plated on M9–galactose–agar. Plates were scored for transductant clones after 72 hr at 30°C. Because the entire mixture was plated, the volumes of helper phage and indicator strain used could vary within a factor of 2 without affecting the final result.

Differential assays of more than one λ*dg* phage in a single gradient were based on the transduction specificities summarized in Table 1. λdg^+ could be assayed uniquely on the *gal* deletion strain S165 because none of the λdg^- phages can transduce this strain to gal^+. λdg^+ transduces S165 by complementation (i.e. by lysogenization). Because λdg^+ phages were never in excess, λdg^- phages carrying the UV16 k^- marker and the four kt^- mutations could be assayed on the e^- strain MS2 which they all transduce by complementation. In the experiment presented in Fig. 6(a), $\lambda dg^-{}_{UV16}$ could be distinguished from $\lambda dg^-{}_{S114}$ by assaying on the t^- strain MS8: $\lambda dg^-{}_{UV16}$ transduces MS8 efficiently by complementation, while $\lambda dg^-{}_{S114}$ transduces MS8 at a negligible frequency by recombination. In the experiment presented in Fig. 6(b), $\lambda dg^-{}_{UV57}$ could be distinguished from $\lambda dg^-{}_{S114}$ by assaying on the k^-su^+ strain MS132 because the efficiency of transduction by $\lambda dg^-{}_{UV57}$

is at least tenfold higher than by $\lambda dg^{-}{}_{S114}$. $\lambda dg^{-}{}_{S114}$ transduces MS132 only by recombination; $\lambda dg^{-}{}_{UV57}$ transduces MS132 by weak complementation and recombination because (i) the $su^{+}{}_{III}$ suppressor will suppress the polarity of the UV57 amber mutation and partially restore kinase gene activity to the transducing phage and (ii) the distance between the UV57 and UV16 sites is sufficiently large to permit frequent recombination (Fig. 1). (Note that the $su^{+}{}_{III}$ suppressor does not suppress the UV57 mutation to gal^{+}: Adhya & Shapiro, 1969.)

TABLE 1

Transduction specificities of the different λdg *phages for various recipient strains*

Recipient strain:	S165 ($\triangle gal$)	MS2 (e^{-})	MS8 (t^{-})	MS132 ($k^{-}su^{+}$)
λdg:				
λdg^{+}	+	+	+	+
$\lambda dg^{-}{}_{UV16}$	—	+	+	
$\lambda dg^{-}{}_{UV57}$	—	—	—	+
$\lambda dg^{-}{}_{S101}$, $\lambda dg^{-}{}_{S104}$, $\lambda dg^{-}{}_{S188}$	—	+		
$\lambda dg^{-}{}_{S114}$	—	+	—	weak +

The symbols indicate the probability that infection of a given strain by a λdg particle (in the presence of excess helper phage) will yield a gal^{+} transductant. (+): $\geq 10^{-1}$; (—): $< 10^{-3}$; (weak +): $\sim 10^{-2}$.

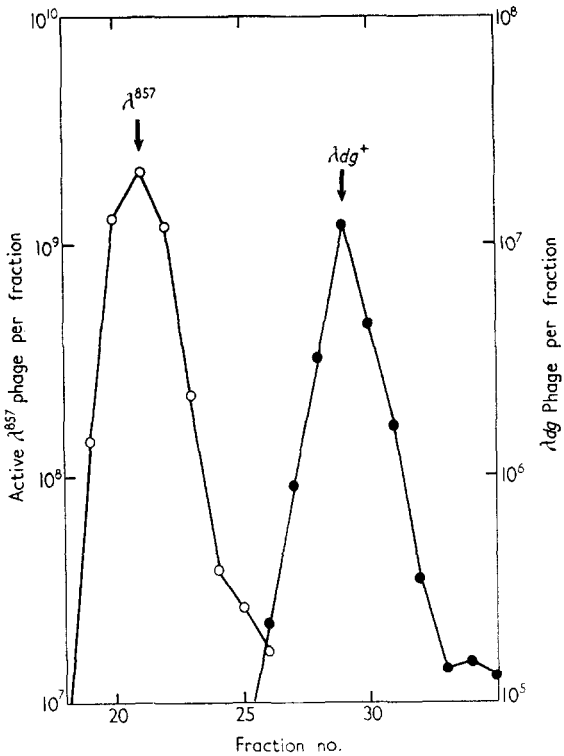


FIG. 2. CsCl density-gradient profile of an HFT lysate of λdg^{+}. 22,000 rev./min, SW50L rotor, miniscus at fraction 54.

—○—○—, Plaque-forming units; —●—●—, transducing phage assayed on strain S165.

3. Results

(a) *Isolation of λdg⁺*

An HFT lysate of a λdg^+ phage with a low buoyant density derived from λ^{857} was obtained as described in Materials and Methods. The equilibrium CsCl density-gradient profile of this lysate is shown in Figure 2. Both the active λ^{857} and λdg^+ phages form single bands. The distance between the two peaks is approximately 15% of the total gradient. Knowing the geometry and speed of the rotor and the mean density of the gradient, the density difference ($\Delta\rho$) between λ^{857} and λdg^+ can be calculated. This has been done for several gradients, and the results vary between $\Delta\rho = 0{\cdot}012$ and $\Delta\rho = 0{\cdot}014$ g cm^{-3}. This corresponds to a DNA difference between λ^{857} and λdg^+ of between 0·12 and 0·14 λ equivalents (Weigle *et al.*, 1959) or at least 6,000 base pairs (Caro, 1965).

(b) *Homogenote formation using a* gal⁻ *point mutation*

A λdg^- phage carrying the UV16 k^- point mutation was isolated by homogenote selection as described in Materials and Methods. An HFT lysate of this λdg^-_{UV16} phage was prepared by superinfection and mixed with an HFT lysate of the parental λdg^+ phage so that the λdg^-_{UV16} particles were in excess. The mixture of lysates was centrifuged in CsCl to equilibrium and differentially assayed for active λ^{857}, λdg^+ and λdg^-_{UV16} phages. The results are shown in Figure 3. There is no significant difference

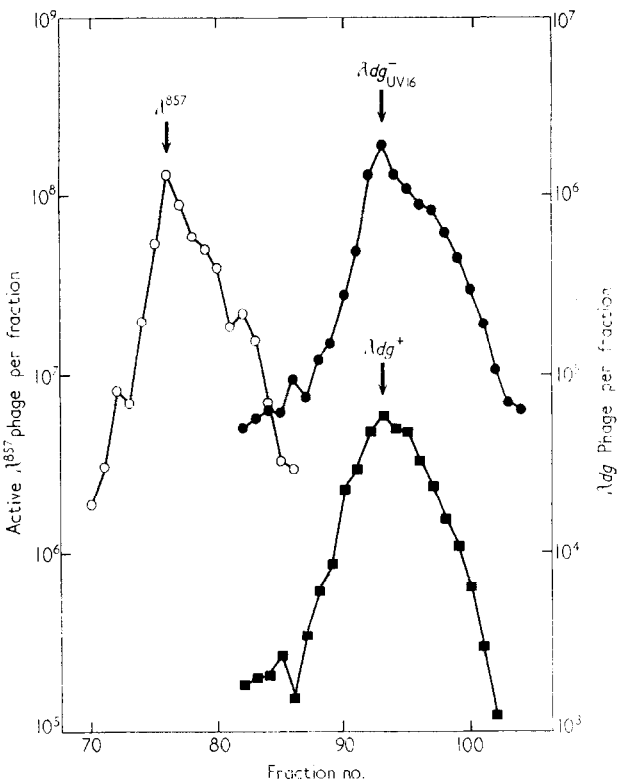


FIG. 3. A mixture of HFT lysates of λdg^-_{UV16} and λdg^+. 22,000 rev./min, SW39 rotor, miniscus at fraction 109.

—○—○—, Plaque-forming units; —●—●—, transducing phage assayed on strain MS2; —■—■—, transducing phage assayed on strain S165.

between the buoyant densities of the two transducing phages. Thus, the process of homogenote formation does not *per se* cause any change in the density of a transducing phage. This observation was previously made by Weigle *et al.* (1959).

(c) *Homogenote formation using spontaneous extreme polar mutations*

λdg^- phages carrying the four spontaneous extreme polar kt^- mutations were also prepared by homogenote selection, and the experiment described in Figure 3 was repeated for each of the four phages. The results are shown in Figure 4. In each case the peak of λdg^- phage is closer to the λ^{857} peak than is the peak of λdg^+ phage. In those gradients where the ratio of λdg^- to λdg^+ particles was relatively low (Fig. 4(b), (c) and (e)), assay of transducing phage on strain MS2 gave a bimodal distribution with one of the peaks corresponding to the λdg^+ phage. The slight shoulder on the $\lambda dg^-{}_{S101}$ peak in Figure 4(a) is also due to λdg^+ phages. Comparison of Figure 4(e) with Figure 4(c) shows that essentially the same results are obtained in the fixed-angle and swinging-bucket rotors. In contrast with the previous experiment, these results show that formation of λdg^- phages carrying these four spontaneous kt^- mutations results in an increase of buoyant density.

(d) λdg^+ *revertants of* λdg^- *phages carrying the S104 and S188 mutations*

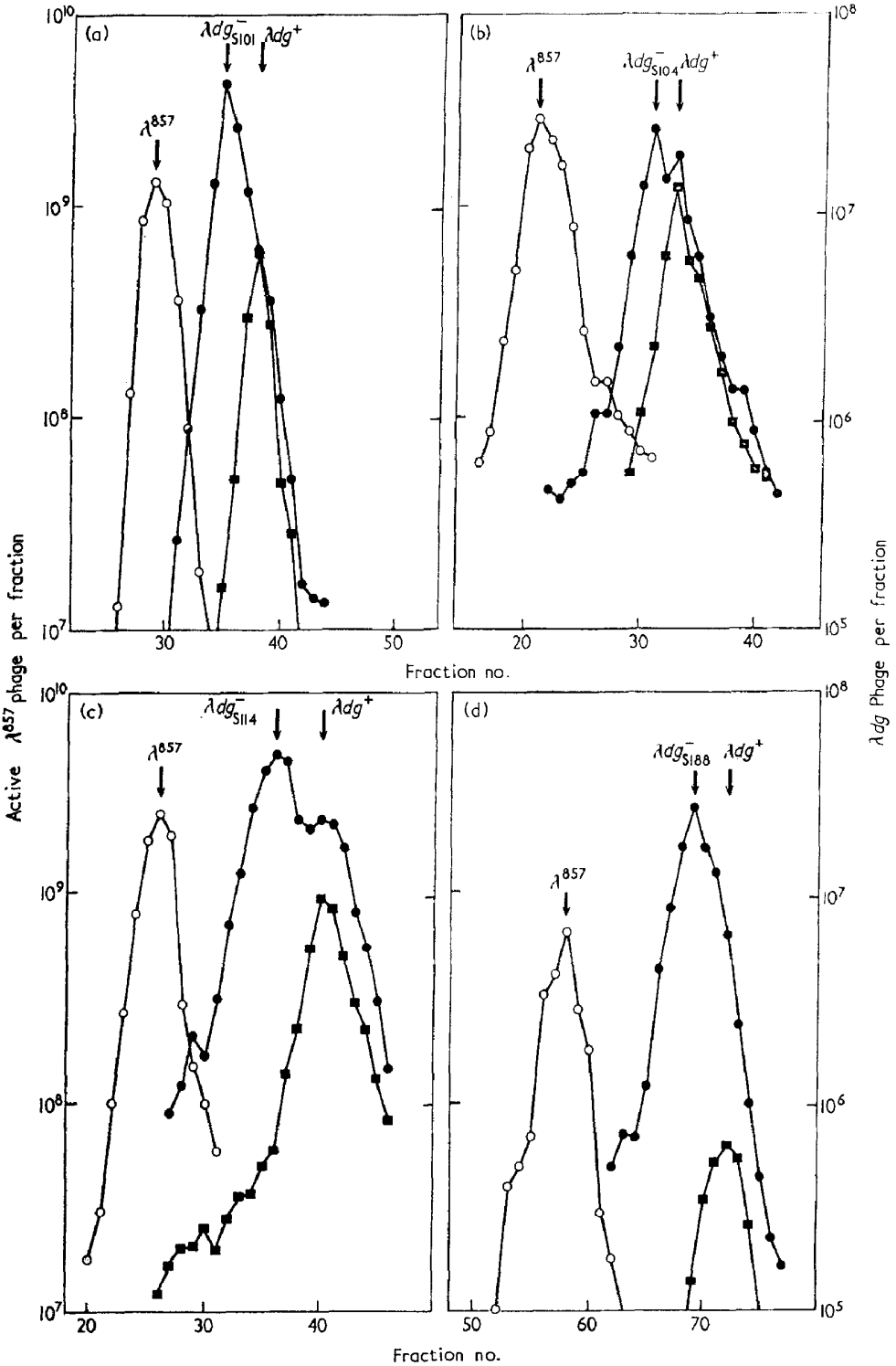
Despite the control experiment presented in Figure 3, the possibility remained that the density increases associated with isolation of λdg^- phages carrying the spontaneous kt^- mutations were somehow artifacts of the homogenote selection procedure. Hence it was necessary to demonstrate that the density differences were specifically associated with the presence of the gal^- mutation on the λdg phage. One way to do this was to isolate λdg^+ revertants of the λdg^- phages and show that the λdg^+ phages have a lower density then the parental λdg^- phages.

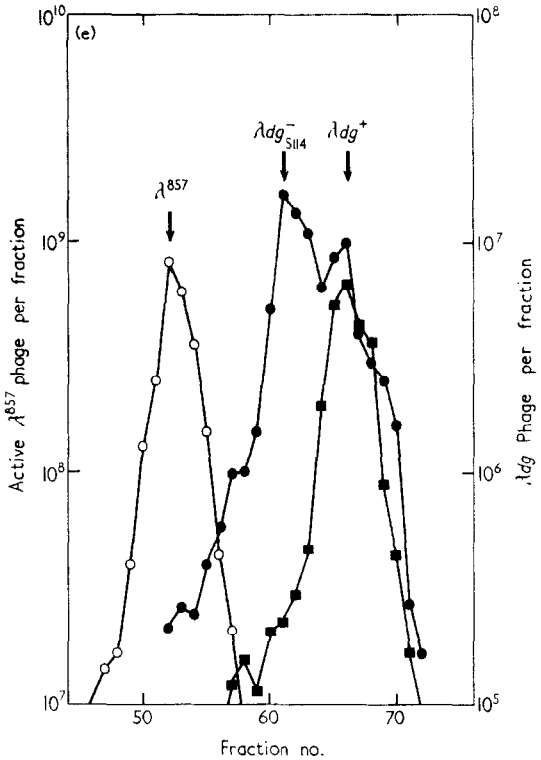
Accordingly, the revertant λdg^+ phages in HFT lysates of $\lambda dg^-{}_{S104}$ and $\lambda dg^-{}_{S188}$ were isolated as described in Materials and Methods.† HFT lysates of these revertant λdg^+ phages were centrifuged to equilibrium together with lysates of the parental λdg^- phages and assayed differentially. The results are presented in Figure 5. It can be seen that reversion of both $\lambda dg^-{}_{S104}$ and $\lambda dg^-{}_{S188}$ to λdg^+ entails a loss of buoyant density. The positions of the revertant λdg^+ peaks suggest that these phages are the same as the original λdg^+ phage.

(e) *Segregation of the density difference associated with the S114 mutation in vegetative crosses*

A second method of demonstrating the specific association between the increased density and gal^- character of λdg^- phages carrying the spontaneous kt^- mutations was to show that both properties segregate together. For this purpose, vegetative crosses between $\lambda dg^-{}_{S114}$ and λdg^- phages carrying the UV16 k^- and UV57e^- amber markers were performed as described in Materials and Methods. Because these two gal^- markers are located on either side of the S114 mutation (Fig. 1), a density marker outside the galactose operon should be detectable among the λdg^+ recombinants in at least one of the two crosses. The density-gradient analyses of the lysates

† This experiment could not be done with the $\lambda dg^-{}_{S101}$ or $\lambda dg^-{}_{S114}$ phages because neither the S101 or S114 mutations ever reverts (Adhya & Shapiro, 1969).


(a)
λdg⁻S101
λdg⁺
λ857
(b)
λ857
λdg⁻S104
λdg⁺
(c)
λdg⁻S114
λdg⁺
λ857
(d)
λdg⁻S188
λdg⁺
λ857
Active λ857 phage per fraction
λdg Phage per fraction
Fraction no.
Fraction no.



arising from these crosses are shown in Figure 6. A schematic representation of each cross is given in the upper left-hand corner. In both crosses the λdg^+ recombinant phages have a lower buoyant density than the λdg^-_{S114} phage. Moreover, the coincidence of the λdg^+ peak with the λdg^-_{UV16} peak in Figure 6(a) and with the λdg^-_{UV57} peak in Figure 6(b) demonstrates that the recombinant λdg^+ phages are identical to the original λdg^+. The shoulder on the λdg^-_{UV57} peak in Figure 6(b) is due to recombinational transduction of strain MS132 by λdg^-_{S114}. Thus, the higher density of the λdg^-_{S114} phage segregates rigorously with the S114 mutation and is not due to a change in either the right- or left-hand end of the transducing phage genome.

Comparable results were obtained in marker-rescue experiments in which an HFT lysate of λdg^-_{S114} was used to infect non-lysogenic strains carrying *gal*$^-$ markers on either side of the S114 mutation.

FIG. 4. Density-gradient analyses of spontaneous *kt*$^-$ mutations.

(a) A mixture of HFT lysates of λdg^-_{S101} and λdg^+. 22,000 rev./min, SW50L rotor, miniscus at fraction 67.

(b) A mixture of HFT lysates of λdg^-_{S104} and λdg^+. 22,000 rev./min, SW50L rotor, miniscus at fraction 90. Phage titres multiplied by 10.

(c) A mixture of HFT lysates of λdg^-_{S114} and λdg^+. 22,000 rev./min, SW50L rotor, miniscus at fraction 103. Active phage titre multiplied by 10, transducing phage titre multiplied by 100.

(d) A mixture of HFT lysates of λdg^-_{S188} and λdg^+. 25,000 rev./min, SW40 rotor, miniscus at fraction 88.

(e) A mixture of HFT lysates of λdg^-_{S114} and λdg^+ 25,000 rev./min, SW40 rotor, miniscus at fraction 87.

—○—○—, Plaque-forming units; —●—●—, transducing phage assayed on strain MS2; —■—■—, transducing phage assayed on strain S165.

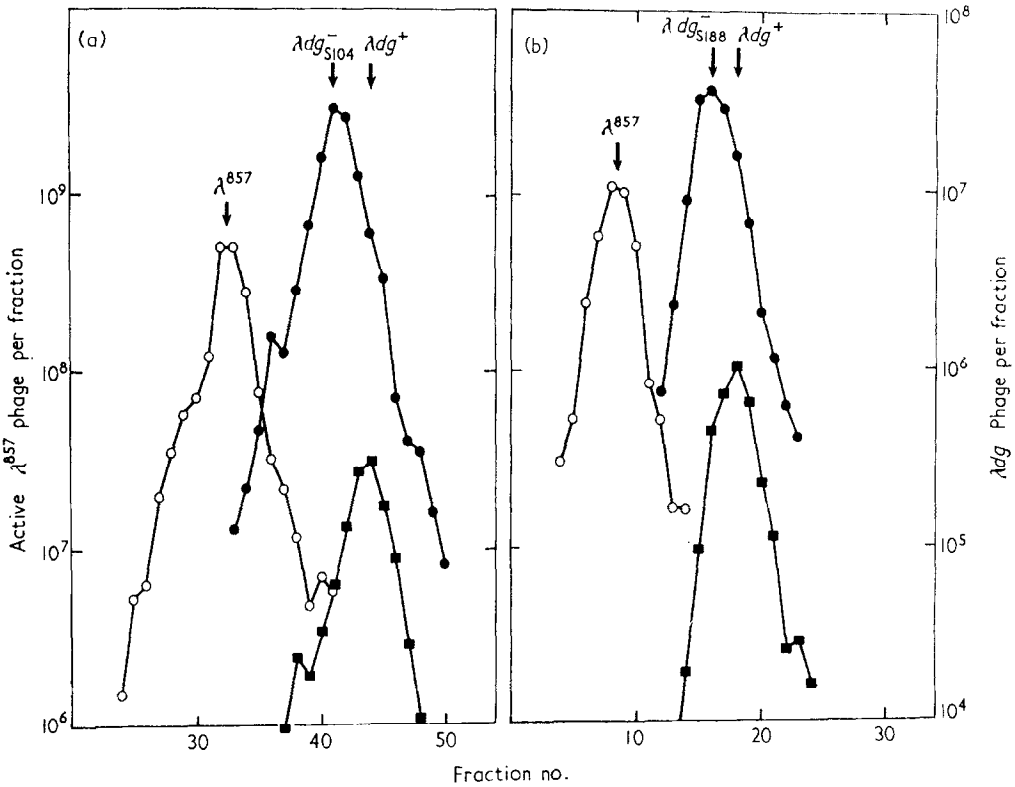


FIG. 5. Density-gradient analyses of revertant galactose operons.

(a) A mixture of HFT lysates of λdg^-_{S104} and λdg^+ revertant phages isolated by transduction of strain S165. 25,000 rev./min, SW40 rotor, miniscus at fraction 70.

(b) A mixture of HFT lysates of λdg^-_{S188} and λdg^+ revertant phages isolated by transduction of strain S165. 22,000 rev./min, SW39 rotor, miniscus at fraction 59.

—○—○—, Plaque-forming units; —●—●—, transducing phage assayed on strain MS2; —■—■—, transducing phage assayed on strain S165.

(f) *Sizes of the various mutations*

From the data presented above, estimates can be made of the amount of inserted material in each mutation. Assuming a difference of 6×10^3 base pairs between the λ^{857} and λdg^+ genomes (see above), linear interpolations based on the positions of λ^{857}, λdg^- and λdg^+ peaks in different experiments give the values in Table 2. At least for the S114 mutation, the observed density difference is reproducible ($\pm$ 25%) from experiment to experiment.

4. Discussion

The density-gradient experiments presented in Figure 4 show that λdg^- phages carrying any one of the four spontaneous extreme polar mutations studied have a higher buoyant density than the λdg^+ phage from which they are derived. Control experiments show that the increased density is not due to the process of homogenote formation and that removal of the spontaneous mutations by reversion or recombination leads to a loss of buoyant density. Although other possibilities are not excluded, the simplest explanation for these results is that the spontaneous extreme polar

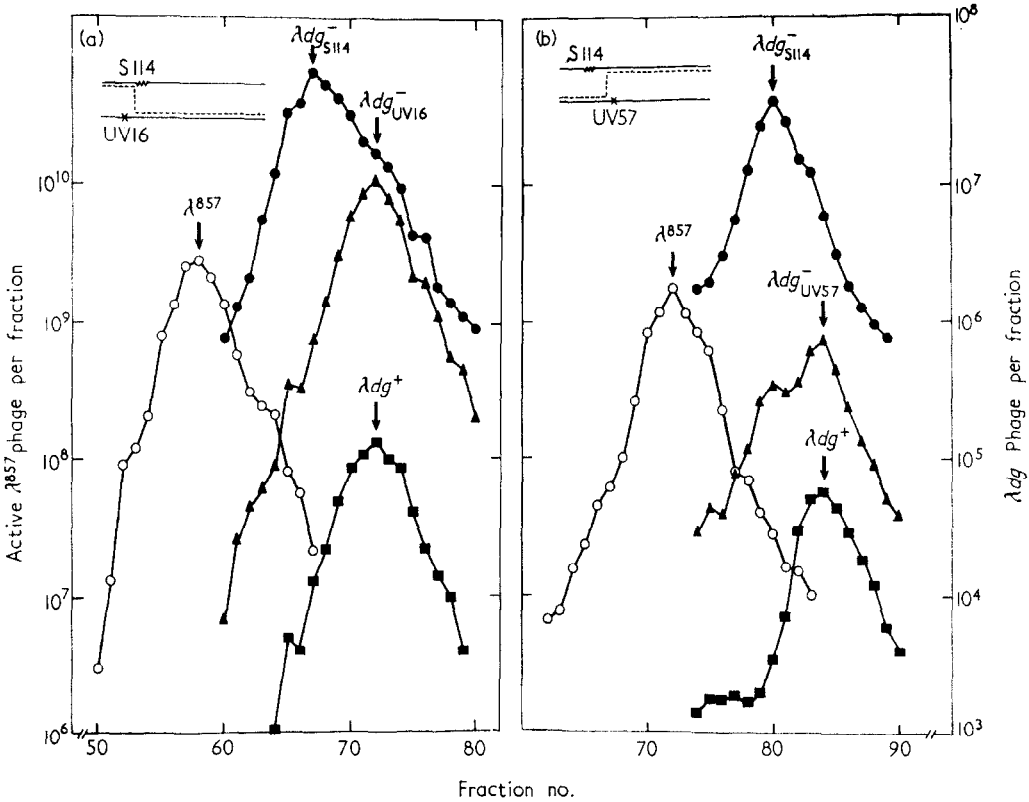


FIG. 6. Segregation of density increase with the S114 mutation in vegetative crosses.

(a) $\lambda dg^-_{S114} \times \lambda dg^-_{UV16}$: an HFT lysate obtained by superinfection of a defective homogenote strain carrying λdg^-_{UV16} as prophage with an HFT lysate of λdg^-_{S114}. 22,000 rev./min, SW39 rotor, miniscus at fraction 102.

(b) $\lambda dg^-_{S114} \times \lambda dg^-_{UV57}$: an HFT lysate obtained by superinfection of a defective homogenote strain carrying λdg^-_{UV57} as prophage with an HFT lysate of λdg^-_{S114}. 22,000 rev./min, SW39 rotor, miniscus at fraction 96.

—○—○—, Plaque-forming units; —●—●—, transducing phage assayed on strain MS2; —■—■—, transducing phage assayed on strain S165; —▲—▲—, transducing phage assayed on strain MS8 (a) or strain MS132 (b).

TABLE 2

Sizes of the four insertion mutations

Mutation	Approximate mutation size (10^3 base pairs)	Gradient
		Figure
S101	2	4(a)
S104	1	4(b)
	1·5	5(a)
S114	1·7	4(c)
	2·1	4(e)
	2·1	6(a)
	2	6(b)
S188	1·3	4(d)
	1·2	5(b)

The approximate mutation size is the number of base pairs of DNA of 50% (G + C) content needed to account for the density difference between each λdg^- phage and the parental λdg^+ phage. For calculations based on Figs 5 and 6, we assume that the revertant or recombinant λdg^+ phage have the same density as the parental λdg^+ phage.

mutations we have examined are the consequences of linear insertion of DNA into the *E. coli* galactose operon. The fact that all four mutations studied fulfill a strong prediction of the insertion hypothesis argues in favour of the validity of the model.

As discussed in the Introduction, the insertion model provides an explanation for the properties of spontaneous extreme polar mutations in the galactose and lactose operons. A possible explanation of why some spontaneous extreme polar mutations appear to be *absolutely* polar (Malamy, 1966; Adhya & Shapiro, 1969) is that they correspond to nucleotide sequences signalling the end of messenger RNA transcription. Although the insertion hypothesis provides a ready explanation for the presence of such transcription signals at many sites within a structural gene, it should be kept in mind that even without special sequences, large insertions should have virtually absolute polar effects (cf. Newton *et al.*, 1965).

The evolutionary importance of insertion mutations depends upon how common they are. Assuming that nearly all spontaneous extreme polar mutations of the galactose operon are the result of insertions, it can be argued that insertion events represent a significant proportion of all spontaneous mutations in *E. coli*. In one experiment, 77 spontaneous mutations of the galactose operon were isolated by selecting for loss of galactokinase activity; of these, at least 14 (18%) have the properties of insertion mutations, and only two are extended deletions (Shapiro, 1967; cf. Adhya & Shapiro, 1969). This high percentage agrees with the results of an analogous but independent experiment of Saedler & Starlinger (1967) in which at least 17 out of 147 spontaneous mutations (12%) have properties ascribable to insertion events. In view of the apparent high frequency of insertion events, it should be pointed out that insertion of DNA fragments containing special sequences can account for the properties of many complex mutations besides polar mutations: for example, gal_3 and its constitutive revertants in *E. coli* (Hill & Echols, 1966; Morse, 1967); unstable partial revertants of the *his-203* mutant of *Salmonella* (Ames, Hartman & Jacob, 1963); the c_{17} mutant of phage λ (Pereira da Silva & Jacob, 1968; Packman & Sly, 1968); and some operator-constitutive mutants in various operons.

I would like to thank Alex Fritsch for his patience and tolerance in teaching me how to make preparative CsCl gradients; Harvey Eisen, Ethan Signer, Sydney Brenner, Gerard Buttin, Maurice Hoffnung and Jon Beckwith for lively and helpful discussions: and Hubert Condamine and Jeffrey Miller for reading the manuscript. I am grateful to Professor Francois Jacob for the hospitality of his laboratory. This work was aided by grants from the Délégation Générale à la Recherche Scientifique et Technique, the Centre National de la Recherche Scientifique and the Commissariat à l'Energie Atomique, and the Jane Coffin Childs Memorial Fund for Medical Research of which I was a Fellow.

REFERENCES

Adhya, S. & Shapiro, J. A. (1969). *Genetics*, in the press.
Ames, B. N., Hartman, P. E. & Jacob, F. (1963). *J. Mol. Biol.* **7**, 23.
Beckwith, J. R., Signer, E. R. & Epstein, W. (1966). *Cold Spr. Harb. Symp. Quant. Biol.* **31**, 393.
Buttin, G. (1963). *J. Mol. Biol.* **7**, 183.
Caro, L. (1965). *Virology*, **25**, 226.
Clark, A. J. (1964). *Z. Vererbungslehre*, **95**, 368.
Hill, C. W. & Echols, H. (1966). *J. Mol. Biol.* **19**, 38.
Kayajanian, G. & Campbell, A. (1966). *Virology*, **30**, 482.
Malamy, M. H. (1966). *Cold Spr. Harb. Symp. Quant. Biol.* **31**, 189.

Morse, M. L. (1967). *Genetics*, **56**, 331.
Morse, M. L., Lederberg, E. M. & Lederberg, J. (1956). *Genetics*, **41**, 758.
Newton, W. A. (1966). *Cold Spr. Harb. Symp. Quant. Biol.* **31**, 181.
Newton, W. A., Beckwith, J. R., Zipser, D. & Brenner, S. (1965). *J. Mol. Biol.* **14**, 290.
Packman, S. & Sly, W. S. (1968). *Virology*, **34**, 778.
Pereira da Silva, L. H. & Jacob, F. (1968). *Ann. Inst. Past.* **115**, 145.
Saedler, H. & Starlinger, P. (1967). *Molec. Gen. Genetics*, **100**, 178.
Shapiro, J. A. (1967). Ph.D. Thesis, University of Cambridge.
Shapiro, J. A. & Adhya, S. (1969). *Genetics*, in the press.
Sussman, R. & Jacob, F. (1962). *C. R. Acad. Sci. Paris*, **254**, 1517.
Weigle, J. (1961). *J. Mol. Biol.* **3**, 393.
Weigle, J., Meselson, M. & Paigen, K. (1959). *J. Mol. Biol.* **1**, 379.

MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE[1,2]

S. E. LURIA[3] AND M. DELBRÜCK
Indiana University, Bloomington, Indiana, and Vanderbilt University, Nashville, Tennessee

Received May 29, 1943

INTRODUCTION

WHEN a pure bacterial culture is attacked by a bacterial virus, the culture will clear after a few hours due to destruction of the sensitive cells by the virus. However, after further incubation for a few hours, or sometimes days, the culture will often become turbid again, due to the growth of a bacterial variant which is resistant to the action of the virus. This variant can be isolated and freed from the virus and will in many cases retain its resistance to the action of the virus even if subcultured through many generations in the absence of the virus. While the sensitive strain adsorbed the virus readily, the resistant variant will generally not show any affinity to it.

The resistant bacterial variants appear readily in cultures grown from a single cell. They were, therefore, certainly not present when the culture was started. Their resistance is generally rather specific. It does not extend to viruses that are found to differ by other criteria from the strain in whose presence the resistant culture developed. The variant may differ from the original strain in morphological or metabolic characteristics, or in serological type or in colony type. Most often, however, no such correlated changes are apparent, and the variant may be distinguished from the original strain only by its resistance to the inciting strain of virus.

The nature of these variants and the manner in which they originate have been discussed by many authors, and numerous attempts have been made to correlate the phenomenon with other instances of bacterial variation.

The net effect of the addition of virus consists of the appearance of a variant strain, characterized by a new stable character—namely, resistance to the inciting virus. The situation has often been expressed by saying that bacterial viruses are powerful "dissociating agents." While this expression summarizes adequately the net effect, it must not be taken to imply anything about the mechanism by which the result is brought about. A moment's reflection will show that there are greatly differing mechanisms which might produce the same end result.

D'HERELLE (1926) and many other investigators believed that the virus by direct action induced the resistant variants. GRATIA (1921), BURNET (1929), and others, on the other hand, believed that the resistant bacterial variants are produced by mutation in the culture prior to the addition of virus. The

[1] Theory by M. D., experiments by S. E. L.

[2] Aided by grants from the DAZIAN FOUNDATION FOR MEDICAL RESEARCH and from the ROCKEFELLER FOUNDATION.

[3] Fellow of the GUGGENHEIM FOUNDATION.

virus merely brings the variants into prominence by eliminating all sensitive bacteria.

Neither of these views seems to have been rigorously proved in any single instance. BURNET'S (1929) work on isolations of colonies, morphologically distinguishable prior to the addition of virus, which proved resistant to the virus comes nearest to this goal. His results appear to support the mutation hypothesis for colony variants. It may seem peculiar that this simple and important question should not have been settled long ago, but a close analysis of the problem in hand will show that a decision can only be reached by a more subtle quantitative study than has hitherto been applied in this field of research.

Let us begin by restating the basic experimental finding.

A bacterial culture is grown from a single cell. At a certain moment the culture is plated with virus in excess. Upon incubation, one finds that a very small fraction of the bacteria survived the attack of the virus, as indicated by the development of a small number of resistant colonies, consisting of bacteria which do not even adsorb the virus.

Let us focus our attention on the first generation of the resistant variant—that is, on those bacteria which survive immediately after the virus has been added. These survivors we may call the "original variants." We know that these bacteria and their offspring are resistant to the virus. We may formulate three alternative hypotheses regarding them.

a. *Hypothesis of mutation to immunity.* The original variants were resistant before the virus was added, and, like their offspring, did not even adsorb it. On this hypothesis the virus did not interact at all with the original variants, the origin of which must be ascribed to "mutations" that occur quite independently of the virus. Naming such hereditary changes "mutations" of course does not imply a detailed similarity with any of the classes of mutations that have been analyzed in terms of genes for higher organisms. The similarity may be merely a formal one.

b. *Hypothesis of acquired immunity.* The original variants interacted with the virus, but survived the attack. We may then inquire into the predisposing cause which effected the survival of these bacteria in contradistinction to the succumbing ones. The predisposing cause may be hereditary or random. Accordingly we arrive at two alternative hypotheses—namely,

b_1. *Hypothesis of acquired immunity of hereditarily predisposed individuals.* The original variants originated by mutations occurring independently of the presence of virus. When the virus is added, the variants will interact with it, but they will survive the interaction, just as there may be families which are hereditarily predisposed to survive an otherwise fatal virus infection. Since we know that the offspring of the original variants do not adsorb the virus, we must further assume that the infection caused this additional hereditary change.

b_2. *Hypothesis of acquired immunity—hereditary after infection.* The original variants are predisposed to survival by random physiological variations in size, age, etc. of the bacteria, or maybe even by random variations in the

point of attack of the virus on the bacterium. After survival of such random individuals, however, we must assume that their offspring are hereditarily immune, since they do not even adsorb the virus.

These alternative hypotheses may be grouped by first considering the origin of the hereditary difference. Do the original variants trace back to mutations which occur independently of the virus, such that these bacteria belong to a few clones, or do they represent a random sample of the entire bacterial population? The first alternative may then be subdivided further, according to whether the original variants do or do not interact with the virus. Disregarding for the moment this subdivision, we may formulate two hypotheses:

1. *First hypothesis* (*mutation*): There is a finite probability for any bacterium to mutate during its life time from "sensitive" to "resistant." Every offspring of such a mutant will be resistant, unless reverse mutation occurs. The term "resistant" means here that the bacterium will not be killed if exposed to virus, and the possibility of its interaction with virus is left open.

2. *Second hypothesis* (*acquired hereditary immunity*): There is a small finite probability for any bacterium to survive an attack by the virus. Survival of an infection confers immunity not only to the individual but also to its offspring. The probability of survival in the first instance does not run in clones. If we find that a bacterium survives an attack, we cannot from this information infer that close relatives of it, other than descendants, are likely to survive the attack.

The last statement contains the essential difference between the two hypotheses. On the mutation hypothesis, the mutation to resistance may occur any time prior to the addition of virus. The culture therefore will contain "clones of resistant bacteria" of various sizes, whereas on the hypothesis of acquired immunity the bacteria which survive an attack by the virus will be a random sample of the culture.

For the discussion of the experimental possibility of distinction between these two hypotheses, it is important to keep in mind that the offspring of a *tested* bacterium which survives is resistant on either hypothesis. Repeated tests on a bacterium at different times, or on a bacterium and on its offspring, could therefore give no information of help in deciding the present issue. Thus, one has to resort to less direct methods. Two main differences may be derived from the hypotheses:

First, if the individual cells of a very large number of microcolonies, each containing only a few bacteria, were examined for resistance, a pronounced correlation between the types found in a single colony would be expected on the mutation hypothesis, while a random distribution of resistants would be expected on the hypothesis of acquired hereditary immunity. This experiment, however, is not practicable, both on account of the difficulty of manipulation and on account of the small proportion of resistant bacteria.

Second, on the hypothesis of resistance due to mutation, the proportion of resistant bacteria should increase with time, in a growing culture, as new mutants constantly add to their ranks.

In contrast to this increase in the proportion of resistants on the mutation hypothesis, a constant proportion of resistants may be expected on the hypothesis of acquired hereditary immunity, as long as the physiological conditions of the culture do not change. To test this point, accurate determinations of the proportion of resistant bacteria in a growing culture and in successive subcultures are required. In the attempt to determine accurately the proportion of resistant bacteria, great variations of the proportions were found, and results did not seem to be reproducible from day to day.

Eventually, it was realized that these fluctuations were not due to any uncontrolled conditions of our experiments, but that, on the contrary, large fluctuations are a necessary consequence of the mutation hypothesis and that the quantitative study of the fluctuations may serve to test the hypothesis.

The present paper will be concerned with the theoretical analysis of the probability distribution of the number of resistant bacteria to be expected on either hypothesis and with experiments from which this distribution may be inferred.

While the theory is here applied to a very special case, it will be apparent that the problem is a general one, encountered in any case of mutation in uniparental populations. It is the belief of the authors that the quantitative study of bacterial variation, which until now has made such little progress, has been hampered by the apparent lack of reproducibility of results, which, as we shall show, lies in the very nature of the problem and is an essential element for its analysis. It is our hope that this study may encourage the resumption of quantitative work on other problems of bacterial variation.

THEORY

The aim of the theory is the analysis of the probability distributions of the number of resistant bacteria to be expected on the hypothesis of acquired immunity and on the hypothesis of mutation.

The basic assumption of the hypothesis of acquired hereditary immunity is the assumption of a fixed small chance for each bacterium to survive an attack by the virus. In this case we may therefore expect a binomial distribution of the number of resistant bacteria, or, in cases where the chance of survival is small, a Poisson distribution.

The basic assumption of the mutation hypothesis is the assumption of a fixed small chance *per time unit* for each bacterium to undergo a mutation to resistance. The assumption of a fixed chance per time unit is reasonable only for bacteria in an identical state. Actually the chance may vary in some manner during the life cycle of each bacterium and may also vary when the physiological conditions of the culture vary, particularly when growth slows down on account of crowding of the culture. With regard to the first of these variations, the assumed chance represents the average chance per time unit, averaged over the life cycle of a bacterium. With regard to the second variation, it seems reasonable to assume that the chance is proportional to the growth rate of the bacteria. We will then obtain the same results as on the simple assump-

tion of a fixed chance per time unit, *if we agree to measure time in units of division cycles of the bacteria,* or any proportional unit.

We shall choose as time unit the average division time of the bacteria, divided by ln 2, so that the number N_t of bacteria in a growing culture as function of time t follows the equations

$$(1) \qquad dN_t/dt = N_t, \quad \text{and} \quad N_t = N_0 e^t.$$

We may then define the chance of mutation for each bacterium during the time element dt as

$$(2) \qquad adt,$$

so that a is the chance of mutation per bacterium per time unit, or the "mutation rate."

If a bacterium is capable of different mutations, each of which results in resistance, the mutation rate here considered will be the sum of the mutation rates associated with each of the different mutations.

The number dm of mutations which occur in a growing culture during a time interval dt is then equal to this chance (2) multiplied by the number of bacteria,[4] or

$$(3) \qquad dm = adtN_t,$$

and from this equation the number m of mutations which occur during any finite time interval may be found by integration to be

$$(4) \qquad m = a(N_t - N_0)$$

or, in words, to be equal to the chance of mutation per bacterium per time unit multiplied by the increase in the number of bacteria.

The bacteria which mutate during any time element dt form a random sample of the bacteria present at that time. For small mutation rates, their number will therefore be distributed according to Poisson's law. Since the mutations occuring in different time intervals are quite independent from each other, the distribution of all mutations will also be according to Poisson's law.

This prediction cannot be verified directly, because what we observe, when we count the number of resistant bacteria in a culture, is not the number of mutations which have occurred, but the number of resistant bacteria which have arisen by multiplication of those which mutated, the amount of multiplication depending on how far back the mutation occurred.

If, however, the premise of the mutation hypothesis can be proved by other means, the prediction of a Poisson distribution of the number of mutations

[4] We assume that the number of resistant bacteria is at all times small in comparison with the total number of bacteria. If this condition is not fulfilled, the total number of bacteria in this equation has to be replaced by the number of sensitive bacteria. The subsequent theoretical developments will then become a little more complicated. For the case studied in the experimental part of this paper the condition is fulfilled.

may be used to determine the mutation rate. It is only necessary to determine the fraction of cultures showing no mutation in a large series of similar cultures. This fraction p_0, according to theory, should be:

$$p_0 = e^{-m}. \tag{5}$$

From this equation the average number m of mutations may be calculated, and hence the mutation rate a from equation (4).

Let us now turn to the discussion of the distribution of the number of resistant bacteria.

The average number of resistant bacteria is easily obtained by noting that this number increases on two accounts—namely, first on account of new mutations, second on account of the growth of resistant bacteria from previous mutations. During a time element dt the increase on the first account will be, by equation (3): $adtN_t$. N_t, the number of bacteria present at time t, is given by equation (1). The increase on the second account will depend on the growth rate of the resistant bacteria. In the simple case, which we shall treat here, this growth rate is the same as that of the sensitive bacteria, and the increment on this account is $\rho\ dt$, where ρ is the average number of resistant bacteria present at time t. We have then as the total rate of increase of the average number of resistant bacteria $d\rho/dt = aN_t + \rho$ and upon integration

$$\rho = taN_t \tag{6}$$

if we assume that at time zero the culture contained no resistant bacteria.

It will be seen that the average number of resistant bacteria increases more rapidly than the total number of bacteria. Indeed the fraction of resistant bacteria in the culture increases proportionally to time. This, as pointed out in the introduction, is a distinguishing feature of the mutation hypothesis but unfortunately, as will be seen in the sequel, is not susceptible to experimental verification due to statistical fluctuations.

The resistant bacteria in any culture may be grouped, for the purpose of this analysis, into clones, taking together all those which derive from the same mutation. We may say that the culture contains clones of various age and size, calling "age" of a clone the time since its parent mutation occurred and "size" of a clone the number of bacteria in a clone at the time of observation. It is clear that size and age of a clone determine each other. If, in particular, we make the simplifying hypothesis that the resistant bacteria grow as fast as the normal sensitive strain, the relation between size and age will be expressed by equation (1), with appropriate meaning given to the symbols.

The relation implies that the size of a clone increases exponentially with its age. On the other hand, the frequency with which clones of different ages may be encountered in any culture must decrease exponentially with age, according to equations (3) and (1).

Combining these two results—namely, that clone size increases exponentially with clone age and that frequency of clones of different age decreases exponentially with clone age—we see that the two factors cancel when the

average number of bacteria belonging to clones of one age group is considered. In other words, at the time of observation we shall have, *on the average*, as many resistant bacteria stemming from mutations which occurred during the first generation after the culture was started as stemming from mutations which occurred during the last generation before observation, or during any other single generation.

On the other hand, for small mutation rates it is very improbable that any mutation will occur during the early generations of a single or of a limited number of experimental cultures. It follows that the average number of resistant bacteria derived from a limited number of experimental cultures will, probably, be considerably smaller than the theoretical value given by equation (6), and, improbably, the experimental value will be much larger than the theoretical value. The situation is similar to the operation of a (fair) slot machine, where the average return from a limited number of plays is probably considerably less than the input, and improbably, when the jackpot is hit, the return is much bigger than the input.

This result characterizes the distribution of the number of resistant bacteria as a distribution with a long and significant tail of rare cases of high numbers of resistant bacteria, and therefore as *a distribution with an abnormally high variance.* This variance will be calculated below.

For such distributions the averages derived from limited numbers of samples yield very poor estimates of the true averages. Somewhat better estimates of the averages may in such cases be obtained by omitting, in the calculation of the theoretical averages, the contribution to these averages of those events which probably will not occur in any of our limited number of samples. We may do this, in the integration leading to equation (6), by putting the lower limit of integration not at time zero, when the cultures were started, but at a certain time t_0, prior to which mutations were not likely to occur in any of our experimental cultures. We then obtain as a *likely average* r of the number of resistant bacteria in a limited number of samples, instead of equation (6),

$$r = (t - t_0)aN_t. \tag{6a}$$

It now remains to choose an appropriate value for the time interval $t - t_0$.

For this purpose we return to equation (4), in which it was stated that the average number of mutations which occur in a culture is equal to the mutation rate multiplied by the increase of the number of bacteria. Let us then choose t_0 such that up to that time just one mutation occurred, on the average, in a group of C similar cultures, or

$$1 = aC(N_{t_0} - N_0).$$

In this equation we may neglect N_0, the number of bacteria in each inoculum, in comparison with N_{t_0}, the number of bacteria in each culture at the critical time t_0. We may also express N_{t_0} in terms of N_t, the number of bacteria at the time of observation, applying equation (1):

$$N_{t_0} = N_t e^{-(t-t_0)}.$$

We thus obtain

$$\hat{t} - t_0 = \ln(N_t Ca). \tag{7}$$

Equations (6a) and (7) may be combined to eliminate $t - t_0$ and to yield a relation between the observable quantities r and N_t on the one hand and the mutation rate a on the other hand, to be determined by this equation:

$$r = aN_t \ln (N_t Ca). \tag{8}$$

This simple transcendental equation determining a may be solved by any standard numerical method. In figure 1, the relation between r and aN_t is plotted for several values of C.

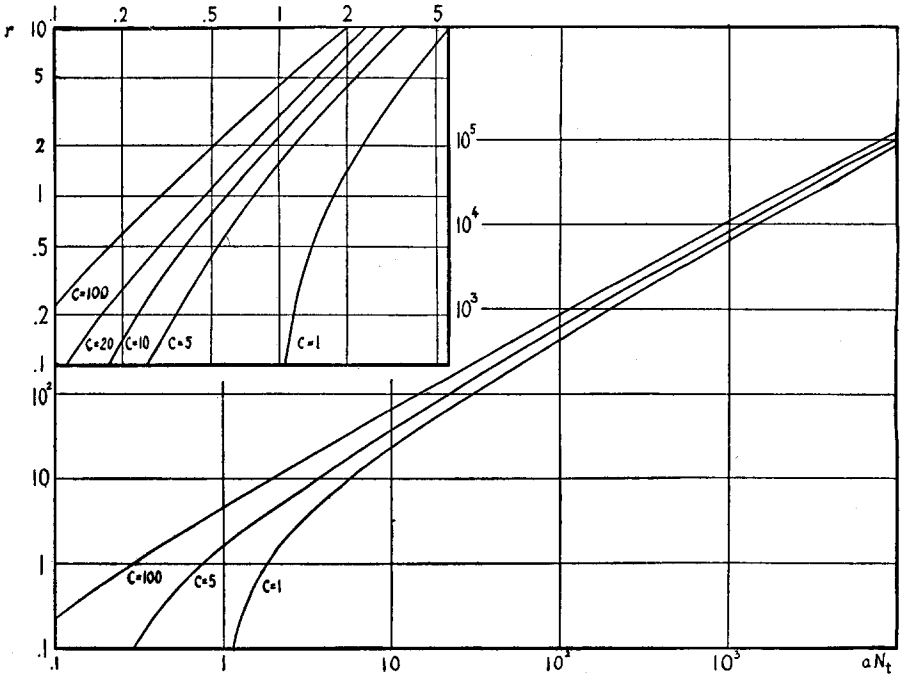


FIGURE 1.—The value of aN_t as a function of r for various values of C. The upper left hand part of the figure gives the curves for low values of aN_t and of r on a larger scale. See text.

Estimates of a obtained from equation (8) will be too high if in any of the experimental cultures a mutation happened to occur prior to time t_0. From the definition of t_0 it will be seen that this can be expected to happen in little more than half of the cases.

While we have thus obtained a relation permitting an estimate of the mutation rate from the observation of a limited number of cultures, this relation is in no way a test of the correctness of the underlying assumptions and, in particular, is not a test of the mutation hypothesis itself. In order to find such tests of the correctness of the assumption we must derive further quantitative relations concerning the distribution of the number of resistant bacteria and compare them with experimental results.

Since we have seen that the mutation hypothesis, in contrast to the hypothesis of acquired immunity, predicts a distribution of the number of resistant bacteria with a long tail of high numbers of resistant bacteria, the determination of the *variance* of the distribution should be helpful in differentiating between the two hypotheses. We may here again determine first the true variance—that is, the variance of the complete distribution—and second the likely variance in a limited number of cultures, by omitting those cases which are not likely to occur in a limited number of cultures.

The variance may be calculated in a simple manner by considering separately the variances of the partial distributions of resistant bacteria, each partial distribution comprising the resistant bacteria belonging to clones of one age group. The distribution of the total number of resistant bacteria is the resultant of the superposition of these independent partial distributions.

Each partial distribution is due to the mutations which occurred during a certain time interval $d\tau$, extending from $(t-\tau)$ to $(t-\tau+d\tau)$. The average number of mutations which occurred during this interval is, according to equation (3),

$$dm = aN_\tau d\tau = aN_t e^{-\tau} d\tau. \tag{9}$$

These mutations will be distributed according to Poisson's law, so that the variance of each of these distributions is equal to the mean of the distribution. We are however not interested in the distribution of the number of mutations but in the distribution of the number of resistant bacteria which stem from these mutations at the time of observation—that is, after the time interval τ. Each original mutant has then grown into a clone of size e^τ. The distribution of the resistant bacteria stemming from mutations occurred in the time interval $d\tau$ has therefore an average value which is e^τ times greater than the average number of mutations, and a variance which is $e^{2\tau}$ times greater than the variance of the number of mutations. Thus we find for the average number of resistant bacteria:

$$d\rho = aN_t d\tau,$$

and for the variance of this number

$$\mathrm{var}_{d\rho} = aN_t e^\tau d\tau.$$

From this variance of the partial distribution, the variance of the distribution of all resistant bacteria may be found simply by integrating over the appropriate time interval—that is, either from time t to time o (τ from o to t), if the true variance is wanted, or from time t to time t_0 (τ from o to $t-t_0$), if the likely variance in a limited number of cultures is wanted. In the first case we obtain:

$$\mathrm{var}_\rho = aN_t(e^t - 1). \tag{10}$$

In the second case we obtain:

$$\mathrm{var}_r = aN_t\left[e^{(t-t_0)} - 1\right]. \tag{10a}$$

Substituting here the previously found value of $(t-t_0)$ and neglecting the second term in the brackets, we obtain:

$$(11) \qquad var_r = Ca^2N_t^2.$$

Comparing this value of the likely variance with the value of the likely average, from equation (8), we see that the ratio of the standard deviation to the average is:

$$(12) \qquad \sqrt{var_r}/r = \sqrt{C}/\ln (N_tCa).$$

It is seen that this ratio depends on the logarithm of the mutation rate and will consequently be only a little smaller for mutation rates many thousand times greater than those considered in the experiments reported in this paper.

In the beginning of this theoretical discussion we pointed out that the hypothesis of acquired immunity leads to the prediction of a distribution of the number of resistant bacteria according to Poisson's law, and therefore to the prediction of a variance equal to the average. On the other hand, if we compare the average, equation (8), with the variance, equation (11), (not, as above, with the square root of the variance), we obtain

$$(12a) \qquad var_r = rN_tCa/\ln (N_tCa).$$

Equation (12a) shows that the likely ratio between variance and average is much greater than unity on the hypothesis of mutation, if (N_tCa), the total number of mutations which occurred in our cultures, is large compared to unity.[5]

It is possible to carry the analysis still further and to evaluate the higher moments of the distribution function of the number of resistant bacteria, or even the distribution function itself. The moments are comparatively easy to obtain, while the calculation of the distribution function involves considerable

[5] In some of the experiments reported in the present paper we did not determine the total number of resistant bacteria in each culture, but the number contained in a small sample from each culture. In these cases the variance of the distribution of the number of resistant bacteria will be slightly increased by the sampling error. The proper procedure is here first to find the average number of resistant bacteria per culture by multiplying the average per sample by the ratio

$$(13) \qquad \frac{\text{volume of culture}}{\text{volume of sample}};$$

second, to evaluate the mutation rate with the help of equation (8); third, to figure the likely variance for the cultures by equation (11); fourth, to divide this variance by the square of the ratio (13) to obtain that part of the variance in the samples which is due to the chance distribution of the mutations. The experimental variance should be greater than this value, on account of the sampling variance. The sampling variance is in all our cases only a small correction to the total variance, and it is sufficient to use its upper limit, that of the Poisson distribution, in our calculations. Consequently, when comparing the experimental with the calculated values, we first subtract from the experimental value the sampling variance, which we take to be equal to the average number of resistant bacteria.

mathematical difficulties. An approximation to the beginning of the distribution function—that is, to its values for small numbers of resistant bacteria—may be obtained by grouping mutations according to the bacterial generation during which they occurred. For instance, the probability of obtaining seven resistant bacteria may be broken down into the sum of the following alternative events: (a) seven mutations during the last generation; (b) three mutations during the last generation and two mutations one generation back; (c) three mutations during the last generation and one mutation two generations back; (d) one mutation during the last generation and three mutations one generation back; (e) one mutation during the last generation, one mutation one generation back and one mutation two generations back.

The probability of each of these events depends only on the mutation rate and on the final number of bacteria.

The grouping of mutations according to the bacterial generation during which they occurred, and the assumption that the bacteria increase in simple geometric progression, simplify the calculation sufficiently to permit numerical computation. On the other hand, the classes with two, four, eight, etc., mutants are artificially favored by this procedure, so that a somewhat uneven distribution results, with too high values for two, four, eight, etc., resistant bacteria (see fig. 2).

MATERIAL AND METHODS

The material used for our experimental study consisted of a bacterial virus α and of its host, *Escherichia coli* B (DELBRÜCK and LURIA 1942). Secondary cultures after apparently complete lysis of B by virus α show up within a few hours from the time of clearing. They consist of cells which are resistant to the action of virus α, but sensitive to a series of other viruses active on B. The resistant cells breed true and can be established easily as pure cultures. No trace of virus could be found in any pure culture of the resistant bacteria studied in this paper. The resistant strains are therefore to be considered as non-lysogenic.

Tests were made to see whether the resistance to virus α was a stable character of the resistant strains. In the first place, it was found that virus α is not appreciably adsorbed by any of the resistant strains. In the second place, when a certain amount of virus α is mixed with a growing culture of a resistant strain, no measurable increase of the titer of virus α occurs over a period of several hours. This is a very sensitive test for the occurrence of sensitive bacteria, and its negative result for all resistant strains shows that reversion to sensitivity must be a very rare event.

Morphologically at least two types of colonies of resistant bacteria may be distinguished. The first type of colony is similar to the type produced by the sensitive strain both in size and in the character of the surface and of the edge. The second type of colony is much smaller and translucent. The difference in colony type is maintained in subcultures. Microscopically the bacteria from these two types of colonies are indistinguishable. They also do not differ

from each other or from the sensitive strain in their fermentation reactions on common sugars and in the characteristics of their growth curves in nutrient broth. In particular, the lag periods, the division times during the logarithmic phase of growth, and the maximum titers attained are identical for the sensitive strain and for the two variants. Both variants, therefore, fulfill the requirements for the applicability of the theory developed above.

In the presentation of our experimental results we have lumped the counts of the two types of colonies together, because: (1) theoretically, this is equivalent to summing the corresponding mutation rates; (2) experimentally, we are not certain whether each of these types does not actually comprise a diversity of variants; (3) experimentally, no correlation appeared to exist between the occurrence of these variants, which shows the independence of the causes of their occurrence.

Cultures of B were grown either in nutrient broth (containing .5 percent NaCl) or in an asparagin-glucose synthetic medium. In the latter, the division time during the logarithmic phase of growth was 35 minutes, as compared with 19 minutes in broth. In synthetic medium, the acidity increased during the time of incubation from pH 7 to pH 5.

In cultures of strain B, between 10^{-8} and 10^{-6} of the bacteria are found usually to give colonies resistant to the action of virus α when samples of such cultures are plated with large amounts of virus. In order to be reasonably certain that the resistant bacteria found in the test had not been introduced into the test culture with the initial inoculum, the test cultures were always started with very small inocula, containing between 50 and 500 bacteria from a growing culture. Thus any resistant bacterium found at the moment of testing (when the culture contains between 10^8 and 5×10^9 bacteria/cc) must be an offspring of one of the sensitive bacteria of the inoculum.

All platings were made on nutrient agar plates. The plating experiments for counting the number of resistant bacteria in a liquid culture of the sensitive strain were done by plating either a portion or the entire culture with a large amount of virus α. The virus was plated first, and spread over the entire surface of the agar. A few minutes later the bacterial suspension to be tested was spread over the central part of the plate, leaving a margin of at least one centimeter. Thus all bacteria were surrounded by large numbers of virus particles.

Microscopic examination of plates seeded in this manner showed that lysis takes place very quickly; only bacteria which at the time of plating were in the process of division may sometimes complete the division. The resistant colonies which appear after incubation are therefore due to resistant bacterial cells present at the time of plating.

The total number of bacteria present in the culture to be tested was determined by colony counts in the usual manner.

The resistant colonies of the large type appear after 12–16 hours of incubation, the colonies of the small type appear after 18–24 hours, and never reach half the size of the former ones. Counts were usually made after 24 and 48 hours.

EXPERIMENTAL

A Test of the Reliability of the Plating Method

In our experiments we wanted to study the fluctuations of the numbers of resistant bacteria found in cultures of sensitive bacteria. It was therefore necessary to show first that the method of testing did not involve any unrecognized variables, which caused the number of resistant colonies to vary from plate to plate or from sample to sample.

Therefore, parallel platings were made using a series of samples from the same bacterial culture. If our plating method is reliable, fluctuations should in this arrangement be due to random sampling only, and the variance from a series of such samples should be equal to the mean.

Table 1 gives the results of three such experiments. It will be seen that in

TABLE 1

The number of resistant bacteria in different samples from the same culture.

SAMPLE NO.	EXP. NO. 10a RESISTANT COLONIES	EXP. NO. 11a RESISTANT COLONIES	EXP. NO. 3 RESISTANT COLONIES
1	14	46	4
2	15	56	2
3	13	52	2
4	21	48	1
5	15	65	5
6	14	44	2
7	26	49	4
8	16	51	2
9	20	56	4
10	13	47	7
mean	16.7	51.4	3.3
variance	15	27	3.8
χ^2	9	5.3	12
P	.4	.8	.2

all three cases variance and mean agree as well as may be expected. There is therefore no reason to assume that the method of sampling or plating introduces any fluctuations into our results besides the sampling error.

Fluctuations of the Number of Resistant Bacteria in Samples from a Series of Similar Cultures

As pointed out in the introduction and in the theoretical part, the hypothesis of acquired immunity and the hypothesis of mutation lead to radically different predictions regarding the distribution of the number of resistant bacteria in a series of similar cultures. The hypothesis of acquired immunity predicts a variance equal to the average, as in sampling, while the mutation hypothesis predicts a much greater variance.

Series of five to 100 cultures were set up in parallel with small equal inocula, and were grown until maximum titer was reached. Three kinds of cultures

were used—namely: (1) 10.0 cc aerated broth cultures; (2) .2 cc broth cultures; (3) .2 cc synthetic medium cultures.

The results of all tests for the number of resistant bacteria are summarized in table 2 and table 3.

TABLE 2

The number of resistant bacteria in series of similar cultures.

EXPERIMENT NO.	1	10	11	15	16	17	21a	21b
Number of cultures	9	8	10	10	20	12	19	5
Volume of cultures, cc	10.0	10.0	10.0	10.0	.2*	.2*	.2	10.0
Volume of samples, cc	.05	.05	.05	.05	.08	.08	.05	.05
Culture No.								
1	10	29	30	6	1	1	0	38
2	18	41	10	5	0	0	0	28
3	125	17	40	10	3	0	0	35
4	10	20	45	8	0	7	0	107
5	14	31	183	24	0	0	8	13
6	27	30	12	13	5	303	1	
7	3	7	173	165	0	0	0	
8	17	17	23	15	5	0	1	
9	17		57	6	0	3	0	
10			51	10	6	48	15	
11					107	1	0	
12					0	4	0	
13					0		19	
14					0		0	
15					1		0	
16					0		17	
17					0		11	
18					64		0	
19					0		0	
20					35			
Average per sample	26.8	23.8	62	26.2	11.35	30	3.8	48.2
Variance (corrected for sampling)	1217	84	3498	2178	694	6620	40.8	1171
Average per culture	5360	4760	12400	5240	28.4	75	15.1	8440
Bacteria per culture	3.4×10^{10}	4×10^{10}	4×10^{10}	2.9×10^{10}	5.6×10^{8}	5×10^{8}	1.1×10^{8}	3.2×10^{10}
Mutation rate	1.8×10^{-8}	1.4×10^{-8}	4.1×10^{-8}	2.1×10^{-8}	1.1×10^{-6}	3.0×10^{-8}	3.3×10^{-8}	3.0×10^{-8}
Standard deviation / Average exp.	1.3	.39	.95	1.8	2.3	2.7	1.7	.71
Standard deviation / Average calc.	.35	.33	.33	.37	.94	.67	1.04	.26

* Cultures in synthetic medium.

It will be seen that in every experiment the fluctuation of the numbers of resistant bacteria is tremendously higher than could be accounted for by the sampling errors, in striking contrast to the results of plating from the same culture (see table 1) and in conflict with the expectations from the hypothesis of acquired immunity.

We want to see next whether these results fit the expectations from the hypothesis of mutation. We must therefore compare the experimental results with the relations developed in the theoretical part, keeping in mind that the theory contains several simplifying assumptions.

First we can compare, according to equation (12), the experimental and the calculated values of the ratio between the standard deviation and the average of the numbers of resistant bacteria. These ratios are included in tables 2 and 3. It is seen that the experimental and theoretical values are reasonably close.

However, in all but one case the experimental ratio is greater than the value calculated from the theory—that is, the variability is even greater than predicted.

TABLE 3

Distribution of the numbers of resistant bacteria in series of similar cultures.

EXPERIMENT NO.	22		23	
Number of cultures	100		87	
Volume of cultures, cc	.2*		.2*	
Volume of samples, cc	.05		.2	
	Resistant bacteria	*Number of cultures*	*Resistant bacteria*	*Number of cultures*
	0	57	0	29
	1	20	1	17
	2	5	2	4
	3	2	3	3
	4	3	4	3
	5	1	5	2
	6– 10	7	6– 10	5
	11– 20	2	11– 20	6
	21– 50	2	21– 50	7
	51– 100	0	51– 100	5
	101– 200	0	101– 200	2
	201– 500	0	201– 500	4
	501–1000	1	501–1000	0
Average per sample	10.12		28.6	
Variance (corrected for sampling)	6270		6431	
Average per culture	40.48		28.6	
Bacteria per culture	2.8×10^{8}		2.4×10^{8}	
Mutation rate	2.3×10^{-8}		2.37×10^{-8}	
Standard deviation / Average (exp.)	7.8		2.8	
Standard deviation / Average (calc.)	1.5		1.5	

* Cultures in synthetic medium.

A part of this discrepancy may be accounted for by the fact that the time t_0, mutations occurring prior to which were disregarded by the theory, was chosen in such a manner that on the average one mutation would occur prior to time t_0. This mutation, if it occurs, will of course tend to increase the variance, and in some of the experiments the high value of the experimental variance can be traced directly to one exceptional culture in which a mutation had evidently occurred several generations prior to time t_0. Unfortunately, there is no general criterion by which one might eliminate such cultures from the statistical analysis, because, in a culture with an exceptionally high count of resistant bacteria, these do not necessarily stem from one exceptionally early mutation, but may also be due to an exceptionally large number of mutations after time t_0.

There may also be other reasons why the observed variances are higher than the expected ones. First of all, the simplifying assumption that the mutation

rate per bacterial generation is independent of the physiological state of the bacteria may be too simple. If the mutation rate is higher for actively growing bacteria than for bacteria near the saturation limit of the cultures, early mutations and big clone sizes will be favored, and therefore higher variations of the numbers of resistant bacteria can be expected. Second, the assumption of a sudden transition from sensitivity to resistance may also be too simple. It is conceivable that the character "resistance to virus" may not fully develop in the bacterial cell in which the mutation occurs, but only in its offspring, after one or more generations. However, if this were the case, cultures with only one or two resistant bacteria should be relatively rare. The last experiment listed in table 3, in which the entire cultures were plated, shows a rather high proportion of cultures with only one resistant bacterium. This seems to show that the

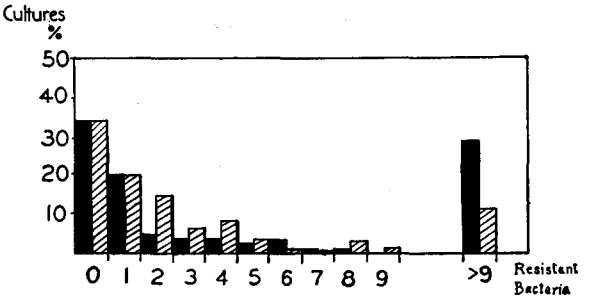


FIGURE 2.—Experimental (Experiment No. 23) and calculated distributions of the numbers of resistant bacteria in a series of similar cultures. Solid columns: experimental. Cross-hatched columns: calculated.

character "resistance to virus" in general does come to expression in the bacterial cell in which the corresponding mutation occurred, as assumed by the theory.

Another way of comparing the experimental results with the theory is to compare the experimental distribution of resistant bacteria with the approximate distribution calculated by the method outlined at the end of the theoretical part. The theoretical distribution has to be calculated from the average number of mutations per culture given by equation (5). Only experiments where the whole culture is tested can therefore be used for such a comparison. This method tests the fitting of the expectations for small numbers of resistant bacteria, in contrast to the comparison of the standard deviations, which involves predominantly the cultures with high numbers of resistant bacteria.

Figure 2 shows the experimental and calculated distributions for Experiment No. 23; the cultures with more than nine resistant bacteria are lumped together in one class, since the distribution has not been calculated for values higher than nine.

It is seen that the fitting for small values is satisfactory. In particular, the

number of cultures with one resistant bacterium very closely fits the expectation. The classes with two, four, eight, etc., resistant bacteria are bound to be favored in the theoretical distribution, as explained in the theoretical part.

The results shown in figure 2 also confirm the assumption that the discrepancy between experimental and calculated standard deviations must be due to an excess of cultures with large numbers of resistant bacteria.

Summing up the evidence, we may say that the experiments show clearly that the resistant bacteria appear in similar cultures not as random samples but in groups of varying sizes, indicating a correlating cause for such grouping, and that the assumption of genetic relatedness of the bacteria of such groups offers the simplest explanation for them.

Mutation Rate

As pointed out in the theoretical part of this paper, mutation rates may be estimated from the experiments by two essentially different methods. The first method makes use of the fact that the number of mutations in a series of similar cultures should be distributed in accordance with Poisson's law; the average number of mutations per culture is calculated from the proportion of cultures containing no resistant bacteria at the moment of the test, according to equation (5).

There are two technical difficulties involved in the application of this method. In the first place, rather large numbers of cultures have to be handled and conditions have to be chosen so that the proportion of resistant bacteria is neither too small nor too large. In the second place, the entire cultures have to be tested, which means, in our method of testing, that cultures of rather small volume have to be used and great care must be taken to plate as nearly as possible the entire culture.

Experiment No. 23 (see table 3) permits an estimate of the mutation rate by this method. Out of 87 cultures, no resistant bacteria were found in 29 cultures, a proportion of .33. From equation (5) we calculate therefore that the average number of mutations per culture in this experiment was 1.10. Since the total number of bacteria per culture was 2.4×10^8, we obtain as the mutation rate, from equation (4),

$$a = .47 \times 10^{-8} \text{ mutations per bacterium per time unit}$$

$$= .32 \times 10^{-8} \text{ mutations per bacterium per division cycle.}$$

This calculation makes use exclusively of the proportion of cultures containing no resistant bacteria. It is therefore inefficient in its use of the information gathered in the experiment.

The second method makes use of the average number of resistant bacteria per culture. The relation of this average number with the mutation rate was discussed in the theoretical part of this paper and was found to be expressed by equation (8). The mutation rates calculated by this method for each experiment are collected in table 4.

TABLE 4

Values of mutation rate from different experiments.

EXPERIMENT NO.	NUMBER OF CULTURES	VOLUME OF CULTURES	MUTATION RATE
		cc	*Mutations per bacterium per time unit*
1	9	10.0	1.8×10^{-8}
10	8	10.0	1.4×10^{-8}
11	10	10.0	4.1×10^{-8}
15	10	10.0	2.1×10^{-8}
16	20	.2*	1.1×10^{-8}
17	12	.2*	3.0×10^{-8}
21a	19	.2	3.3×10^{-8}
21b	5	10.0	3.0×10^{-8}
22	100	.2*	2.3×10^{-8}
23	87	.2*	2.4×10^{-8}
Average			2.45×10^{-8}

* Cultures in synthetic medium.

It will be seen that the values of the mutation rate obtained by the second method are all higher than the value found by the first method. This discrepancy may be traced back to the same cause as the discrepancy between the calculated and observed values of the standard deviation of the numbers of resistant bacteria. This, we found, was due to an excess of early mutations, giving rise to big clones of resistant bacteria. These big clones do not affect the mutation rate calculated by the first method, but they do affect the results of the second method, which is based on the average number of resistant bacteria.

One sees in table 4 that the mutation rate calculated by the second method does not vary greatly from experiment to experiment. In particular, it will be noted that there is no significant difference between the values obtained from cultures in broth and from cultures in synthetic medium, notwithstanding the considerable difference of metabolic activity and of growth rate of the bacteria in these two media. This shows that the simple assumption of a fixed small chance of mutation per physiological time unit is vindicated by the results. It may also be noted in table 4 that there is no significant difference between the mutation rates obtained from 10 cc cultures and those obtained from .2 cc cultures, or between the experiments with many and those with few cultures. The variability of the value of the mutation rate seems to be solely due to the peculiar probability distribution of the number of resistant bacteria in series of similar cultures predicted by the mutation theory.

At this point an experiment may be mentioned by which it was desired to find out whether or not mutations occur in a culture after the bacteria have ceased growing. A culture was grown to saturation and was then tested repeatedly for resistant bacteria and for total number of bacteria over several

days. The proportion of resistant bacteria did not change, even when the sensitive bacteria began to die, showing that the resistant bacteria have the same death rate in aging cultures as the sensitive bacteria.

DISCUSSION

We consider the above results as proof that in our case the resistance to virus is due to a heritable change of the bacterial cell which occurs independently of the action of the virus. It remains to be seen whether or not this is the general rule. There is reason to suspect that the mechanism is more complex in cases where the resistant culture develops only several days after lysis of the sensitive bacteria.

The proportion of mutant organisms in a culture and the mutation rate are far smaller in our case than in other studied cases of heritable bacterial variation. The possibility of investigation of such rare mutations is in our case merely the result of the method of detecting the mutant organisms. In other cases, the variants are detected by changes in the colony type which is produced by the mutant organism, either in the pigmentation or in the character of the surface or the edge of the colony. Often, colonies of intermediate character occur, and it is difficult to decide whether they are mixed colonies or stem from bacteria with intermediate character. This is particularly true of cases where the mutation rate is high and where reverse mutation occurs. Fairly high mutation rates, however, are a prerequisite of any study of colony variants, since the number of colonies that can be examined is limited by practical reasons.

The study of mutations causing virus resistance is free of these difficulties. The segregation of the mutant from the normal organisms occurs in the one-cell stage by elimination of the normal individuals, and the character of the colony which develops from a mutant organism is of secondary importance. Owing to the total elimination of the normal individuals, the number of organisms which may be examined is very much higher than for any other method; more than 10^8 bacteria may be tested on a single plate. Since the mutations to virus resistance are often associated with other significant characters, the method may well assume importance with regard to the general problems of bacterial variation.

It must not be supposed that the peculiar statistical difficulties encountered in our case are restricted to cases of very low mutation rates. The essential condition for the occurrence of the peculiar distribution studied in the theoretical part of this paper is the following: *the initial number of bacteria in a culture must be so small that the number of mutations which occur during the first division cycle of the bacteria is a small number.* This will always be true, however great the mutation rate, if one studies cultures containing initially a small number of organisms.

In a series of very interesting studies of the color variants of *Serratia marcescens*, BUNTING (1940a, 1940b, 1942; BUNTING and INGRAHAM 1942) succeeded to some extent in obviating the statistical difficulties by always using

inocula of about 100,000 bacteria. In some of her cases this number was sufficiently high to result in numerous mutations during the first division cycle of the bacteria. In other cases the number was apparently not high enough, since the author reports troublesome variations of the fractions of variants in successive subcultures. In those cases where the size of the inocula was high enough, the author succeeded in deriving reproducible values for the mutation rates from the study of single cultures, followed through numerous subcultures. In these cases it is sufficient to apply the equations of the theory referring to the *average* numbers of mutants as a function of time. It is clear, however, that this method is applicable only in cases of mutation rates of at least 10^{-4} per bacterium per division cycle.

In our case, as in many others, the virus resistant variants do not exhibit any striking correlated physiological changes. There is therefore little opportunity for an inquiry into the nature of the physiological changes responsible for the resistance to virus. Since the offspring of the mutant bacteria, when isolated after the test, are unable to synthesize the surface elements to which the virus is specifically adsorbed in the sensitive strain, one might suppose that this loss is a direct effect of the mutation. However, it is also conceivable that the loss occurs upon contact with virus, since it is detected only after such contact (hypothesis b_1). In some of the cases studied by BURNET (1929), where the mutational change to resistance is correlated with a change of phase, from smooth to rough or vice versa, the change of the surface structure must be a direct result of the mutation, since the mutant colonies may be picked up prior to the resistance test and, when tested, exhibit the typical change of affinity of the surface structure. These findings make it more probable that the loss of surface affinity to virus is a direct effect of the mutation.

The alteration of specific surface structures due to genetic change is a phenomenon of the widest occurrence. The genetic factors determining the antigenic properties of erythrocytes are well known. There is evidence (WEBSTER 1937; HOLMES 1938; STEVENSON, SCHULTZ, and CLARK 1939) that resistance or sensitivity to virus in plants and animals is correlated with, or even dependent on, genetic changes, possibly affecting the antigenic make-up of the cellular surface. The proof that resistance to a bacterial virus may be traced to a specific genetic change may assume importance, therefore, with regard to the general problems of virus sensitivity and virus resistance.

SUMMARY

The distribution of the numbers of virus resistant bacteria in series of similar cultures of a virus-sensitive strain has been analyzed theoretically on the basis of two current hypotheses concerning the origin of the resistant bacteria.

The distribution has been studied experimentally and has been found to conform with the conclusions drawn from the hypothesis that the resistant bacteria arise by mutations of sensitive cells independently of the action of virus.

The mutation rate has been determined experimentally.

LITERATURE CITED

BUNTING, M. I., 1940a A description of some color variants produced by *Serratia marcescens*, strain 274. J. Bact. **40**: 57–68.

1940b The production of stable populations of color variants of *Serratia marcescens* #274 in rapidly growing cultures. J. Bact. **40**: 69–81.

1942 Factors affecting the distribution of color variants in aging broth cultures of *Serratia marcescens* #274. J. Bact. **43**: 593–606.

BUNTING, M. I., and L. J. INGRAHAM, 1942 The distribution of color variants in aging broth cultures of *Serratia marcescens* #274. J. Bact. **43**: 585–591.

BURNET, F. M., 1929 Smooth-rough variation in bacteria in its relation to bacteriophage. J. Path. Bact. **32**: 15–42.

DELBRÜCK, M., and S. E. LURIA, 1942 Interference between bacterial viruses. I. Arch. Biochem. **1**: 111–141.

GRATIA, A., 1921 Studies on the d'Herelle phenomenon. J. Exp. Med. **34**: 115–131.

D'HERELLE, F., 1926 The Bacteriophage and Its Behavior. Baltimore: Williams and Wilkins.

HOLMES, F. O., 1938 Inheritance of resistance to tobacco-mosaic disease in tobacco. Phytopathology **28**: 553–561.

STEVENSON, F. J., E. S. SCHULTZ, and C. F. CLARK, 1939 Inheritance of immunity from virus X (latent mosaic) in the potato. Phytopathology **29**: 362–365.

WEBSTER, L. T., 1937 Inheritance of resistance of mice to enteric bacterial and neurotropic virus infections. J. Exp. Med. **65**: 261–286.

The origin of mutants

John Cairns, Julie Overbaugh & Stephan Miller

Department of Cancer Biology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

Nucleic acids are replicated with conspicuous fidelity. Infrequently, however, they undergo changes in sequence, and this process of change (mutation) generates the variability that allows evolution. As the result of studies of bacterial variation, it is now widely believed that mutations arise continuously and without any consideration for their utility. In this paper, we briefly review the source of this idea and then describe some experiments suggesting that cells may have mechanisms for choosing which mutations will occur.

WHEN populations of single cells are subject to certain forms of strong selection pressure, variants emerge bearing changes in DNA sequence that bring about an appropriate change in phenotype. If the selection is being applied to cells growing in liquid, the population with the original genotype will sooner or later be supplanted by the variants; if the cells are immobilized on a plate, the variants will form colonies or papillae, rising above the rest of the population. Our problem is to determine how many of these variants are arising as a direct and specific response to the selection pressure (would not have occurred in its absence) and how many are 'spontaneous' (would have arisen even in the absence of selection).

Luria and Delbrück were the first to attempt this distinction[1]. They studied the origin of the phage-resistant mutants that are found when cultures of *Escherichia coli* are plated out in the presence of bacteriophage T1. They considered two extreme models. In the first model, the mutants are assumed to arise after plating, in response to the attack by the bacteriophage; the mutations occur in a small fraction of the bacteria exposed to the virus, and the number of mutants found in each of a large group of independent cultures should form a Poisson distribution whose mean is simply the product of the number of bacteria plated and the probability that a bacterium can become resistant before it is killed by the virus. In the second model, the mutants are assumed to arise spontaneously, as the result of a rare error in replication that has a constant probability of occurring in the life of each bacterium; in this case, the mutational events will be distributed randomly throughout each culture's previous history (scattered at random through its family tree), and the final number of mutants should be highly variable because occasional cultures will, by chance, have suffered a mutation early in their growth and will therefore end up containing a very large number of mutants. When put to the test, the number of T1-resistant mutants was observed to vary greatly from culture to culture, with the occasional culture containing up to 100 times the median value for the group as a whole. Luria and Delbrück concluded that mutation to phage resistance is the result of rare spontaneous events, occurring during the prior growth of cultures before they had been exposed to any selection.

By various ingenious sampling procedures such as 'replica plating', which tested cells' siblings rather than the cells themselves[2,3], it subsequently became possible to prepare pure cultures of mutants (for example, streptomycin-resistance mutants) without at any time having exposed the cells or their ancestors to any selection pressure. These experiments constituted the final proof that at least some forms of bacterial mutation are occurring spontaneously, before the bacteria can have any indication of each mutation's possible utility. They had, however, shown only that some mutants do arise spontaneously, in the absence of selection. In fact, we now know that mutations to phage and streptomycin resistance are not expressed until several generations after the change in DNA sequence has occurred[4]; thus the resistant mutants one isolates by spreading cultures on selective plates are the result of events that must have occurred many generations before one applies the selection that demonstrates their existence. So these classical experiments could not have detected (and certainly did not exclude) the existence of a non-random, possibly product-oriented form of mutation.

To look for such a process, we have studied mutations that are immediately expressed, and where the selection pressure rewards mutants by letting them multiply but allows all the other, non-mutant cells to survive so that they can at least have the opportunity to perform directed mutation. For this analysis, it was necessary to calculate the expected composite distribution, where some mutants are arising spontaneously (during the prior growth of each culture) and some are arising later, after the cells are put on selective plates.

Mutant distribution in independent cultures

We have to consider two distributions and their combination. The first distribution is created when mutants arise at random while the population in each culture is increasing by binary fission; at this stage, before any selection has been applied, mutants and non-mutants are postulated to multiply at the same rate; when the cultures are put out on selective plates, the mutants are able to form colonies. In these circumstances, the distribution of mutants among a set of independent cultures will be determined by a single parameter, the constant probability of mutation per cell per generation. Lea and Coulson[5] were the first to derive an exact solution. The distribution has certain awkward properties[6,7] but the following argument leads to a useful approximation for its upper end. For a culture undergoing binary fission, the number of cells (n) present at any particular moment plus the number present in all previous generations (that is, n plus $n/2+n/4+n/8+\cdots$) is twice as great as the number present in all previous generations ($n/2+n/4+n/8+\cdots$); so the chance that any particular mutation occurred at least g generations ago is twice the chance that it occurred at least $g+1$ generations ago. If, therefore, we just consider those very rare cultures that suffered a mutation so early in their history that in each culture one single clone (that is, one single mutational event) accounts for virtually all the mutants finally present, we see that the proportion of such cultures that contain at least 2^g cells will be twice the proportion with at least 2^{g+1}; this describes exactly the upper end of the distribution. In fact, if the cultures finally contain N cells and have suffered an average of m mutational events per culture (the mutation rate is $m/2N$ per cell), the probability (P_j) of finding a 'jackpot' of at least j mutants approaches m/j. This approximation had previously been derived for the case where m is very small[8], but actually it holds for the upper end of the distribution irrespective of the value of m. Figure 1*a* shows the exact relation between log x and log P_x (the logarithm of the proportion of cultures with x or more mutants) for several values of m.

A very different distribution will be created if, instead, all the

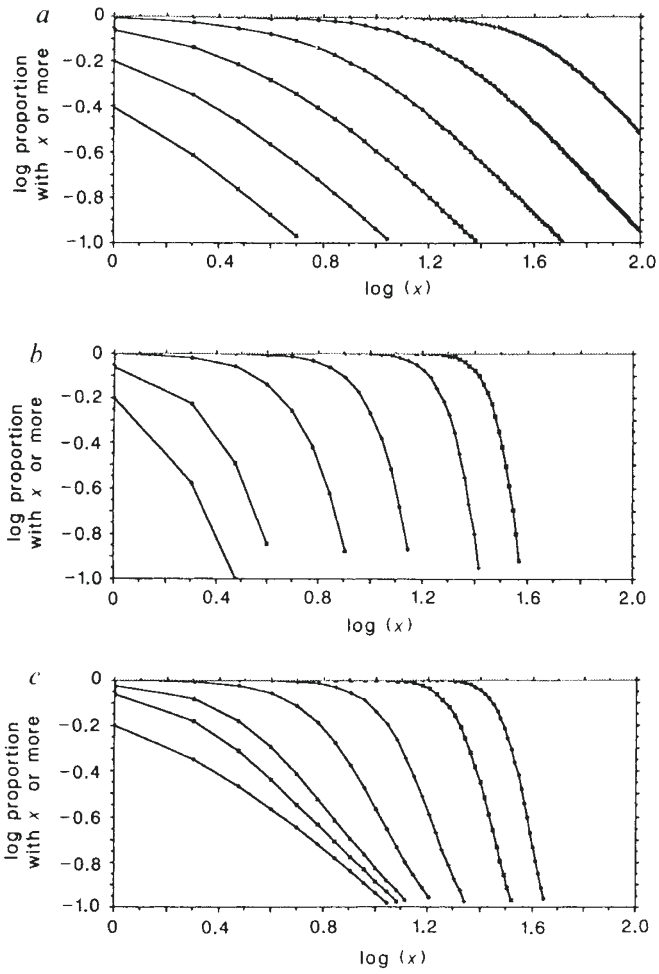


Fig. 1 The expected distributions of mutants among sister cultures that have experienced various kinds of mutational process. Each shows how the logarithm of the probability that a culture has at least x mutants ($\log_{10} P_x$) declines as $\log_{10} x$ increases. *a*, The case where the mutational events are distributed randomly among the entire lineage of each culture (calculated from Lea and Coulson Eq-15, ref. 5). The six examples shown are where the mean number of events per culture (m) is 0.5, 1, 2, 4, 8 and 16. *b*, The case where the mutational events are distributed randomly among the last generation of cells (that is they form a Poisson distribution). The six examples shown are where the mean number of events per culture (μ) is 1, 2, 5, 10, 20 and 30. *c*, Some combinations of the two previous distributions (where m events occur at random throughout the previous lineage of each culture and μ additional events occur in the final generation, among the cells that are being tested for the presence of the mutation). The examples show the expected distribution when $m = 1$ and $\mu = 0$, 1, 2, 5, 10, 20 and 30.

mutants arise after plating, among the bacteria exposed to the selection pressure. The numbers in independent cultures should form a Poisson distribution, and some examples are shown, for comparison, in Fig. 1*b*. The distributions are much narrower and log P_x falls rapidly as log x increases, especially when the average number of mutations per culture is high.

If both forms of mutation are occurring, the distribution of log P_x will be a composite of the two previous distributions. The lower end of the distribution will tend to be dominated by the mutations arising after selection, so that log P_x will initially fall steeply with increasing x; at high values, the distribution will be dominated by the rare jackpots produced by the mutations that occurred early during prior growth, and P_x will be inversely proportional to x. Various composite distributions are shown in Fig. 1*c*.

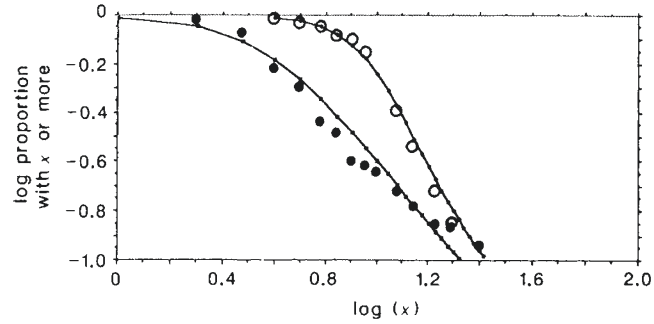


Fig. 2 The observed distribution of Lac$^+$ revertants for *E. coli* Δ *lac pro F' lacZ*$_{am}$ (filled circles) and for a derivative bearing a deletion of *uvrB* and *bio* (open circles). Sixty 3-ml cultures of each strain, in M9 medium plus 0.05% yeast extract, were grown overnight with aeration at 37 °C, and then each culture was centrifuged, resuspended in M9 and plated on M9-lactose plates. The colonies were counted after 48 h. The curves show the expected distributions for $m = 1.6$, $\mu = 2$ (on the left), and $m = 1.6$, $\mu = 7$ (on the right). (Roughly a third of the revertants, produced by these strains, are amber suppressors, and these form slightly smaller colonies at 48 h.)

Mutations affecting lactose fermentation

Bacterial variation in colony morphology was observed soon after the invention of plating procedures. One of the earliest studied examples was the formation of papillae of lactose-fermenting mutants, seen when non-fermenting strains of *E. coli* are plated out on rich media containing lactose, and it was attributed simply to a process of random variation and survival of the fittest[9]. This is a mutation that presumably is expressed as soon as it occurs, and so it is suitable for our present purposes.

Following the publication of tables for the Luria–Delbrück distribution[5], Ryan studied the distribution of Lac$^+$ variants in replicate cultures of Lac$^-$ strains spread on lactose-minimal plates and, surprisingly, he found that in some strains it was narrower than expected[10]. We have confirmed this. The shape of the distribution of Lac$^+$ revertants of strains bearing an amber mutation in *lacZ* depends on the genotype of the strain and on the medium in which it was growing before being plated; in certain circumstances, the distribution looks rather like some of those shown in Fig. 1*c*, as if there are two periods when mutations are occurring—first during the period of growth (producing the kind of distribution shown in Fig. 1*a*) and then in stationary phase (producing the kind of distribution shown in Fig. 1*b*). It is the second of these that is determined by the medium and genotype. In separate experiments, not described here, we have shown that such stationary phase mutation seems to be under the control of genes in the *uvrB-bio* region of the chromosome, but we have so far not been able to derive any simple picture of the regulatory process.

When multiple cultures of *E. coli lacZ*$_{am}$ and of the same strain bearing a deletion in the *uvrB-bio* region are grown in yeast extract or in M9-glycerol and then plated on minimal-lactose plates, the two strains appear to have the same mutation rate during growth (they are indistinguishable at the upper ends of their distributions), but the *uvrB* strain adds a large number of extra mutants that are apparently part of a Poisson distribution (Fig. 2). Interestingly, these extra mutants tend to be slower to produce colonies than the mutants that are members of jackpots (that is, slower than the mutants already in existence at the time of plating). This suggests that the extra mutant colonies are the result of late events, occurring on the plates when the bacteria are in the presence of lactose and under strong selection pressure to become Lac$^+$. The *uvrB* strain produces, each day about 3–5 times as many late mutants as the *uvr*$^+$ strain and it is therefore the strain of choice, when testing for the possible existence of directed mutation.

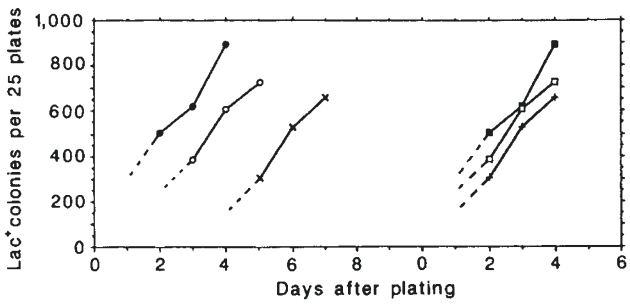


Fig. 3 The accumulation of Lac$^+$ revertants of *E. coli* Δ *uvrB-bio* Δ*lac pro F'lacZ*$_{am}$ after exposure to lactose on M9 plates. Six large cultures were grown overnight in M9 plus 0.1% glycerol. Fifteen 1-ml aliquots from each culture were plated with 2.5 ml top-layer agar on to M9-agar plates; the 90 plates were then overlaid with top-layer agar to ensure that all bacteria were buried in the agar; five plates from each culture were overlaid immediately with top-layer agar containing lactose, five were overlaid at 24 h, and five at 72 h. One of the six cultures proved to have a small jackpot of mutants, and so its plates were discarded; the plates from the remaining five cultures were counted each day, and the total counts are shown in real time on the left and in relation to time after the addition of lactose on the right.

To determine the role of lactose in the production of late Lac$^+$ mutants, aliquots of several replicate cultures of the *uvrB* strain were plated on M9 plates without lactose. When the plates were overlaid immediately with top-layer agar containing lactose, they showed an average of 20 colonies by 48 h and then gained an extra 8 colonies a day thereafter. If, however, the addition of lactose was delayed one or three days, the whole time course of appearance of colonies was delayed one or three days; this delay was seen even when the plates contained the non-hydrolysable inducer of the *lac* operon, isopropyl-thiogalactoside. In short, the accumulation of late Lac$^+$ mutants occurred only in the presence of lactose (Fig. 3).

One trivial explanation could be that this Lac$^-$ strain can grow slowly in lactose (either because its amber mutation in *lacZ* is leaky or because the lactose is impure), and that is why it can continue to produce mutants on plates containing lactose. We have therefore tested for the presence of another rapidly expressed mutation, valine-resistance (Val^R), by overlaying the lactose plates at various times with agar containing glucose and valine, and we find that the number of Val^R colonies countable two days after overlaying with valine actually goes slightly downwards with time (just as did the number of Lac$^+$ colonies in the absence of lactose). In other words, a population of bacteria that is accumulating Lac$^+$ revertants on a lactose-minimal plate is not, at the same time, accumulating Val^R mutants.

This experiment suggests that populations of bacteria, in stationary phase, have some way of producing (or selectively retaining) only the most appropriate mutations. For *E. coli*, under pressure to revert one particular amber codon, 'appropriate mutation' does not add much to the high background mutation rate arising from errors during replication. The next two examples, however, suggest that it plays a much larger role when the selection pressure is for complex changes in genotype that must seldom if ever arise from chance errors in replication.

The Shapiro deletion

In order to study the process of spontaneous deletion, Shapiro[11] constructed a strain in which *araC* (the positive regulatory element controlling the arabinose operon) is placed upstream of *lacZ* but separated from it by a short segment of Mu bacteriophage DNA that contains transcription terminating signals. Both the arabinose and the lactose operons are missing in this strain, so it is Lac$^-$ and Ara$^-$; if, however, it can delete the intervening Mu segment it will be able to grow on lactose, provided arabinose is present—a phenotype we will refer to as Lac(Ara)$^+$. Shapiro has shown that the deletion occurs in populations of cells plated out on lactose-arabinose-minimal plates. The Lac(Ara)$^+$ mutants are not, however, the result of spontaneous deletions occurring during the prior growth of the cultures but are apparently arising on the plates, because all the colonies appear much later than the colonies that are formed when cultures are seeded with a few pre-existing Lac(Ara)$^+$ cells and then grown up and plated. Indeed, the frequency of spontaneous deletion during growth is probably less than 10^{-11}, because immediate-colony-formers have never been found in cultures grown in the absence of selection.

Because this appears to be potentially a case of directed mutation, we have investigated it a little further. The delay in the formation of Lac(Ara)$^+$ colonies on selective plates is not simply some problem, in stationary phase, with the expression of *araC* and *lacZ* following their fusion, for we have prepared fused strains that are Lac(Ara)$^-$ because they have point mutations in *lacZ* or *araC*, and these strains show no such delay in the appearance of their Lac(Ara)$^+$ revertant colonies. Thus the act of gene fusion is indeed a late event, occurring long after the cells are put on lactose-arabinose-minimal plates. The event is influenced by some special property of the segment of Mu DNA that is being deleted, because the deletion does not occur when there is Mu DNA elsewhere in the chromosome[11]. Even so, we would like to know if the deletion is actually dependent on the presence of lactose and arabinose.

We have therefore studied the apperance of Lac(Ara)$^+$ cells (cells that are immediate-colony-formers on lactose-arabinose-minimal plates) in liquid cultures that are kept gently aerated after reaching stationary phase. The results can be summarized as follows. (1) In rich media containing both lactose and arabinose, large numbers of Lac(Ara)$^+$ cells start to appear after 3-4 days at 30 °C. (2) Under these conditions, even minority populations (marked with rifamycin-resistance) contribute to the final population of Lac(Ara)$^+$ cells, and from this we have calculated that the proportion of bacteria that are becoming Lac(Ara)$^+$ must be at least 10^{-8}. (3) In rich media without lactose (or without arabinose), Lac(Ara)$^+$ cells do not accumulate at a detectable rate. (4) When lactose is added to a stationary culture that has arabinose but not lactose, Lac(Ara)$^+$ cells promptly start to accumulate; judging from their rate of increase, the first Lac(Ara)$^+$ cell must be formed within 1-2 h of the addition of lactose.

This therefore seems to be another example of the production of appropriate mutations in response to selection. In one sense, it is a rather clear case, because this particular mutation does not occur at a measurable rate in the absence of selection; so bacteria, in stationary phase, do apparently have access to some process that either can prevent useless mutations from occurring or can destroy unsuccessful mutants soon after they arise. At the same time, it is perhaps a rather special case, because it involves the movement (deletion) of a piece of Mu DNA and does not occur if there is Mu elsewhere in the chromosome.

So far we have only considered examples of spontaneous mutation where bacteria are under pressure to reverse defects that have been previously introduced into their genes by mutagenesis or by genetic engineering. For our final example, we will briefly consider the response of *E. coli* to more natural forms of selection pressure.

Activation of the cryptic genes in *E. coli*

The usual way of distinguishing the various pathogenic and nonpathogenic *Enterobacteriaceae* is to observe which sugars they can use as a source of energy; for example, *E. coli* ferments lactose, whereas the various members of the *Shigella* and *Salmonella* families do not. The standard test is to inoculate a small volume of peptone plus sugar, and observe turbidity, *p*H and

CO_2 production after 1-2 days. When this system of classification was being prepared, early this century, certain bacterial species were observed to be 'late' fermenters of some sugars. Thus it is a regular feature of some bacteria that it takes a week or more before the population of cells, produced by growth on the peptone, are able to generate a variant that can hydrolyse the sugar; for example, *Sh. sonnei* is classified as a 'late' fermenter of lactose, and many strains of *E. coli* are 'late' fermenters of salicin (an aromatic β-glucoside). The genes coding for the appropriate enzymes are there, but they are not readily accessible.

Bacteria apparently have an extensive armoury of such 'cryptic' genes that can be called upon for the metabolism of unusual substrates. The mechanism of activation varies. In some cases it is simply by the movement of an insertion sequence into a position upstream of the cryptic gene[12], but in others it may require several changes in base sequence. It is these latter examples that are of special interest. For instance, *E. coli* turns out to have a cryptic gene (*ebgA*^{0}) that it can call upon to hydrolyse lactose, if the usual gene for this purpose (*lacZ*) has been deleted[13]. The activation of *ebg* requires at least two mutations, one in the repressor (*ebgR*) and one at a particular site in the gene coding for the enzyme (*ebgA*) to make the enzyme capable of hydrolysing lactose[14]. During growth, each of these point mutations occurs at a frequency of less than 10^{-8}; neither on its own will allow a *lacZ* deletion strain to use lactose[15], yet colonies of such a strain, that have grown to stationary phase on a plate containing lactose, will produce Lac$^+$ papillae in about two weeks. That such events ever occur seems almost unbelievable, but we have also to realize that what we are seeing probably gives us only a minimum estimate of the efficiency of the process, since in these cases the stimulus for change must fairly quickly disappear once a few mutant clones have been formed that can exploit the novel sugar. It is difficult to imagine how bacteria are able to solve complex problems like these—and do so without, at the same time, accumulting a large number of neutral and deleterious mutations—unless they have access to some reversible process of trial and error.

Discussion

The main purpose of this paper is to show how insecure is our belief in the spontaneity (randomness) of most mutations. It seems to be a doctrine that has never been properly put to the test. We describe here a few experiments and some circumstantial evidence suggesting that bacteria can choose which mutations they should produce. But we realize that this is too important an issue to be settled by three or four rather ambiguous experiments.

The origin of genetic variation has been the subject of bitter controversy, throughout the nineteenth century and into the first half of the twentieth. Is all variation essentially random, like thermal noise? Or can the genome of an individual cell profit by experience? At its extremes, it was an argument between reductionists and romantics—between those who sought to explain the evolution and behaviour of the biosphere in terms of the laws of physics, and those who wished to make the success of evolution just another manifestation of the mysteriousness of living things.

The early triumphs of molecular biology strongly supported the reductionists. The DNA double helix is an extraordinarily stable structure, but it is subject to spontaneous degradation and to errors in replication, and these changes are, in a sense, due to thermal noise. Furthermore, the discovery of all the elements that lie between DNA sequence and protein structure gave rise to the central dogma of molecular biology, and this doctrine denies any possible effect of a cell's experience upon the sequence of bases in its DNA.

Curiously, when we come to consider what mechanism might be the basis for the forms of mutation described in this paper we find that molecular biology has, in the interim, deserted the reductionist[16]. Now, almost anything seems possible. In certain systems, information freely flows back from RNA into DNA; genomic instability can be switched on under conditions of stress, and switched off when the stress is over; and instances exist where cells are able to generate extreme variability in localized regions of their genome. The only major category of informational transfer that has not been described is between proteins and the messenger RNA (mRNA) molecules that made them. If a cell discovered how to make that connection, it might be able to exercise some choice over which mutations to accept and which to reject.

Since this is the kind of versatility and adaptability we seem to be seeing in these experiments with *E. coli*, it is worth considering briefly how such a connection might be made. In a very direct way, the cell could produce a highly variable set of mRNA molecules and then reverse-transcribe the one that made the best protein; for this, it would have to have a special organelle (perhaps like the gag-pol complex of retroviruses) which contains reverse transcriptase plus some element that somehow monitors the protein product and determines whether the mRNA should go on being translated or should be transcribed into DNA; this could be an efficient process and makes an attractive hypothesis because it gives us an unusual explanation for the origin of retroviruses. Less efficient would be the production of reverse transcripts at random, but even that could achieve the desired end result, if reverse transcription were switched on temporarily, just at the time the cell resumed growth; any cell that started growing would acquire a reverse transcript of the variant sequence that made it able to grow, and so the sequence would eventually be incorporated by recombination into the genome of one of the cell's descendents (indeed, the same result could perhaps be achieved starting with DNA copies of segments of the genome rather than RNA transcripts). Each of these processes would allow individual cells to subject a subset of their informational macromolecules to the forces of natural selection. Each could, in effect, provide a mechanism for the inheritance of acquired characteristics.

We thank Victoria Cairns and Rainer Riester for writing a program for calculating the composite distributions of mutants, Patricia Foster and Richard Peto for comments on the manuscript, and Eric Eisenstadt for the hospitality of his laboratory. The work was supported by the National Science Foundation.

Received 2 May; accepted 11 July 1988.

1. Luria, S. E. & Delbrück, M. *Genetics* **28**, 491-511 (1943).
2. Lederberg, J. & Lederberg, E. M. *J. Bact.* **63**, 399-406 (1952).
3. Cavalli-Sforza, L. L. & Lederberg, J. *Genetics* **41**, 367-381 (1956).
4. Hayes, W. *The Genetics of Bacteria* (Blackwell, Oxford, 1964).
5. Lea, D. E. & Coulson, C. A. *J. Genet.* **49**, 264-285 (1949).
6. Armitage, P. *J. R. statist. Soc. B.* **14**, 1-40 (1952).
7. Mandelbrot, B. *J. appl. Prob.* **11**, 437-444 (1974).
8. Luria, S. E. *Cold Spring Harb. Symp. quant. Biol.* **16**, 463-470 (1951).
9. Hadley, P. *J. infect. Dis.* **40**, 1-312 (1927).
10. Ryan, F. J. *Nature* **169**, 882-883 (1952).
11. Shapiro, J. A. *Molec. Gen. Genet.* **194**, 79-90 (1984).
12. Reynolds, A. E., Felton, J. & Wright, A. *Nature* **293**, 625-629 (1981).
13. Campbell, J. H., Lengyel, J. A. & Langridge, J. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1841-1845 (1973).
14. Hall, B. G. *J. Bact.* **129**, 540-543 (1977).
15. Hall, B. G. & Zuzel, T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3529-3533 (1980).
16. Reanney, D. C. *Nature* **307**, 318-319 (1984).

Week 3

Suppressors

Gene Expression

Proc. Nat. Acad. Sci. USA
Vol. 72, No. 7, pp. 2738–2742, July 1975
Genetics

Conditional-lethal mutations that suppress genetic defects in morphogenesis by altering structural proteins

(protein interaction/temperature-sensitive mutants/cold-sensitive mutants/phage P22 morphogenesis/reversion)

JONATHAN JARVIK AND DAVID BOTSTEIN

Department of Biology, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Communicated by Boris Magasanik, April 21, 1975

ABSTRACT **An analysis of revertants of missense mutants in phage P22 has shown: *(i)* New temperature-sensitive (TS) and cold-sensitive (CS) phenotypes are often acquired concomitant with reversion. *(ii)* In many cases, these new phenotypes are due to second-site mutations (suppressors) that correct the original defect. *(iii)* Sometimes the suppressor mutation is not in the same gene as the original mutation. *(iv)* Extragenic suppressors are almost always in genes whose products are known to interact physically with the original gene products. *(v)* The suppressor mutations typically retain their TS or CS phenotypes when crossed into wild-type genetic backgrounds. *(vi)* Some TS and CS mutants derived by reversion can themselves be reverted to produce additional mutations.**

We have shown that genetic reversion of missense mutants can be of value in producing new temperature-sensitive and cold-sensitive mutations affecting related functions. We suggest that our approach can be extended to organisms with large genomes.

Temperature-sensitive and cold-sensitive mutants allow an investigator to modulate protein activity *in vivo* and *in vitro* simply by varying the temperature. Such mutants thereby greatly aid in the analysis of biochemical and developmental events at the molecular, cellular, and organismic levels (1–7).

For conditional-lethal mutants, nonpermissive conditions provide an absolute selection for further mutation, since only revertant individuals can grow. Such revertants need not be true wild type; rather, they may have acquired suppressors—new mutations that act so as to correct, replace, or bypass the original defect.

The analysis of suppressors has proved to be of great value in diverse biological investigations (8). However, such analysis is operationally difficult if the suppressors under investigation lack characteristic phenotypes of their own—phenotypes that are expressed independent of the original mutations.

In this paper, we demonstrate, using bacteriophage P22, that new mutations with temperature-sensitive and cold-sensitive phenotypes frequently appear among the revertants of existing missense mutants. Our experiments were undertaken with two principles in mind. *(i)* Suppressors are often missense mutations (8, 9). *(ii)* Missense mutations often confer cold-sensitive or temperature-sensitive phenotypes (2). We therefore expected, and found, that some revertants of existing missense mutants would contain suppressor mutations that produce cold-sensitive or temperature-sensitive phenotypes in and of themselves.

Abbreviations: TS, temperature-sensitive phenotype; *ts*, genetic determinant of TS phenotype; CS, cold-sensitive phenotype; *cs*, genetic determinant of CS phenotype.

MATERIALS AND METHODS

Phage, Bacteria, and Media. The following mutations of bacteriophage P22 have been described (4,10): c_1-7, 13$^-$*am*H101 (in our notation, the gene designation precedes the allele designation; thus, 13$^-$*am*H101 means that *am*H101 is an allele of gene 13), 1$^-$*am*H21, 1$^-$*am*H58, 8$^-$*am*N26, 1$^-$*am*N23, 2$^-$*cs*H22, 2$^-$*cs*H59, 1$^-$*cs*H139, 5$^-$*cs*H126, 12$^-$*ts*12.1, 3$^-$*ts*RH203, 3$^-$*ts*3.1, 2$^-$*ts*2.1, 1$^-$*ts*1.1, 5$^-$*ts*N26, 5$^-$*ts*N105, 8$^-$*ts*N102, and 10$^-$*ts*10.1. 1$^-$*cs*RH21D was selected as an *am*$^+$ revertant of *am*H21. 8$^-$*cs*RN26D was selected as an *am*$^+$ revertant of *am*N26. 1$^-$*cs*H137 was erroneously reported to be in gene 8 in an earlier publication (4). Bacterial strains are derivatives of *Salmonella typhimurium* LT2. DB53 and its amber suppressor derivative DB74, and DB21 and its amber suppressor derivative DB28, have been described (10). DB7000 is *leu*$^-$*am*, a derivative of DB21. DB7004 is a derivative of DB7000 that contains the *su*$^+$ allele of DB74. DB7002 is a derivative of DB7000 that contains the *su*$^+$ allele of DB28. LB broth, nutrient agar, λ plates, and buffered saline have been described (11).

Genetic Mapping, Complementation Tests, and Genetic Crosses. All newly isolated *cs* and *ts* mutations were mapped by the methods of efficiency of plating or permissive rescue on prophage deletion strains (12). Any mutation that was not shown by deletion mapping to be between known markers in a known gene was assigned to a complementation group after tests against known amber alleles by the complementation in liquid culture method (10). Genetic crosses were performed as described (10).

Isolation of Independently Arising Revertants. Permissively grown plaques were touched with sterile needles and streaked on plates which were then incubated nonpermissively. Revertant plaques which grew on these plates were picked with a sterile capillary tube and suspended in buffered saline; the phage were tested for plaque-forming ability on the hosts and at the temperatures described in *Results*.

Nomenclature for Revertants. The phage to be reverted is called the parental mutant. A revertant is named by enclosing the name of the parental mutant in parentheses, preceding it with the letter "r," and succeeding it by a letter or number. All this is preceded by a designation of the revertant's phenotype (TS for temperature-sensitive and CS for cold-sensitive). This name stands for a phage of a particular phenotype, but it does not strictly define a genotype. The name of the new mutation should indicate its phenotype in a wild-type background. In naming the new mutation, the parentheses are dropped and lower case letters are used. For example, the first temperature-sensitive revertant of *cs*H137 isolated was given the name TSr(*cs*H137)A. The revertant

Table 1. Survey of revertants of missense mutants

Parental mutant	Conditions for selection of revertants	No. of revertants examined	Revertants with acquired CS or TS phenotypes	Gene in which new mutation is located	Is the new mutation a suppressor?
2⁻*cs*H22	18°	5	TSr(*cs*H22)A	Gene 2	ND
2⁻*cs*H59	18°	4	TSr(*cs*H59)A	Gene 2	ND
1⁻*cs*RH21D	18°	3	TSr(*cs*RH21D)3	Gene 1	No
1⁻*csrr*RH21D3A	18°	14	TSr(*csrr*RH21D3A)1	Gene 1	Yes
1⁻*cs*H137	18°	36	TSr(*cs*H137)A	Gene 5	Yes
			TSr(*cs*H137)B	Gene 5	Yes
			TSr(*cs*H137)C	Gene 5	Yes
			TSr(*cs*H137)D	Gene 5	Yes
1⁻*cs*H139	18°	30	TSr(*cs*H139)A	Gene 1	Yes
8⁻*cs*RN26D	18°	10	TSr(*cs*RN26D)1	Gene 8	ND
5⁻*cs*H126	18°	33	—	—	—
5⁻*csrr*H58G1	18°	48	TSr(*csrr*H58G1)A	Gene 5	ND
12⁻*ts*12.1	37°	40	—	—	—
3⁻*ts*3.1	41°	96	—	—	—
3⁻*ts*RH203	41°	32	—	—	—
2⁻*ts*2.1	37°	88	—	—	—
1⁻*ts*1.1	37°	58	—	—	—
1⁻*tsr*RH21D3	41°	14	CSr(*tsr*H21D3)A	Gene 1	No
8⁻*ts*N102	37°	32	CSr(*ts*N102)A	Gene 8	ND
5⁻*ts*N26	37°	94	—	—	—
5⁻*tsr*H58G	37°	114	CSr(*tsr*H58G)1	Gene 5	Yes
			CSr(*tsr*H58G)2	ND	ND
			CSr(*tsr*H58G)3	ND	ND
5⁻*tsr*H58E	37°	49	—	—	—
5⁻*ts*N105	37°	33	—	—	—
5⁻*tsr*H137B	41°	16	—	—	—
5⁻*tsr*H137C	34°	22	CSTSr(*tsr*H137C)19	Gene 12	Yes
1⁻*am*H58	Nonpermissive Su⁺ host at 30°	12	TSr(*am*H58)A	Gene 5	Yes
			TSr(*am*H58)D	Gene 5	Yes
			TSr(*am*H58)E	Gene 5	Yes
			TSr(*am*H58)G	Gene 5	Yes
			TSr(*am*H58)H	Gene 5	Yes
8⁻*am*N26	Nonpermissive Su⁺ host at 30°	21	CSTSr(*am*N26)K	Gene 5	Yes
1⁻*am*H21	Nonpermissive Su⁺ host at 30°	16	CSr(*am*H21)E	Gene 1	No

Revertants were selected and tested as described in *Materials and Methods*. "ND" in the table means "not determined." 1⁻*am*H58 fails to grow at any temperature on the Su⁺ host DB7004, but it does grow at 18°, 30°, and 41° on the Su⁺ host DB7002. 1⁻*am*H21 shows the same pattern of growth. 8⁻*am*N26 does not grow on DB7004 at any temperature, but it does grow on DB7002 at 30° and 41°. A mutation is called a suppressor only if it has been separated from its parental mutation by recombination. Thus, some revertants [e.g., TSr(*cs*H22)A] may contain suppressors but not yet have been identified as such. All four *tsr*H137 mutations are separable from one another by recombination, as are all five *tsr*H58 mutations. *ts*12.1, *ts*RH203, *ts*2.1, *ts*N102, and *cs*H137 phages carry c_1-7 (clear plaque) mutations. *cs*H22, *cs*H59, *csrr*H58G, and *ts*3.1 phages carry c_1-7 and *am*H101 (lysis defective) mutations. All other phages are wild type.

contains a new mutation at a different site from *cs*H137. If this mutation were shown to confer a temperature-sensitive phenotype when crossed away from *cs*H137, it would be named *tsr*H137A. (If the new mutation had been shown to be at the same genetic site as *cs*H137, it would likewise be called *tsr*H137A, since we would know that it confers a temperature-sensitive phenotype in a wild-type background). If *tsr*H137A were to produce a cold-sensitive revertant, that phage would be named CSr(*tsr*H137A)1; if the new mutation were shown to confer cold-sensitivity on a wild-type background, it would be named *csrr*H137A1, and so on.

RESULTS

Generation of New Mutants. We surveyed spontaneous independent revertants of three kinds of missense mutants for the presence of new temperature-sensitive and cold-sensitive phenotypes. CS⁺ revertants of cold-sensitive mutants were tested for temperature-sensitivity. TS⁺ revertants of temperature-sensitive mutants were tested for cold-sensitivity. Finally, revertants of a third class of mutant—amber mutants that do not grow in one of our standard amber-suppressing hosts—were selected for growth on the "nonpermissive" amber-suppressing host, and these revertants were tested for cold-sensitivity and temperature-sensitivity. (We presume that for this third class of mutant a defective protein with a missense amino acid at the amber position is made in the nonpermissive amber-suppressing host; the amber mutants are, therefore, phenotypically missense mutants in these hosts.) For each revertant with a newly acquired cold-sensitive or temperature-sensitive phenotype, we located the genetic determinant of the phenotype by deletion mapping, complementation tests, and, in some cases, by three-factor crosses.

Table 1 details the results of our survey. Of nine cold-sensitive mutants examined, eight yielded temperature-sensi-

Table 2. Phenotypes conferred by suppressors in parental and wild-type backgrounds

	18°	21°	26°	30°	34°	37°	41°
TSr(*cs*H137)B	+	+	+	+	+	(+)	–
*ts*rH137B	+	+	+	+	+	(+)	–
TSr(*cs*H137)C	+	+	+	+	–	–	–
*ts*rH137C	+	+	+	(+)	–	–	–
TSr(*cs*H137)D	+	+	–	–	–	–	–
*ts*rH137D	+	+	+	–	–	–	–
TSCSr(*am*N26)K	–	(+)	+	+	+	(+)	–
*ts*rN26K	(+)	+	+	+	+	(+)	–
TSr(*am*H58)A	+	+	+	+	+	+	–
*ts*rH58A	+	+	+	+	+	+	–
TSr(*am*H58)D	+	+	+	+	+	+	–
*ts*rH58D	+	+	+	+	+	+	–
TSr(*am*H58)E	+	+	+	+	+	–	–
*ts*rH58E	+	+	+	+	+	–	–
TSr(*am*H58)G	+	+	+	+	+	–	–
*ts*rH58G	+	+	+	+	+	–	–
TSr(*am*H58)H	+	+	+	+	+	+	–
*ts*rH58H	+	+	+	+	+	+	–
CSr(*ts*rH58G)1	–	–	+	+	+	+	–
*cs*rrH58G1	–	+	+	+	+	+	+

The upper member of each pair is the revertant itself; the lower member is a phage that carries the suppressor in a wild-type background.

Backcrosses were performed as follows. For *ts*rH137 phages: A 1⁻*am*N23 mutation was crossed into the TSr strain to produce a phage with an AM,TS phenotype. The AM,TS phage was then crossed with wild type and an AM⁺,TS recombinant was recovered. The Am⁺,TS phage was checked to be sure that the parental *cs*H137 allele was absent by crossing it with a 1⁻*cs*H137 5⁻*ts*N26 double mutant phage and testing TS⁺ progeny for a CS⁺ phenotype (*am*N26 maps between the *ts*r mutation and *cs*H137). CS⁺,TS⁺ progeny indicate that the parental *cs* allele is absent from the AM⁺,TS phage. For *ts*rH58 and *ts*rN26 phages: The *am ts*r revertant was plated at permissive temperature on a prophage deletion lysogen (10) that contains the *am*⁺ allele but not the *ts*⁺ allele. *Am*⁺ recombinants were recovered. For *cs*rrH58G1: The phage, which is TS⁺ at 37° but still TS at 41°, was crossed with a 10⁻*ts*10.1 phage, and a fully TS⁺ recombinant that retains the CS phenotype was recovered.

tive mutants by reversion. Of 13 temperature-sensitive mutants examined, four yielded cold-sensitive mutants by reversion. Of three amber mutants examined, three yielded cold-sensitive or temperature-sensitive mutants by reversion. We conclude that *the probability is high that a missense mutant can yield new TS or CS mutants by reversion.*

Some of the New Mutations Are Suppressors. At least seven of the mutants in our survey yielded revertants with new mutations at genetic sites other than the parental sites. These revertants therefore carry suppressors. After separating some of these suppressors from their parental mutations by recombination, we examined the backcrossed phages for growth at various temperatures (Table 2). Clearly, the mutations confer their temperature-sensitive or cold-sensitive phenotypes in wild-type backgrounds, although some of them grow at greater extremes of temperature than with the parental mutations present. We conclude that in each case the mutant phenotype does not depend upon the presence of the parental mutation, and in this respect *the mutants behave like standard temperature-sensitive and cold-sensitive mutants.*

The Chain of Revertants. In several cases, a mutation that was generated by reversion was used as a parental mutation to generate further mutations. These mutations, therefore, comprise a *chain of revertants* in which each member is related by reversion to its predecessor. Three such chains are shown in Table 3. One of the chains contains only extragenic revertants, one contains only intragenic revertants, and one contains both. They illustrate clearly that several conditional-lethal mutations can be generated from a single missense mutant by sequential rounds of reversion.

Table 3. Three chains of revertants

1⁻*am*H58 → 5⁻*ts*rH58G → 5⁻*cs*rrH58G1 → 5⁻TSr(*cs*rrH58G1)A

1⁻*cs*RH21D → 1⁻*ts*rRH21D3 → 1⁻*cs*rrRH21D3A → 1⁻TSr(*cs*rrRH21D3A)1

1⁻*cs*H137 → 5⁻*ts*rH137C → 12⁻*cs*rrH137C19

DISCUSSION

The analysis of suppressors—mutations that can remedy the defects caused by other mutations—has been of value in the study of protein interaction, protein synthesis, and gene control (8, 14). However, a severe liability to the general use of suppressor analysis has derived from the fact that a suppressor typically does not have a phenotype in its own right. A suppressor whose only phenotype is an effect on another mutation's phenotype can only be observed in the presence of that other mutation, and so the examination of the suppressor itself is difficult. In the experiments reported in this paper, we have confined our attention to phage that have acquired, in the process of reversion, *new mutations that restore growth at one temperature and are lethal at another.* The conditional-lethal phenotypes of the new mutations, in turn, allowed us to show by standard genetic methods that in many instances the new mutation is in a gene other than the one in which the original mutation lies. Having confined our attention to conditional-lethal mutations, we are in a position to take advantage of their additional virtues: they can be crossed into diverse genetic backgrounds at will, they can be used in temperature-shift experiments *in vivo* and *in vitro* (15), and they can be used in schemes of enzyme or protein-complex purification (16, 17).

What is the probability that a revertant has a new phenotype?

Of nine CS mutants examined, eight yielded temperature-sensitive revertants, whereas only four of 13 TS mutants tested yielded cold-sensitive revertants (Table 1). In addition, we had to examine only a few CS⁺ revertants of each cold-sensitive mutant in order to obtain temperature-sensitives, but we had to examine many TS⁺ revertants of each temperature-sensitive mutant in order to find any cold-sensitives at all. If we pool the data for all nine CS mutants, we find that they yielded about one temperature-sensitive revertant per 20 total revertants (nine CS⁺TS revertants out of 183 CS⁺). A pooling of the data for the 13 TS mutants shows about one cold-sensitive revertant per 100 total revertants (six TS⁺CS out of 790 TS⁺). As shown by Tables 1 and 2, most of the revertants with new phenotypes carry suppressors. It must be stressed that these frequencies were obtained by examining independently arising revertants. Screening many revertants from a single mutant stock has not, in general, been so successful, probably because most of the revertants in each stock have a common ancestor (18).

We do not know why CS mutants revert to temperature-sensitivity more readily than TS mutants revert to cold-sen-

sitivity. It might simply reflect a relative rarity of genetic sites that can mutate to produce cold-sensitivity. Or it might indicate an intrinsic difference in the structural nature of cold-sensitive and temperature-sensitive mutant proteins. For example, temperature-sensitive proteins might typically be in disordered conformations at nonpermissive temperature, whereas cold-sensitive proteins might typically be in fixed, but inactive, conformations (19, 20). The structure at nonpermissive temperature, in turn, ought to influence the kinds of second-site mutations that could restore function (21).

How do our suppressors work?

We do not yet know how any of our suppressors restores growth. It is worth noting, however, that for all but one of our extragenic revertants, the suppressor is in a gene whose product is known to interact physically with the parental gene product (13). It is therefore an attractive possibility for the mechanism of suppression that the parental mutation and the suppressor mutation alter sites of mutual protein/protein interaction. By this hypothesis, the parental mutation destroys or distorts an interaction, and the suppressor produces a *compensating alteration* that restores the interaction. An intragenic suppressor produces a compensating alteration in the parental protein itself (9); an extragenic suppressor produces a compensating alteration in another protein that is in physical contact with the parental protein (22, 23). The hypothesis that the functional defect in a mutant protein can be corrected by physical interaction with another mutant protein finds support in the observation that antibodies can restore activity to some mutant enzymes *in vitro* (24), and in the many demonstrations that subunit interaction can restore activity lost by mutation in multimeric enzymes (25, 26).

Three of our parental mutants—1$^-$*cs*H137, 1$^-$*am*H58, and 8$^-$*am*N26—give rise to suppressors in gene 5. How might we interpret this in terms of the protein interaction hypothesis? The product of gene 5 (designated P5) is the major structural protein of the phage capsid, and the product of gene 1 (P1) is a structural protein that physically associates with the capsid (13). A suitable alteration in the structure of P5 might restore an interaction between P5 and P1 that was lost to the parental 1$^-$ mutant. Likewise, P5 and P8 interact, and so an alteration in P5 might restore a P5/P8 interaction that was lost to the parental 8$^-$ mutant. Thus, the parental/suppressor pairs that we obtained may indicate sites relevant to molecular interaction between P5 and P1 and between P5 and P8.

General advantages of experiments such as ours

By reversion of existing missense mutants, we have acquired many new *cs* and *ts* mutations in phage P22. These mutations should be of use in our ongoing study of P22 structure and function, and their further analysis may tell us much about the protein/protein interactions that occur during P22 morphogenesis. But here we wish to emphasize some ways in which experiments like ours might be useful in studying *any* organism.

1. Reversion Generates Mutations Suitable for Temporal Sequencing. We recently described a generally applicable method for determining the temporal order of events in a biological pathway (4). This method requires both *ts* and *cs* mutations in the genes of interest. The generation of both *ts* and *cs* mutations that affect related functions is intrinsic to the reversion system described in this paper, and so reversion is excellently suited for the acquisition of mutations for use in temporal sequencing.

2. Reversion Does Not Entail Mutagenesis. A hazard of mutagenic treatment is the possible occurrence of double mutants, and the more severe the mutagenesis the greater the risk. Since only revertant individuals will grow in nonpermissive conditions, reversion of conditional-lethal mutants is selective by nature, and so one can use reversion to isolate new mutations without the use of mutagenesis.

3. Reversion Is Function-Specific and Might Be Used to Advantage in Complex Organisms. Reversion by suppression is intrinsically function-specific, since the new mutation must remedy a very particular defect caused by the parental mutation. Even in a complex organism, the number of essential genes that can mutate to suppress any particular mutant phenotype is surely small—much smaller than the total number of essential genes. We therefore believe that the isolation of conditional-lethal suppressors could, in some cases, allow one to zero in on small subsets of essential genes (some of whose products are physically or functionally related) in organisms such as yeast, nematodes, and *Drosophila*. [We do not expect that *every* suppressor would affect a protein that is intimately related to the parental gene product (8), but we do suggest that a substantial number would.] Finally, suppressor analysis ought to be especially useful for investigating proteins that engage in strong noncovalent interactions with one another. These include not only the structural proteins (such as those in phage particles), but also the many enzymes that function as members of multi-protein complexes (27, 28).

We thank Larry Soll, Jonathan King, Van Jarvik, Ken Lew, Miriam Susskind, and George Weinstock for useful discussions. This work was supported in part by Grants VC18B and VC18C from the American Cancer Society and in part by Grant R01-GM21253-01 from the National Institutes of Health.

1. Horowitz, N. & Leupold, U. (1951) *Cold Spring Harbor Symp. Quant. Biol.* **16**, 65–72.
2. Campbell, A. (1961) *Virology* **14**, 22–32.
3. Edgar, R. & Lielausis, I. (1964) *Genetics* **49**, 649–662.
4. Jarvik, J. & Botstein, D. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2046–2050.
5. Hartwell, L., Culotti, J., Pringle, J. & Reid, B. (1974) *Science* **183**, 46–51.
6. Suzuki, D. (1970) *Science* **170**, 695–706.
7. Brenner, S. (1974) *Nature* **248**, 785–787.
8. Hartman, P. & Roth, J. (1973) *Adv. Genet.* **17**, 1–105.
9. Yanofsky, C., Ito, J. & Horn, V. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 151–162.
10. Botstein, D., Chan, R. & Waddell, C. (1972) *Virology* **49**, 268–282.
11. Chan, R. & Botstein, D. (1972) *Virology* **49**, 257–267.
12. Chan, R., Botstein, D., Watanabe, T. & Ogata, Y. (1972) *Virology* **50**, 883–898.
13. King, J., Lenk, E. & Botstein, D. (1973) *J. Mol. Biol.* **80**, 697–731.
14. Floor, E. (1970) *J. Mol. Biol.* **47**, 293–306.
15. Luftig, R. & Lundh, N. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1636–1640.
16. Gefter, M., Hirota, Y., Kornberg, T., Wechsler, J. & Barnoux, C. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 3150–3153.
17. Wright, M., Wickner, S. & Hurwitz, J. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3120–3124.
18. Luria, S. & Delbruck, M. (1943) *Genetics* **28**, 491–511.
19. Tai, P., Kessler, D. & Ingraham, J. (1969) *J. Bacteriol.* **97**, 1198–1134.

20. Guthrie, C., Nashimoto, H. & Nomura, M. (1969) *Proc. Nat. Acad. Sci. USA* **63**, 384–391.
21. Cox, J. & Strack, H. (1971) *Genetics* **67**, 5–17.
22. Aperion, D. & Schlessinger, D. (1967) *J. Bacteriol.* **94**, 1275–1276.
23. Tomizowa, J. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. D. (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.), pp. 549–552.
24. Rotman, B. & Celada, F. (1968) *Proc. Nat. Acad. Sci. USA* **60**, 660–667.
25. Fincham, J. (1966) *Genetic Complementation* (W. A. Benjamin, New York).
26. Crawford, I., Sikes, S., Belser, N. & Martinez, L. (1970) *Genetics* **65**, 201–211.
27. Case, M. & Giles, N. (1974) *Genetics* **77**, 613–626.
28. Jaenicke, R. & Helmreich, E. (eds.) (1972) *Protein–Protein Interactions* (Springer-Verlag, New York).

Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 968–972, February 1989
Genetics

PrlA is important for the translocation of exported proteins across the cytoplasmic membrane of *Escherichia coli*

(*lacZ*/β-galactosidase fusion proteins/overproduction lethality/signal sequence/protein secretion)

KRISTINA L. BIEKER AND THOMAS J. SILHAVY*

Department of Biology, Princeton University, Princeton, NJ 08544

Communicated by Jon Beckwith, October 20, 1988 (received for review August 9, 1988)

ABSTRACT **Strains of *Escherichia coli* in which *lacZ* (specifies β-galactosidase) is fused to genes that specify exported proteins such as LamB (λ receptor) exhibit unusual phenotypes. In particular, such strains are killed by high-level expression of the LacZ hybrid protein. Previous results suggest that this overproduction phenotype is the consequence of a lethal jamming of the cellular protein export machinery and this hypothesis is supported by the observed accumulation of the precursor forms of many noncytoplasmic proteins within the moribund cell. Under conditions in which protein export is compromised, biochemical and immunocytochemical analyses indicate that these hybrid proteins can be found in transmembrane orientation. To identify the cellular component rendered rate-limiting by the LacZ hybrid protein under jamming conditions we have utilized signal sequence mutations, which block entry of the hybrid protein into the export pathway, and a dominant suppressor of these lesions, *prlA4*. Data obtained with a series of merodiploids heterozygous and homozygous for *prlA*$^+$ and *prlA4* show that PrlA is the component sequestered by hybrid jamming. Taken together, these results suggest that PrlA is a component of the export machinery that functions in the translocation of proteins across the cytoplasmic membrane.**

In *Escherichia coli*, *lacZ* (specifies β-galactosidase) fusions have been particularly useful in the study of protein export, because strains in which this gene is fused to sequences specifying a noncytoplasmic protein often exhibit unusual phenotypes. First, the function of the β-galactosidase moiety is inhibited if the chimera is directed efficiently to a membrane location. Presumably in this context, the enzyme cannot oligomerize into the active tetrameric species. Second, the cellular protein export machinery cannot deal effectively with sequences of this large cytoplasmic enzyme. Accordingly, high-level synthesis of a LacZ hybrid protein can lead to a lethal "jamming" of the export apparatus. These characteristic phenotypes have been exploited to obtain export-defective signal sequence mutations in the gene to which *lacZ* is fused and to identify components of the export machinery such as *secA* (a peripheral inner membrane protein) and *secB* (a cytoplasmic protein) (for review see ref. 1).

The signal sequence mutations obtained using selections based on gene fusions have been employed to further our knowledge of the components of the export apparatus by utilizing the technique of interactive suppressors. For example, mutations that alter the *lamB* (encodes the outer membrane protein LamB) or *malE* [encodes the periplasmic maltose-binding protein (MBP)] signal sequence can confer a negative phenotype when recombined onto the corresponding wild-type gene. This permits selection for suppressors that alter a component of the export machinery and restore recognition of the mutant signal sequence. One gene identified by this method is *prlA* (also known as *secY*, encodes a 49,000-dalton integral inner membrane protein) (2, 3).

Recently, biochemical evidence supporting a direct role for PrlA and SecA in protein export has been presented (4, 5). Despite these advances the mechanistic role of these proteins in the export process is presently unclear and a number of fundamental questions remain to be addressed. We have exploited a suppressor allele of *prlA* and a *lamB–lacZ* gene fusion carrying a signal sequence mutation to investigate the LacZ-mediated jamming of the export apparatus. Our results indicate that the fusion protein sequences sequester PrlA rendering its function rate-limiting for export and they support the view that PrlA is a component of the translocator—i.e., the part of the export apparatus through which proteins cross the inner membrane.

MATERIALS AND METHODS

Media and Chemicals. Media and chemicals have been described elsewhere (6).

Strains and Plasmids. All strains are derivatives of *E. coli* K-12 strain MC4100 and are described in Table 1 and *Results* (2, 7). Strain constructions were done by using standard genetic techniques (6). The plasmid pMLB1107 is a pBR322-derived vector in which transcription of an inserted fragment can be regulated by the *lac* promotor. In addition, this vector contains the Lac repressor gene (Mike Berman, personal communication). pRLA41 was constructed by cloning a single 2440-base-pair fragment carrying the 3′ region of the P*spc* operon and a small 5′ segment of the P*α* operon into a unique *Pst* I site at codon 11 of *lacZ* on pMLB1107. In addition to the *prlA4* gene, the inserted fragment carries the upstream ribosomal protein genes specifying L15 and L30 (8). In pRLA41, *prlA4* expression is controlled by the *lac* promoter and thus inducible with isopropyl β-D-thiogalactoside (IPTG).

Maltose Sensitivity Assay. Sensitivity to maltose was quantitated by the disk assay as described (9) except that 10-μl samples of 0.5%, 1%, 2%, 5%, and 10% maltose were applied to the disks. To induce the plasmid-borne *prlA4* gene, 20 mM IPTG was added to the F-top agar. All disk assays were repeated a minimum of 10 times with the appropriate controls.

Pulse-Labeling and Immunoprecipitation. Cells were grown and LamB and MBP synthesis was induced as described (10) except that cells were induced with 0.2% maltose for 60 min. Under this condition, the pleiotropic defects exhibited by maltose-sensitive *lamB–lacZ* fusion strains are minimal. Cell death, for example, is not evidenced until induction times approach 3 hr. Pulse–chase assays and immunoprecipitations were done as described (11).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MBP, maltose-binding protein; IPTG, isopropyl β-D-thiogalactoside.

*To whom reprint requests should be addressed.

Table 1. Maltose-sensitive disk assay

Strain	Genotype	Diameter of sensitivity, mm
	Assayed with 0.5% maltose	
pop3186	*prlA*$^+$ *lamB–lacZ42-1*	9.8
pop3186/pRLA41	*prlA*$^+$ *lamB–lacZ42-1/ prlA4* (uninduced)	5.2
pop3186/pRLA41	*prlA*$^+$ *lamB–lacZ42-1/ prlA4* (induced)	0.3
pop3186/pMLB-1107.2	*prlA*$^+$ *lamB–lacZ42-1/ lacZ* (induced)	10.1
	Assayed with 5.0% maltose	
pop3186	*prlA*$^+$ *lamB–lacZ42-1*	22
SE1073	*prlA*$^+$ *lamB17D–lacZ42-1*	0
BKR73A4	*prlA4 lamB17D–lacZ42-1*	10
SE1073/pRLA41	*prlA*$^+$ *lamB17D–lacZ42-1/prlA4* (uninduced)	0
BKR73A4/pRLA41	*prlA4 lamB17D–lacZ42-1/prlA4* (uninduced)	10
BKR73A4/pRLA41	*prlA4 lamB17D–lacZ42-1/prlA4* (induced)	6
BKR73A4/F'141	*prlA4 lamB17D–lacZ42-1/F' prlA*$^+$	0

Data shown are from a representative series of assays described in the text. Uninduced or induced in parentheses refers to the absence or presence of IPTG in the top agar to vary the level of *prl4* expression.

SDS/PAGE and Autoradiography. Sample electrophoresis was performed by using 10% polyacrylamide gels that were then fluorographed. Autoradiograms were scanned and peaks were integrated as described (10).

RESULTS

Background and Rationale. High-level synthesis of certain LamB–LacZ and MalE–LacZ hybrid proteins is lethal to the cell (7). Several lines of evidence indicate that this lethality is the consequence of a general block in protein export. First, synthesis of these hybrids causes the cytoplasmic accumulation of the precursor form (contains the signal sequence) of many different exported proteins (12, 13). Second, signal sequence mutations that prevent hybrid protein export relieve this lethality and high-level synthesis no longer results in generalized precursor accumulation (7).

Although the molecular mechanism of the block in export is not yet clear, the phenomenon has been studied in some detail in several different laboratories. Results of these studies indicate that the cell recognizes the hybrid protein as an exported protein and initiates the localization process. Export proceeds efficiently to a stage at which the signal sequence is removed by leader peptidase at the periplasmic face of the inner membrane and sequences of the hybrid molecule are accessible to externally added proteases (refs. 14 and 15; B. Rassmussen and T.J.S., unpublished data). Jamming of the export machinery occurs subsequent to signal sequence cleavage but prior to complete translocation of the entire molecule leaving the hybrid stuck in transmembrane fashion. Given this topology, it is reasonable to propose that jamming occurs at the level of the translocator, and although this has not been proven directly, we think it likely. Indeed, Voorhout *et al.* (16) have reached the same conclusion based on immunocytochemical analysis.

If the hybrid protein is incompletely translocated and thus jammed part way through the export apparatus of the inner membrane, then it follows that specific cellular components must become limiting for protein translocation from the cytoplasm under jamming conditions. These components must be those that comprise the part of the export machinery used for the actual translocation reaction.

Overproduction of PrlA Relieves Maltose Sensitivity. The overproduction lethality conferred by *lamB–lacZ* fusions such as 42-1 (contains the signal sequence plus 181 codons of mature LamB) is evidenced phenotypically as maltose sensitivity because maltose induces high-level synthesis of the hybrid protein. Maltose sensitivity can be simply scored by using a disk test analogous to that employed for determining antibiotic sensitivity, and the degree of sensitivity can be quantitated by measuring the diameter of the zone of growth inhibition. Accordingly, this test provides a means to identify conditions in which maltose sensitivity is altered.

We reasoned that increased production of the export component that is rendered rate-limiting by hybrid protein jamming would result in decreased maltose sensitivity. If so, then it should be possible to identify the component in question by introducing the relevant gene on a multicopy plasmid vector into the *lamB–lacZ* fusion strain and scoring maltose sensitivity using the disk assay. We began our studies using *prlA*, as its gene product is the only component of the export machinery that is known to be located within the cytoplasmic membrane, the location of the partially translocated hybrid protein.

To investigate the possible role of PrlA as a limiting factor in hybrid jamming, the maltose-sensitive phenotype of pop3186 (*lamB–lacZ42-1*) was assayed under conditions of PrlA4 overproduction. Overproduction of PrlA4 was attained by using the high-copy plasmid pRLA41, in which expression of the suppressor allele is driven by the *lac* promoter. As can be seen in the upper part of Table 1, the presence of the plasmid greatly reduced sensitivity to maltose under uninduced and induced conditions, resulting in a decrease by a factor of 2 in the diameter of sensitivity to 0.5% maltose under uninduced conditions and almost complete resistance to this concentration of maltose under conditions of *prlA4* induction. The plasmid pMLB1107, lacking the *prlA4* insert, had no effect on maltose sensitivity.

The observation that overproduction of PrlA relieves maltose sensitivity suggests that PrlA is the limiting export component under conditions of high-level hybrid protein synthesis. However, from these data alone, it is impossible to rule out a decrease in maltose sensitivity caused by overproduction of other proteins encoded within the plasmid insert or an indirect effect on other components of the export apparatus caused by the overproduction of PrlA. We do not think the presence of the suppressor allele affects the maltose sensitivity in these experiments since the presence of the *prlA4* mutation in single copy has no known effects on proteins with wild-type signal sequences or on the maltose sensitivity conferred by the fusion with the wild-type signal sequence (data not shown).

The prlA4 Suppressor Can Restore Maltose Sensitivity. To more selectively probe the role of PrlA in hybrid jamming we used a genetic approach combining an export-defective signal sequence mutation with the *prlA4* suppressor allele of *prlA*. Signal sequence mutations such as *lamB17D*, when present in a *lamB–lacZ* gene fusion, block entry of the hybrid protein into the export pathway and cause the hybrid protein to accumulate in the cytoplasm (7). Accordingly, they confer a maltose-resistant phenotype to the fusion strain. As shown in the lower part of Table 1, this phenotype is reversed by the *prlA4* suppressor—i.e., maltose sensitivity is restored. The

decreased maltose sensitivity that is observed under suppressing conditions (strain BKR73A4) relative to the wild-type (strain pop3186) is probably the result of incomplete suppression of the 17D signal sequence mutation by the *prlA4* suppressor. Indeed, strains that carry *lamB–lacZ* hybrid genes with signal sequence mutations that cause a more pronounced export defect than *lamB17D* and/or that are suppressed less well by *prlA4* remain maltose-resistant in the presence of the suppressor (unpublished data).

Maltose Sensitivity Is Recessive in a prlA$^+$/prlA Diploid. Initial characterization of *prlA4* showed that the suppressor is dominant to the wild-type gene in diploid analysis (2). This dominance reflects the fact that the suppressor mutation is a missense mutation that results in a functionally altered gene product (10). An experiment demonstrating this dominance is shown in Fig. 1. Here it can be seen that the signal sequence mutation *lamBS78* blocks export of an otherwise wild-type LamB protein causing the accumulation of the precursor form of the molecule. When the *prlA4* suppressor was provided in trans on the plasmid pRLA41, processing of the LamBS78 precursor to mature product occurs with nearly wild-type kinetics. This result verifies the dominance of *prlA4* and demonstrates that the suppressor gene on the plasmid is fully functional. Indeed, the suppression observed with *prlA4* on the plasmid is better than that observed in haploid strains that carry *prlA4* at the normal chromosomal locus. Presumably this reflects significant basal level production of PrlA from the multicopy plasmid (note that this occurs even in the absence of induction with IPTG). When a similar dominance test (*prlA$^+$/prlA4*) was performed with a strain carrying the *lamB17D–lacZ* fusion, we found that *prlA4* behaved in a recessive manner—i.e., maltose sensitivity was not restored and the diploid strain remained completely maltose-resistant (lower part of Table 1). This result is striking because we predict that the signal sequence mutation present in the gene fusion should be suppressed in the diploid strain. Indeed, in a *prlA4/prlA4* diploid, maltose sensitivity is observed (lower part of Table 1). The maltose sensitivity of the *prlA4/prlA4* diploid indicates that the recessive nature of *prlA4* in the diploid test cannot be explained by simple overproduction of PrlA. In addition, we can conclude that the maltose resistance observed in the *prlA$^+$/prlA4* strain is not due to effects of protein products other than PrlA encoded by pRLA41. Further support for these conclusions comes from the observation that the maltose-sensitive strain BKR73A4 (*lamB17D–lacZ, prlA4*) can be rendered maltose-resistant by the introduction of an F′ factor (F′141) that carries *prlA$^+$* (lower part of Table 1).

Maltose Sensitivity Correlates with Precursor Accumulation. To confirm that the maltose-sensitive and -resistant phenotypes observed in the *lamB–lacZ* fusion strains carrying different alleles of *prlA* accurately reflect the presence or absence of a jammed export complex, we measured the translocation of a representative wild-type protein, MBP, using a standard pulse–chase assay. This assay monitors the rate of processing of the precursor form of MBP to the mature form, a reaction catalyzed by leader peptidase at the periplasmic face of the inner membrane. Normally, translocation and processing are quite rapid; however, if the export apparatus is jammed by the LamB–LacZ hybrid protein, the precursor form of MBP accumulates in the cell.

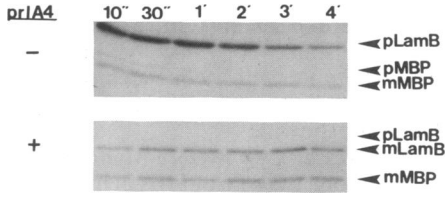


FIG. 1. Effect of the *prlA4* suppressor on the processing of LamB with an altered signal sequence. Cells were pulse labeled with [^{35}S]-methionine and immunoprecipitations were performed. The strain BKR104 was assayed with the vector pMLB1107.2 (*Upper*) and the plasmid carrying the suppressor allele of *prlA*, pRLA41 (*Lower*). In both cases, plasmid-borne genes are not induced with IPTG. "p" denotes the precursor form of the designated protein; "m" designates the mature product. Time points shown indicate the time of the trichloroacetic acid precipitation following the addition of chase.

The pulse–chase data for the various strains following a 60-min induction with maltose are presented in Fig. 2. The degree of precursor accumulation in the parent strain pop3186 (*lamB–lacZ prlA$^+$*) is shown in Fig. 2*A*. Under these conditions the precursor form of MBP predominates up to the 2-min time point. When the 17D signal sequence mutation is present on the hybrid gene (strain SE1073, Fig. 2*B*), precursor accumulation is virtually abolished as would be expected since the hybrid protein fails to enter the export pathway. When the suppressor allele *prlA4* is present together with the *lamB17D–lacZ* fusion (strain BKR73A4) precursor MBP accumulates to the same degree as that caused by the wild-type fusion in the *prlA$^+$* background (compare Fig. 2 *C* with *A*). This finding, in combination with the maltose-sensitive phenotypes described above, indicates that the block in export normally relieved by the signal sequence mutation has been restored in the presence of the suppressor allele *prlA4*. Fig. 2*D* shows the results obtained with the *lamB17D–lacZ prlA$^+$/prlA4* diploid. Again, as predicted from the maltose-resistant phenotype of the strain, nearly normal export of MBP is observed. Thus, the maltose-sensitive phenotype correlates with precursor accumulation and the recessive nature of *prlA4* with respect to this phenotype is confirmed.

A Plausible Model. To account for the seemingly paradoxical behavior of *prlA4* in diploid analysis (compare Figs. 1 and 2), we offer the explanation shown in Fig. 3. According to this view, protein translocation across the inner membrane occurs via PrlA or a complex that includes PrlA. LacZ hybrid proteins interfere with the translocation reaction by direct interaction with, and sequestration of, PrlA (Fig. 3*A*). Signal sequence mutations would prevent this interaction by preventing entry of the hybrid protein into the export pathway (Fig. 3*B*) and consequently, no general block to export would be predicted. When the suppressor allele, *prlA4*, is present, recognition of the fusion protein by PrlA would be restored and a general block to the export of noncytoplasmic proteins would be observed (Fig. 3*C*). Fig. 3*D* shows the predicted situation for a strain that carries the suppressor and wild-type alleles of *prlA*. Under these conditions the hybrid protein with an altered signal sequence may sequester the suppressor PrlA4, creating a jammed complex, as is the case in Fig. 3*C*. However, because the wild-type PrlA fails to recognize the mutant hybrid, it would remain free to participate in general protein export, as is the case in Fig. 3*B*. The continued availability of wild-type PrlA would allow expression of a

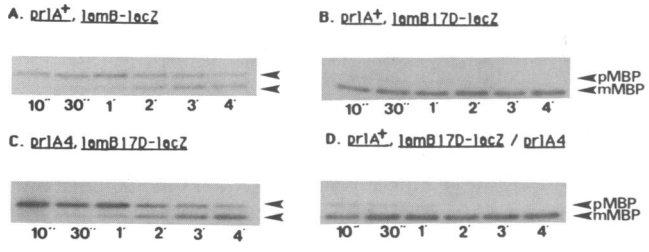


FIG. 2. Processing of MBP following induction of the *LamB–LacZ* fusions with maltose. The samples were prepared as described in the legend to Fig. 1. Strains are as follows: pop3186 (*A*); SE1073 (*B*); BKR73A4 (*C*); SE1073/pRLA41 (*D*).

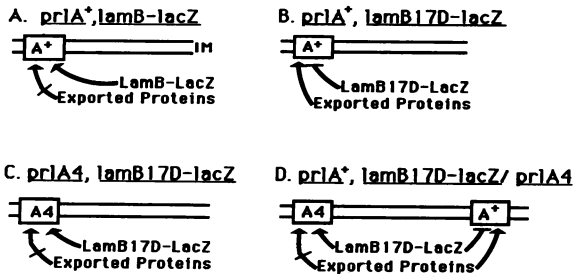


FIG. 3. Model for the maltose sensitivity phenotypes observed in the presence of *lamB–lacZ* fusion alleles when combined with wild-type and suppressor alleles of *prlA*. →, Recognition by PrlA; —|, no recognition by PrlA; ↛, recognition by PrlA blocked due to fusion jamming. The corresponding precursor accumulation assays are shown in Fig. 2. IM, inner membrane.

maltose-resistant phenotype and thus account for the recessive nature of the suppressor with respect to this phenotype. Moreover, because this model separates jamming from the suppression of signal sequence mutations, it can also account for the dominant suppression of signal sequence mutations carried on an otherwise wild-type *lamB* gene (Fig. 1).

The LamB17D–LacZ Hybrid Protein Selectively Blocks Export at PrlA4. The model shown in Fig. 3 predicts that the LamB17D–LacZ hybrid protein would specifically block export of proteins with mutant signal sequences at PrlA4 but that proteins that can be translocated via the wild-type PrlA would be unhindered. To test this hypothesis we took advantage of a second signal sequence mutation in *lamB*, *lamBS78*, which is not exported in wild-type, *prlA*$^+$, strains but is efficiently exported in a *prlA*$^+$/*prlA4* diploid (Fig. 1). Strain BKR7313/pRLA41 harbors the *lamB17D–lacZ* fusion and the *lamBS78* allele and both the wild-type and suppressor alleles of *prlA*. By using this strain we can monitor the effects of the LamB17D–LacZ hybrid protein on export via the PrlA4 suppressor (by assaying processing of LamBS78) and export via wild-type PrlA (by assaying the processing of MBP). This experimental scheme is shown in Fig. 4*A*.

The top autoradiogram in Fig. 4*B* shows the data obtained in a pulse–chase assay of LamBS78 and MBP export following a 60-min induction with maltose of strain BKR7313/pRLA41 (*lamBS78 lamB17D-lacZ prlA*$^+$/*prlA4*). Even at the 4-min time point no processing of LamBS78 was observed, indicating that LamBS78 did not reach the periplasmic face of the inner membrane under these conditions. This is in contrast to MBP, whose processing is apparent at 10 sec and complete by the 2-min time point in the same strain under identical conditions. It is also in marked contrast to the rapid processing of LamBS78 that is observed in the isogenic strain lacking the fusion (BKR104/pRLA41, Fig. 1). Taken together, these results support the model presented in Fig. 4*A* and demonstrate that the LamB17D–LacZ hybrid protein specifically blocks export at PrlA4.

Strain BKR7313/pRLA41 is diploid for two different chromosomal regions: *lamB* (*lamBS78 lamB17D–lacZ*) and *prlA* (*prlA*$^+$ *prlA4*). Accordingly, appropriate controls for the experiment described in the preceding paragraph involve strains in which each of these regions is altered in turn. In the first control, a strain that is isogenic except for the presence of *lamB*$^+$ instead of *lamBS78* was used. In this strain (BKR7310/pRLA41), mature LamB and MBP were observed at the 10-sec time point with nearly complete processing of both proteins by 2 min (Fig. 4*B*). As expected, this strain is maltose-resistant and we presume that the slight export defect observed for both proteins relates to the inherent leakiness of the 17D signal sequence mutation (10). The second control involves a strain isogenic to BKR7313/pRLA41 except that both *prlA* alleles are *prlA4*. Pulse–chase

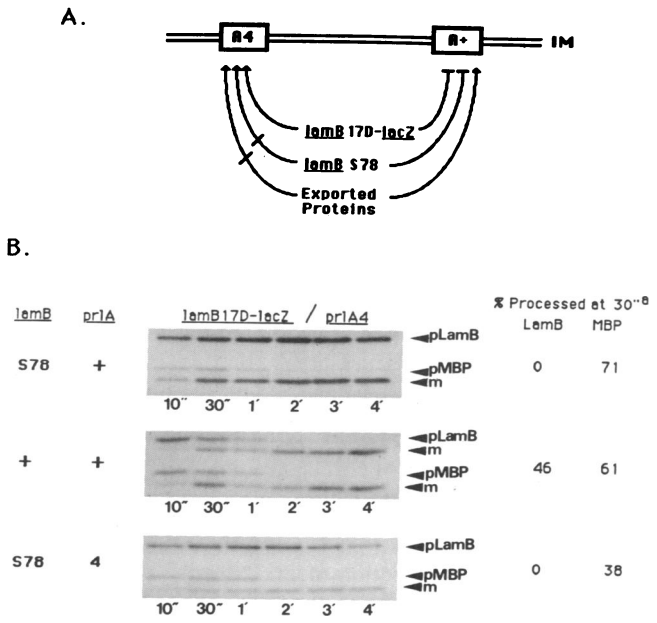


FIG. 4. Experimental design and assay for selective precursor accumulation in strains diploid for *prlA* and *lamB*. (*A*) The predicted situation following maltose induction of the strain BKR7313/pRLA41. Symbols are as described in the legend to Fig. 3. (*B*) The samples were prepared as described in the legend to Fig. 1. Strains assayed are (top to bottom) BKR7313/pRLA41, BKR7310/pRLA41, and BKR7314/pRLA41. The *lamB* and *prlA* alleles indicated on the left are chromosomal; the *lamB17D–lacZ* fusion is present on the phage λSE73 (7) and *prlA4* is carried on the plasmid pRLA41 as indicated across the top of the figure. Superscript a, % of LamB and MBP processed to the mature form was determined by densitometry.

assays reveal a complete block to LamBS78 export and a decrease by a factor of 2 in the amount of mature MBP that is present at the 30-sec time point relative to BKR7313/pRLA41. Possible explanations for the differential sensitivity of LamBS78 and wild-type MBP to jamming by the LamB17D–LacZ hybrid protein are considered in the *Discussion*. It is of note, however, that if induction time (with maltose) is increased to 2 hr instead of 1 hr, no processing of MBP is seen before the 2-min time point, and it is reasonable to expect that increased levels of the mutant hybrid protein are required to effectively jam a strain that contains higher amounts of PrlA4 (see Table 1). In any event, this strain is maltose-sensitive, as predicted, since all available export sites can be jammed.

DISCUSSION

A central question posed by the process of protein export is the mechanism by which the large hydrophilic molecules are translocated across a hydrophobic lipid bilayer. One necessary step in the elucidation of this mechanism is the identification of the cellular components that participate in this translocation reaction. In the *E. coli* system, genetic and biochemical approaches have been successfully used to address this issue. However, most of the cellular components identified are soluble cytoplasmic proteins and, in the few cases where a function can be assigned, this role appears to involve antifolding and perhaps piloting activities (17, 18). Because PrlA is the only truly integral inner membrane protein yet identified and because a role in translocation would seem to necessitate such a cellular location, it is commonly assumed that PrlA is a component of the translocator. The data we present provide evidence that this assumption may be correct.

The approach we used to probe the nature of the translocator is genetic and rests heavily on *lamB–lacZ* gene fusions and *prlA4*, a dominant suppressor of signal sequence mutations. The *lamB–lacZ* gene fusion used in these studies is known to jam the export apparatus of the cell, causing a generalized accumulation of precursors to exported proteins (7, 12). Evidence summarized in "Background and Rationale" suggests that this jamming occurs at the level of the translocator, leaving the hybrid protein stuck in a transmembrane fashion.

Because PrlA is clearly an integral inner membrane protein and since *prlA4* is a dominant suppressor of signal sequence mutations, we can use this allele to confer selectivity to that portion of the export machinery that is located in the membrane. Such selectivity is possible because PrlA4 will recognize proteins with altered signal sequences as being destined for export, whereas the wild-type counterpart, PrlA, will not. This selectivity permits meaningful interpretation of the diploid analysis described.

Two lines of evidence are presented to show that PrlA is the cellular component that is sequestered by hybrid jamming. First, overproduction of PrlA decreases the maltose sensitivity of *lamB–lacZ* fusion strains. Second, and perhaps more convincingly, *prlA4* will restore the maltose-sensitive phenotype conferred by a *lamB–lacZ* fusion carrying a signal sequence mutation. However, this restoration is masked if the cell also contains a *prlA*$^+$ allele. Because *prlA4* is a dominant suppressor of signal sequence mutations we conclude that its recessive nature with respect to maltose sensitivity reflects the fact that wild-type PrlA is insensitive to jamming by the mutant hybrid protein and therefore operates in normal fashion for protein export. Indeed, we can show a selective block in protein export via PrlA4 with little effect on PrlA$^+$ in the *prlA*$^+$*/prlA4* diploid strain (Fig. 4). Given these results and the topology of the hybrid protein discussed in "Background and Rationale," we propose that PrlA participates directly in the translocation reaction. Biochemical analysis will ultimately be required to verify our proposal directly.

Our results further suggest that PrlA is the only cellular component that becomes rate-limiting for protein export under conditions of hybrid jamming. This does not mean that PrlA is the only component of the translocator, although this is one intriguing possibility. Other proteins, if present in sufficient quantities, could be part of the translocator complex as well. It has been observed that synthesis of SecA is increased substantially under jamming conditions (19) and therefore may always be present in excess. Such regulation could exist for other components of the translocator as well.

We have also found that the export of proteins with altered signal sequences via PrlA4 is more sensitive to the jamming effects of hybrid proteins than is the export of wild-type proteins (Fig. 4). Several explanations for this result can be offered; we present but one. It seems likely that signal sequences perform multiple functions during the export process (10) and the signal sequence mutations we employ may well be multiply defective. For example, the signal sequence may be recognized by a cytoplasmic export factor. Since the *prlA4* suppressor would not restore interaction of the mutant signal sequence with such a cytoplasmic factor, the export of the mutant protein under suppressing conditions would be compromised and therefore more susceptible to interference from the hybrid molecule.

Genetic analysis of export components such as PrlA is hindered because of the essential nature of the function performed. Accordingly, the types of mutations that can be studied are limited to conditional lethals or suppressor alleles. Our results show that in *prlA*$^+$*/prlA4* diploids, the function of PrlA4 can be examined specifically by following the export of proteins (including LacZ hybrid proteins) with altered signal sequences. Since PrlA4 is not essential under these conditions, the types of mutations that can be identified and studied are greatly expanded and we are hopeful that this will provide a means to probe the complex functions of this protein in greater detail.

We thank Scott Emr, Nancy Trun, and members of the laboratory for helpful discussions and comments throughout the course of this work. We also thank Mike Berman for the gift of the plasmid pMLB1107 and Johann Lim for the construction of pRLA41. This research was supported by Public Health Service Grant GM34821.

1. Benson, S., Hall, M. & Silhavy, T. (1985) *Annu. Rev. Biochem.* **54,** 101–134.
2. Emr, S., Hanley-Way, S. & Silhavy, T. J. (1981) *Cell* **23,** 79–88.
3. Ito, K. (1984) *Mol. Gen. Genet.* **197,** 204–208.
4. Fandl, J. & Tai, P. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7448–7452.
5. Cabelli, R., Chen, L., Tai, P. C. & Oliver, D. (1988) *Cell* **55,** 683–692.
6. Silhavy, T. J., Berman, M. & Enquist, L. (1984) *Experiments with Gene Fusions* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
7. Emr, S. & Silhavy, T. J. (1980) *J. Mol. Biol.* **141,** 63–90.
8. Schultz, J., Silhavy, T. J., Berman, M., Fiil, N. & Emr, S. (1982) *Cell* **31,** 227–235.
9. Benson, S., Bremer, E. & Silhavy, T. J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3830–3834.
10. Stader, J., Benson, S. & Silhavy, T. J. (1986) *J. Biol. Chem.* **261,** 15075–15080.
11. Trun, N. & Silhavy, T. J. (1987) *Genetics* **116,** 513–521.
12. Bassford, P., Silhavy, T. J. & Beckwith, J. (1979) *J. Bacteriol.* **139,** 19–31.
13. Hall, M., Schwartz, M. & Silhavy, T. J. (1982) *J. Mol. Biol.* **156,** 93–112.
14. Wolfe, P., Wickner, W. & Goodman, J. (1983) *J. Biol. Chem.* **258,** 12073–12080.
15. Rasmussen, B., Bankaitis, V. & Bassford, P. (1984) *J. Bacteriol.* **160,** 612–617.
16. Voorhout, W., De Kroon, T., Leunissen-Bijvelt, J., Verkleij, A. & Tommassen, J. (1988) *J. Gen. Microbiol.* **134,** 599–604.
17. Collier, D., Bankaitis, V., Weiss, J. & Bassford, P. (1988) *Cell* **53,** 273–283.
18. Crooke, E., Guthrie, B., Lecker, S., Lill, R. & Wickner, W. (1988) *Cell* **54,** 1003–1011.
19. Oliver, D. & Beckwith, J. (1982) *Cell* **30,** 311–319.

J. Mol. Biol. (1959) **1**, 165-178

The Genetic Control and Cytoplasmic Expression of "Inducibility" in the Synthesis of β-galactosidase by *E. Coli*†

ARTHUR B. PARDEE‡, FRANÇOIS JACOB AND JACQUES MONOD

Institut Pasteur, Paris and University of California, Berkeley, California, U.S.A.

(*Received 16 March 1959*)

A number of extremely closely linked mutations have been found to affect the synthesis of β-galactosidase in *E. coli*. Some of these (*z* mutations) are expressed by loss of the capacity to synthesize active enzyme. Others (*i* mutations) allow the enzyme to be synthesized constitutively instead of inducibly as in the wild type. The study of galactosidase synthesis in heteromerozygotes of *E. coli* indicates that the *z* and *i* mutations belong to different cistrons. Moreover the constitutive allele of the *i* cistron is recessive over the inducible allele. The kinetics of expression of the i^+ (inducible) character suggest that the *i* gene controls the synthesis of a specific substance which represses the synthesis of β-galactosidase. The constitutive state results from loss of the capacity to synthesize active repressor.

1. Introduction

Any hypothesis on the mechanism of enzyme induction implies an interpretation of the difference between "inducible" and "constitutive" systems. Conversely, since specific, one-step mutations are known, in some cases, to convert a typical inducible into a fully constitutive system, an analysis of the genetic nature and of the biochemical effects of such a mutation should lead to an interpretation of the control mechanisms involved in induction. This is the subject of the present paper.

It should be recalled that the metabolism of lactose and other β-galactosides by intact *E. coli* requires the sequential participation of two distinct factors:

(1) The galactoside-permease, responsible for allowing the entrance of galactosides into the cell.

(2) The intracellular β-galactosidase, responsible for the hydrolysis of β-galactosides.

Both the permease and the hydrolase are inducible in wild type *E. coli*. Three main types of mutations have been found to affect this sequential system:

(1) $z^+ \rightarrow z^-$: loss of the capacity to synthesize β-galactosidase;

(2) $y^+ \rightarrow y^-$: loss of the capacity to synthesize galactoside-permease;

(3) $i^+ \rightarrow i^-$: conversion from the inducible (i^+) to the constitutive (i^-) state.

The $i^+ \rightarrow i^-$ mutation always affects *both* the permease and the hydrolase. All these mutations are extremely closely linked: so far all independent occurences of each of these types have turned out to be located in the "*Lac*" region of the *E. coli* K 12 chromosome. However, the mutations appear to be *independent* since all the different phenotypes resulting from combinations of the different alleles are observed (Rickenberg, Cohen, Buttin & Monod, 1955; Cohen & Monod, 1957; Cohn, 1957).

† This work has been aided by a grant from the Jane Coffin Childs Memorial Fund.

‡ Senior Postdoctoral Fellow of the National Science Foundation (1957–58).

It should also be recalled that conjugation in *E. coli* involves the injection of a chromosome from a ♂ (Hfr) into a ♀ (F^-) cell, and results generally in the formation of an incomplete zygote (merozygote) (Wollman, Jacob & Hayes, 1956). Recombination between ♂ and ♀ chromosome segments does not take place until about 60 to 90 min after injection; moreover segregation of recombinants from heteromerozygotes occurs only after several hours, thus allowing ample time for experimentation.

In order to study the interaction of these factors, their expression in the cytoplasm and their dominance relationships, we have developed a technique which allows one to determine the kinetics of β-galactosidase synthesis in merozygotes of *E. coli*, formed by conjugation of ♂ (Hfr) and ♀ (F^-) cells carrying different alleles of the factors *z*, *y* and *i* (Pardee, Jacob & Monod, 1958). Before discussing the results obtained with this technique, we shall summarize some preliminary observations on the genetic structure of the "*Lac*" region in *E. coli* K 12.

2. Materials and Methods

(a) *Bacterial strains*

A ♂ (Hfr) strain (no. 4,000) of *E. coli* K 12 was used in most experiments. It was derived from strain 58,161 F^+, and was selected for early injection of the "*Lac*" marker (Jacob & Wollman, 1957). This strain is streptomycin sensitive (S^s), requires methionine for growth and carries the phage λ. A second Hfr strain (no. 3,000), isolated by Hayes (1953), was used in some experiments. This strain is S^s, requires vitamin B_1, and does not carry λ prophage. Other Hfr strains carrying mutations for galactosidase (*z*), inducibility-constitutivity (*i*), and permease (*y*) were isolated from the Hayes strain after u.v. irradiation. These markers were also put into ♀ (F^-) strains, by appropriate matings and selection of the desired recombinants.

A synthetic medium (M 63) was commonly used. It contained per liter: 13·6 g KH_2PO_4 2·0 g $(NH_4)_2SO_4$, 0·2 g $MgSO_4 . 7H_2O$, 0·5 mg $FeSO_4 . 7 H_2O$, 2·0 g glycerol, and KOH to make pH 7·0. If amino-acids were required, they were added at a concentration of 10 mg/l. of the L-form. For mating experiments, the above stock medium was adjusted to pH 6·3 and vitamin B_1 (0·5 mg/l.) was added prior to use. Aspartate (0·1 mg/ml.) was generally added at the time of mating, according to Fisher (1957).

(b) *Mating experiments*

The desired volume of fresh medium was inoculated with an overnight culture (grown in the same medium) to an initial density of approximately 2×10^7 bacteria/ml. This culture was aerated by shaking at 37°C in a water bath. Turbidity was measured from time to time; and when the density reached 1 to 2×10^8 bacteria/ml., the experiment was started. Usually small volumes of ♂ and ♀ bacteria were mixed in a large Erlenmeyer flask, with the ♀ strain in excess (e.g. 3 ml. ♂ plus 7 ml. ♀ in a 300 ml. flask). The mixed bacteria were agitated very gently so that the motion of the liquid was barely perceptible. From time to time samples were removed for enzyme assay and plating on selective media, usually lactose-B_1-streptomycin agar, for measurement of recombinants. Under these conditions, in a mating of ♂ z^+Sm^s by ♀ z^-Sm^r, up to 20 % of the ♂ population formed z^+Sm^r recombinants (as tested by selection on lactose-streptomycin agar). More often 5 to 10 % recombinants were found.

Streptomycin (Sm)† was used in many mating experiments, to block enzyme synthesis by z^+Sm^s ♂ cells. Controls showed that the synthesis of β-galactosidase was blocked in these strains immediately upon addition of 1 mg/ml. of Sm. Incorporation of ^{35}S from $^{35}SO_4^=$ as well as increase of turbidity were also suppressed by this treatment. This concentration of Sm had no effect on Sm-resistant (Sm^r) mutants. In some experiments, virulent phage (T6) was used to kill the ♂ cells, thus preventing remating.

† The following abbreviations are used in this paper:

Sm = streptomycin
IPTG = *iso*propyl-thio-β-D-galactoside
ONPG = *o*-nitrophenyl-β-D-galactoside
TMG = methyl-thio-β-D-galactoside

It should be noted that if streptomycin was added initially, it significantly reduced the number of recombinants (e.g., 75 % fewer colonies were formed on lactose-B_1-streptomycin plates after 80 min mating in the presence of 1 mg/ml. streptomycin) relative to mating in the absence of streptomycin; but the antibiotic had little effect on enzyme formation by zygotes if added at the commencement of the experiment or after the z^+ locus had been injected.

When galactosidase synthesis had to be induced in zygotes, *iso*propyl-thio-β-D-galactoside (IPTG) was used at 10^{-3}M, a concentration at which this inducer is known to be active even in the absence of permease (Rickenberg *et al.*, 1956).

(c) *Recombination studies*

The blender technique of Wollman & Jacob (1955) was used to determine the times of penetration of markers into the zygotes. It should be noted that this treatment reduces enzyme-forming capacity in zygotes by 30 to 60 %. Recombinant colonies, selected on appropriate selective media, were restreaked on the selector medium and replica plating was used to determine unselected characters. Tests for galactosidase synthesis (with or without induction) were performed on maltose-synthetic agar plates with or without IPTG, using filter paper impregnated with ONPG, according to Cohen-Bazire & Jolit (1953).

Transductions were performed with phage 363, according to Jacob (1955).

(d) *β-galactosidase assay*

For this enzyme assay, 1 ml. aliquots of culture were pipeted into tubes containing 1 drop of toluene. The tubes were shaken vigorously and were incubated for 30 min at 37°C. They were then brought to 28°C; 0·2 ml. of a solution of M/75 *o*-nitrophenyl-β-D-galactoside in M/4 sodium phosphate (pH 7·0) was added, and the tubes were incubated a measured time, until the desired intensity of color had developed. The reaction was halted by addition of 0·5 ml. of 1 M-Na_2CO_3, and the optical density was measured at 420 mμ with the Beckman spectrophotometer. A correction for turbidity could be made by multiplying the optical density at 550 mμ by 1·65 and subtracting this value from the density at 420 mμ. One unit of enzyme is defined as producing 1 mμ-mole *o*-nitrophenol/minute at 28°C, pH 7·0. The units of enzyme in the sample can be calculated from the fact that 1 mμ-mole/ml. *o*-nitrophenol has an optical density of 0·0075 under the above conditions (using 10 mm light-path).

(e) *Chemicals*

o-nitrophenyl-β-D-galactoside (ONPG), methyl-thio-β-D-galactoside (TMG) and *iso*propyl-thio-β-D-galactoside (IPTG) were synthesized at the Institut Pasteur by Dr. D. Türk. Other chemicals were commercial products.

3. Genetic Structure of the "Lac" Region

Figure 1 presents the structure of the "*Lac*" region, as it can be sketched from the data available at present. This complex locus, as established long ago by Lederberg (1947) and confirmed by the blender experiments of Wollman & Jacob (1955), lies at about equal distances from the classical markers *TL* and *Gal*. The closest known markers are *Proline* (left) and *Adenine* (or *T6*) (right). As shown in the map, the several (about 10) occurrences of the y^- mutation all lie together probably at the left of the segment, while the different z^- mutations and the i^- mutant are packed together at the other end. No attempt has been made to establish the order of individual y^- mutations. The order of the z^- mutations relative to each other and to the i^- marker is unambiguously established, as shown, except for the z_U^- mutation, whose position is largely undetermined. Several independent occurrences of the i^- mutation have been isolated. They all appear to be closely linked to the i_3^- marker, but they have not been mapped, for lack of adequate methods of selection i^+ recombinants. The evidence for this structure is briefly as follows:

(1) The frequency of recombination between z and y mutations is very low:

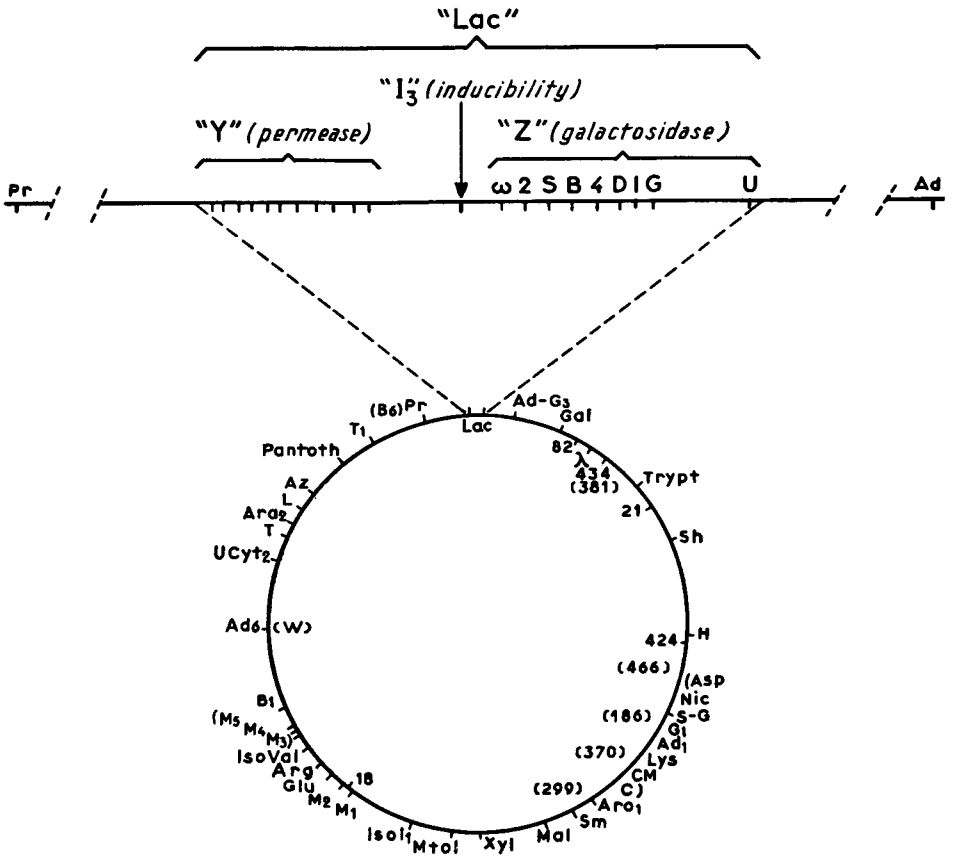


FIG. 1. Fine structure of the "*Lac*" segment.

The "*Lac*" segment is shown enlarged and positioned with respect to the rest of the *E. coli* K 12 linkage group for which the circular model (Jacob & Wollman, 1958) has been adopted.

roughly 1/100th of the frequency of recombination between *TL* and *Gal*. The frequency of recombination between individual *z* markers is about one order of magnitude lower.

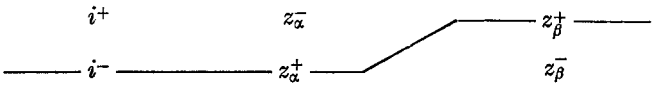
(2) When y^+z^+ recombinants are selected (by growth on lactose-agar) in crosses of the type:

$$y^+i^-z^- \times y^-i^+z^+$$

the i^+ marker remains associated with z^+ 85 % of the time.

(3) The frequency of cotransduction of *i* with *z* (selecting for z^+ alone) is very high (> 90 %), while the frequency for *i* and *y* is also high, although definitely lower (about 70 %). (These data are somewhat ambiguous, because of the heterogeneity of the clones resulting from a transduction.)

(4) The selection of z^+ recombinants in crosses involving different z^- mutants, and *i* as unselected marker, invariably results in about 90 % of the progeny being either i^- or i^+, depending on the particular z^- mutants used. Assuming this result to be due to the position (left or right) of *i* with respect to the *z* group:



a linear order can be established, without contradictions, for the eight markers shown. This however leaves an ambiguity as to whether *i* lies *between* the *y* and the *z* groups, or outside.

Let us emphasize that this sketch of the *Lac* region is preliminary and very incomplete, and that the results concerning the relationships of certain markers are not understood. For instance, the z^-_U marker recombines rather freely with all the other mutants shown (both *y* and *z*) yet, by cotransduction tests, it is closely linked to *i* (25 % cotransduction). It should also be mentioned that certain of the z^- mutants (z^-_ω ; z^-_2; z^-_G) have apparently lost the capacity to synthesize *both* the galactosidase *and* the permease. Yet these mutations do not seem to be deletions. We shall not attempt, here, to interpret this finding, since we shall center our attention on the interaction between the *i* marker and the *z* region.†

A question which should now be considered is whether we may regard the *z* region as possessing the specific structural information concerning the galactosidase molecule. The fact that so far all the independent mutations resulting in loss of the capacity to synthesize galactosidase were located in this region might not constitute sufficient evidence‡. However, it has been found by Perrin, Bussard & Monod (1959, in preparation) that several of the z^- mutants synthesize, instead of active galactosidase, an antigenically identical, or closely allied, protein. Moreover several of these mutant proteins are different from one another by antigenic and other tests. These findings appear to prove that the *z* region indeed corresponds to the "structural" genetic unit for β-galactosidase.

4. β-Galactosidase Synthesis by Heteromerozygotes

(a) *Preliminary experiments*

The feasibility and significance of experiments on the expression and interaction of the *z*, *y* and *i* factors depended primarily on whether *E. coli* merozygotes are physiologically able to synthesize significant amounts of enzyme very soon after mating. It was equally important to determine whether the mating involved any cytoplasmic mixing. These questions were investigated in a series of preliminary experiments.

Since the physical separation of *E. coli* zygotes from unmated or exconjugant parent cells cannot be achieved at present, test conditions must be set up, such that the zygotes only, but not the parents, can synthesize the enzyme. This is obtained when the following mating:

$$♂\ z^+y^+i^+Sm^s \times ♀\ z^-_s y^+i^+Sm^r \qquad \text{(A)}$$

is performed in the presence of inducer (IPTG) and of 1 mg/ml. of streptomycin. The ♀ lack the z^+ factor; the ♂ are inhibited by streptomycin (cf. Methods); the zygotes are not, because they inherit their cytoplasm from the ♀ cells (see below

† Interaction of *i* with the *y* region is of course equally interesting, but since determinations of activity are much less sensitive with the galactoside-permease than with the galactosidase, we have used the latter almost exclusively.

‡ In addition to the mutants shown on Fig. 1, 20 other galactosidase-negative mutants, as yet unmapped, have been found to belong to the same segment by contransduction tests. None was found outside. Lederberg *et al.* (1951), however, have isolated some lactose-"non-fermenting" mutants (as tested on EMB-lactose agar) which are located at other points on the *E. coli* chromosome. In our hands, one of these mutants (Lac^-_3) formed normal amounts of both galactosidase and galactoside-permease (although it did form white colonies on EMB-lactose). Another one (Lac^-_7) formed reduced, but significant, amounts of both. A third (Lac^-_2) which is a galactosidase-negative, appears to belong to the "*Lac*" segment, by cotransduction tests.

pages 170 and 171), and because the type of ♂ used transfers the Sm^s gene to only a very small percentage of the cells. Under these conditions, enzyme is formed in the mated population with a time course and in amounts showing that the synthesis can be due only to zygotes having received the z^+ factor. Figure 2 shows the

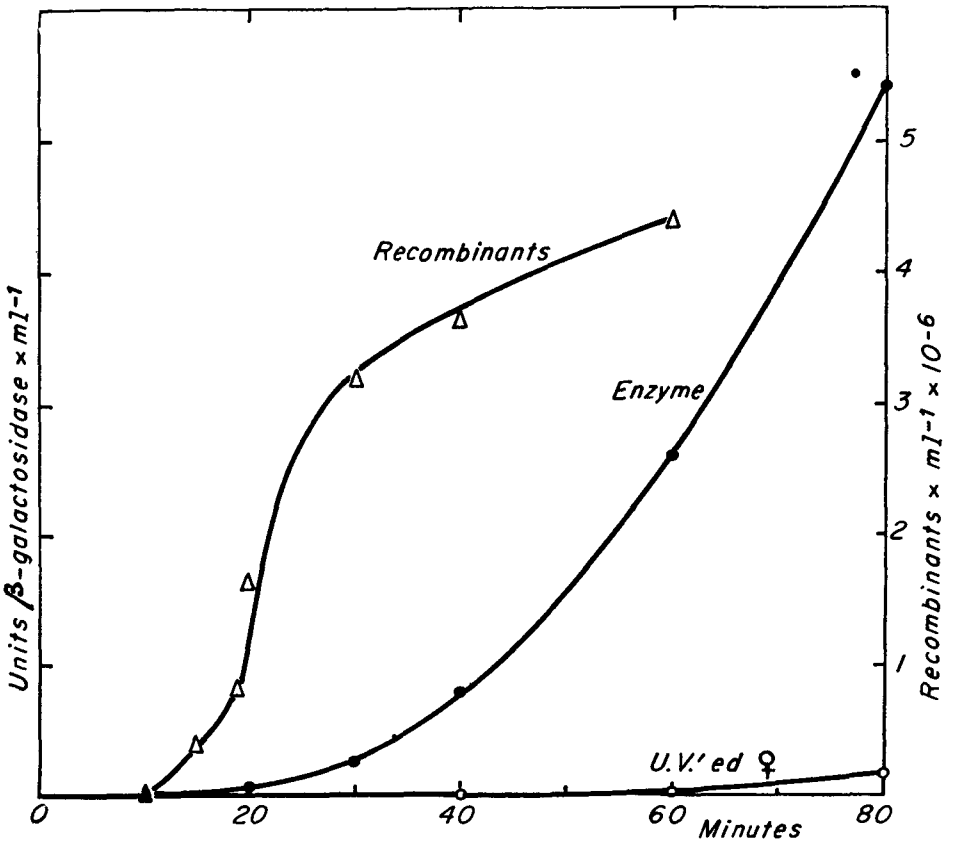


FIG. 2. Enzyme formation and appearance of recombinants in mating A.
Mating in presence of streptomycin (1 mg/ml.) and IPTG (10^{-3}M). A control with u.v.-treated ♀ cells (0·01 % survival) is shown. Recombinants (z^+Sm^r) selected by plating on Sm-lactose agar after blending (separate experiment with the same ♂ culture).

kinetics of galactosidase accumulation, compared with the appearance of z^+Sm^r recombinants, determined on aliquots of the same population (cf. Methods). The latter curve corresponds, as shown by Wollman & Jacob (1955), to the distribution of times of penetration of z^+ genes in the zygote population. It will be remarked that enzyme synthesis commences just within a few minutes after the first z^+ genes enter into zygotes. Assuming that the number of zygotes having received a z^+ gene is 4 to 5 times the number of recovered z^+Sm^r recombinants, and taking into account the fact that normal cells are on the average trinucleate (i.e., have three z^+ genes), the rate of enzyme synthesis per injected z^+ appears nearly normal.

This rapid expression of the z^+ factor poses the problem whether cytoplasmic constituents are injected from the ♂ into the zygote. This already appeared unlikely from the previous observations of Jacob & Wollman (1956). We reasoned that if there occurred any significant cytoplasmic mixing, such a mixing should allow the

♂ cells to feed the ♀ cells with any small metabolites which the ♂ had and the ♀ lacked. This condition is obtained in the following mating:

$$♂\ z^+Sm^s\ maltose^+ \times ♀\ z^-Sm^r\ maltose^-$$

if it is performed in presence of maltose as sole carbon source, using a ♂ which virtually does not inject the *maltose*$^+$ gene. It results in a very strong inhibition of enzyme synthesis (and recombinant formation) showing that the ♂ cannot effectively

TABLE 1

Enzyme formation in nutritionally deficient zygotes

Deficiency	Rate of enzyme formation †			Mean % inhibition of recombinant formation
	Control	Deficient	Mean % inhibition	
Carbon source ‡	1·6	0·4	73	75
	0·66	0·20		
Arginine §	0·28	0·02	96	65
	0·36	0·01		

† Units of enzyme × hr^{-1}.

‡ ♂ z^+Sm^s *maltose*$^+$ × ♀ z^-Sm^r *maltose*$^-$ mated in presence of inducer and Sm, with glycerol plus maltose (control) or maltose as sole carbon source.

§ ♂ z^+Sm^s *Arg*$^+$ × ♀ z^-Sm^r *Arg*$^-$ mated in presence of inducer and Sm with and without arginine (10 μg/ml.).

feed the ♀. An even stronger effect is observed when the ♀ requires arginine, the ♂ not, and mating takes place in absence of arginine (again on condition that the *Ar*$^+$ gene is not injected by the ♂) (Table 1). These observations indicate that even small molecules do not readily pass from the ♂ into the ♀ cell during conjugation.†

It therefore appears that cytoplasmic fusion or mixing does not occur to an extent which might allow cross-feeding. That the contribution of the ♂ is exclusively genetic, and does not involve cytoplasmic constituents of a nature, or in amounts, significant for our purposes, is however only proved by the results of the opposite matings, which we shall consider in the next section.

(b) *Expression and interaction of the alleles of the z and i factors*

We should first consider which of the alleles of the *z* factors are dominant, and whether they all belong to a single cistron. Experiments of the type described above (mating A) were performed with each of the eight z^- mutants, used as ♀ cells, receiving a z^+ from the ♂. Enzyme was synthesized to similar extents in all cases, showing that the z^- mutants in question were all recessive. Each of the mutants was also mated (as ♂) to a z^- ♀. No enzyme was synthesized by any of these double recessive heterozygotes where the mutations were in the *trans* position

† However such leakage may occur when the concentration of a compound is exceptionally high in the ♂. This happens when a ♂ with the constitution $z^-i^-y^+$ is used in the presence of lactose. The constitutive permease then may concentrate lactose up to 20 % of the cells' dry-weight (Cohen & Monod, 1957). Adequate tests have shown that this lactose does flow from the ♂ into a permease-less ♀ during conjugation.

$$\frac{z^+_\alpha z^-_\beta}{z^-_\alpha z^+_\beta}$$

showing that all the (tested) z^- mutants belong to the same cistron as defined by Benzer (1957).

The next and most critical problem is whether the z and i factors also belong to the same unit of function (gene or cistron) or not. Let us recall that cells with the constitution z^+i^+ synthesize enzyme in presence of inducer only, while z^+i^- cells synthesize enzyme without induction, and z^-i^+ or z^-i^- cells do not synthesize enzyme under any condition. The extremely close linkage of z and i mutations suggests that they may belong to the same unit. If this were so, they would not be able to interact through the cytoplasm, but could act together only when in *cis* position within the same genetic unit. The heterozygote, z^+i^+/z^-i^- would then be expected not to synthesize galactosidase constitutively.

In order to test this expectation, the following mating:

$$\text{♂ } z^+i^+ \times \text{♀ } z^-_2 i^-_3 \qquad \text{(B)}$$

was performed *in absence of inducer*. The ♂ cannot synthesize enzyme, because they are i^+. The ♀ cannot because they are z^-. The zygotes however do synthesize enzyme (Fig. 3): during the first hour following mating the synthesis is, if anything, even more rapid and vigorous than when both parents are i^+ and inducer is used, as in mating (A).

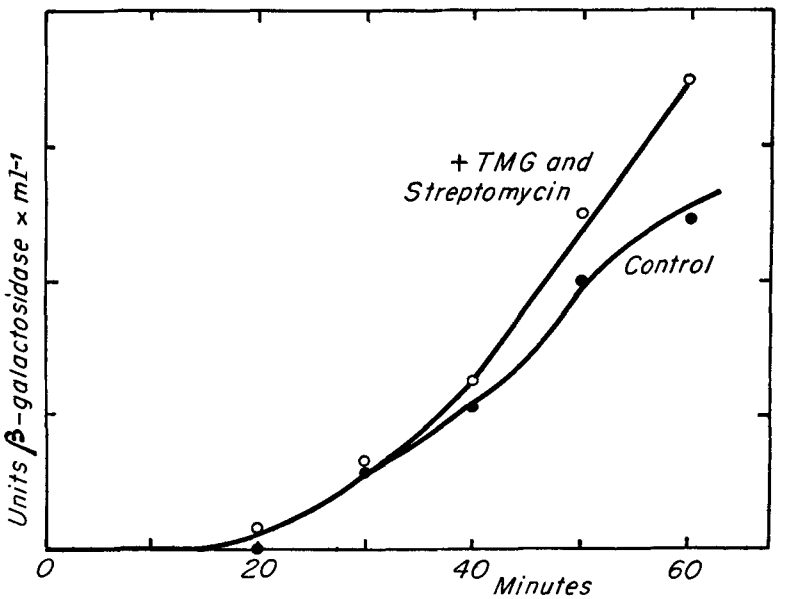


FIG. 3. Enzyme formation during first hour in mating B.
Mating under usual conditions. To an aliquot streptomycin (0·8 mg/ml.) was added at 20 minutes, and TMG at 25 minutes, to allow comparison of synthesis with and without inducer.

Such a mating therefore allows immediate and complete interaction of the z^+ from the ♂ with the i^- from the ♀. The possibility that the interaction depends upon actual recombination yielding z^+i^- in *cis* configuration is excluded because: (a) the synthesis begins virtually immediately after injection whereas genetic recombination is known (Jacob & Wollman, 1958) not to occur until 60 to 90 min after injection; (b) the factors z and i are so closely linked that recombination is an exceedingly rare

event (less than 10^{-4} of the zygotes) while the rate of enzyme synthesis is of an order indicating that most or all of the zygotes participate.

The possibility should also be considered that, rather than taking place through the cytoplasm, the interaction requires actual *pairing* of the homologous chromosome segments. This is excluded by the fact that the following mating:

$$\sigma\, z_2^- i_3^- \times \varphi\, z^+ i^+ \qquad \text{(C)}$$

when performed in the *absence* of inducer, yields no trace of enzyme, at any time after mixing, although conjugation and chromosome injection occur normally as shown by adequate controls involving other markers. The zygotes obtained in matings B and C are genetically identical, except that the wild type alleles (z^+i^+) are in relative excess (about 3 to 1) in (B), while the mutant alleles are in similar excess in (C). This quantitative difference cannot account for the absolute contrast of the results of the reciprocal matings, one allowing vigorous constitutive synthesis, the other none at all. This can only be attributed to the fact that the cytoplasm of the zygote is entirely furnished by the ♀ cell, with no significant contribution from the ♂. Therefore the $i^- \rightarrow z^+$ interaction must be considered to take place through the cytoplasm.

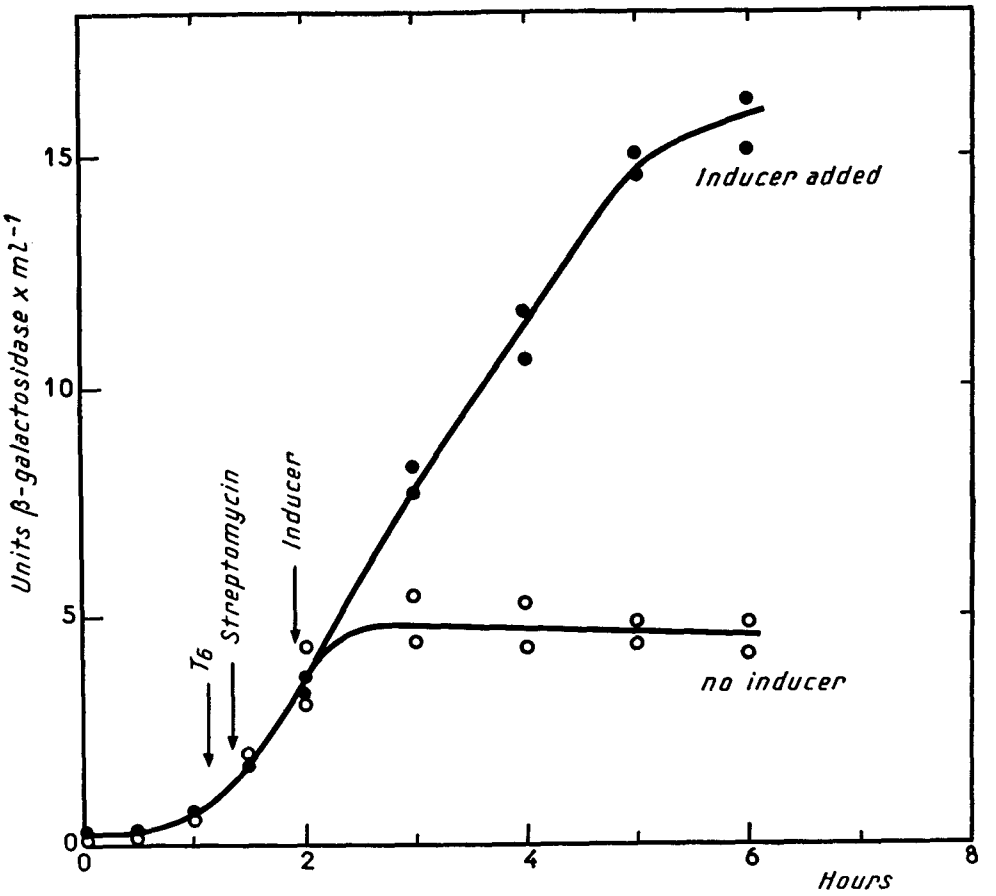


FIG. 4. Enzyme formation in mating D.

Mating performed under usual conditions in quadruplicate in absence of inducer. At times indicated, a suspension of phage T6 (20ϕ/B final concentration) and streptomycin (1 mg/ml.) were added to all of the cultures and TMG (2×10^{-3}M) was added to two of them (black circles) while the other two (white circles) received no addition.

This result may also be expressed by saying that the *i* factor sends out a cytoplasmic message which is picked up by the *z* gene, or gene products. Postulating, as we must, that this message is borne by a specific compound synthesized under the control of the *i* gene, we may further assume that one of the alleles of the *i* gene provokes the synthesis of the message, while the other one is inactive in this respect. If these assumptions are adequate, one of the alleles should be absolutely dominant over the other, but the dominance should become expressed only gradually when the cytoplasm of the zygotes came from the recessive parent, while it should be expressed immediately when the cytoplasm came from the dominant parent.

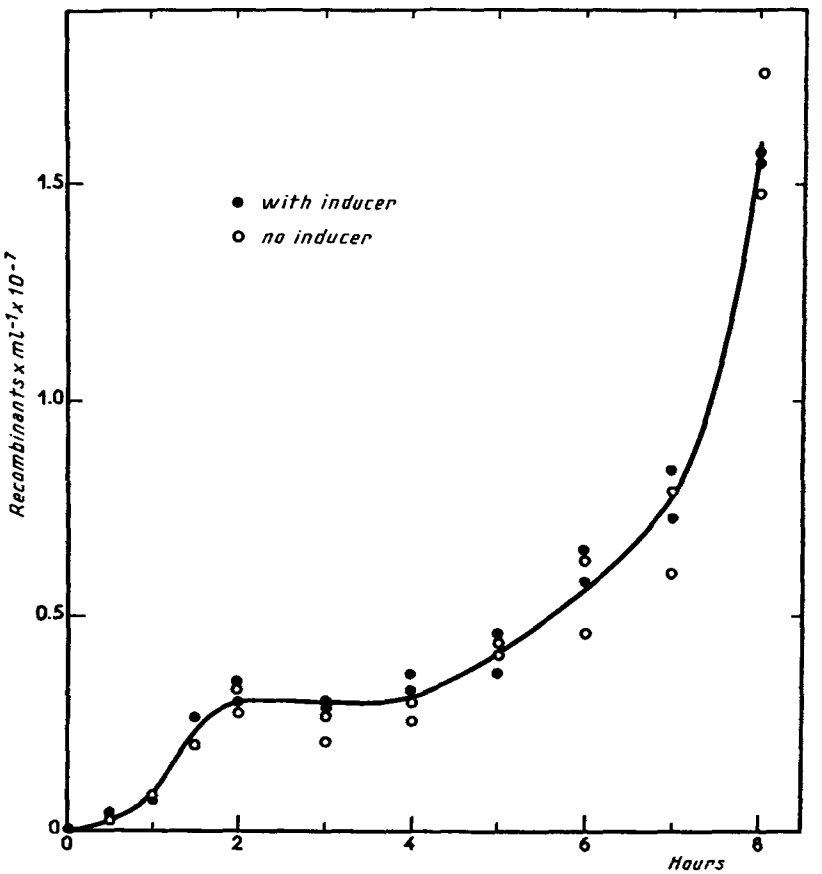


FIG. 5. Recombinant appearance in mating D.

Formation of z^+Sm^r recombinants tested by plating aliquots of the four cultures used in the experiment above (Fig. 4) on lactose-Sm agar. Portions of the culture were diluted 1000-fold and shaken vigorously at 100 minutes to prevent further mating. The increase up to the second hour is due to increasing numbers of zygotes. The increase after the fourth hour is due to *multiplication* of segregants (Wollman, Jacob & Hayes, 1956).

The fact that in matings of type (C) no enzyme is synthesized, even several hours after mating, means that the constitutive (i^-) allele from the ♂ is never expressed. This suggests that the dominant allele is the inducible (i^+). If so, the i^+ should eventually become expressed in matings of type (B)—i.e., the zygotes, initially constitutive (since their cytoplasm comes from the i^- parent), should eventually become inducible. To test this prediction, the following mating was performed:

$$\sigma^{\prime} z^{+}i^{+}Sm^{s}\ T6^{s} \times \varphi\ z_{2}^{-}i_{3}^{-}Sm^{r}\ T6^{r} \qquad \text{(D)}$$

and the synthesis of enzyme, in the absence and in the presence of inducer, was followed over several hours (in order to block induction of the ♂ and remating, a mixture of streptomycin and T6 phage was used). Figure 4 shows that, in the absence of inducer, enzyme synthesis stops about 90 min (or earlier) after entry of the $z^{+}i^{+}$ genes into the ♀ cells. When inducer is added at this stage, enzyme synthesis is resumed, showing that the initially constitutive $z^{+}i^{+}/z^{-}i^{-}$ zygotes have not been inactivated, but have become inducible.

It should be asked whether this conversion to inducibility, rather than occurring in the heterozygotes, might not correspond to the segregation of homozygous $z^{+}i^{+}Sm^{r}T6^{r}$ recombinants with concomitant disappearance of the heterozygotes. This is excluded because the earliest homozygous recombinants only appear 2 hr after the time when constitutive synthesis ceases† (Fig. 5).

From these observations we may conclude that the constitutive (i^{-}) allele is inactive, while the i^{+} is dominant, provoking the synthesis of a substance responsible specifically for the inducible behaviour of the galactosidase enzyme-forming-center.

5. Discussion and Conclusions

(1) The conclusions which can be directly drawn from the evidence presented above may be summarized as follows:

The synthesis of β-galactosidase and galactoside-permease in *E. coli* is controlled by three extremely closely linked genes (cistrons), *z*, *i* and *y*. The *z* gene determines, in part at least, the structure of the galactosidase protein molecule. The *y* gene probably does the same for the permease molecule, but there is no evidence on this point. The *i* gene in its active form controls the synthesis of a product which, when present in the cytoplasm, prevents the synthesis of β-galactosidase and galactoside-permease, unless inducer is added externally (inducible behaviour). When the *i* gene-product is absent or inactive as a result of mutation within the gene, no external inducer is required for β-galactosidase and galactoside-permease synthesis (constitutive behaviour). The *i* gene product is very highly specific, having no effect on any other known system.

(2) While proving that the interaction of the *i* and *z* factors involves a specific cytoplasmic messenger, the data presented here do not, by themselves, give any indication as to the mode of action of this compound. Two alternative models of this action should be considered.

According to one, which we shall call the "inducer" model, the activity of the galactosidase-forming system‡ requires the presence of an inducer, both in the constitutive and in the inducible organism. Such an inducer (a galactoside) is synthesized by *both* types of organisms. The i^{+} gene controls the synthesis of an enzyme which destroys or inactivates the inducer: hence the requirement for external inducer in the wild type. The i^{-} mutation inactivates the gene (or its product, the enzyme) allowing accumulation of endogenous inducer. This model accounts for the dominance of inducibility over constitutivity, and for the kinetics of conversion of the zygotes.

† It may also be recalled that, according to Anderson & Maze (1957), heterozygosis prevails for many generations in the descendants of *E. coli* zygotes.

‡ By this term we designate the system of all cellular constituents *specifically* involved in galactosidase synthesis. This includes the *z* gene and its cytoplasmic products.

According to the other, or "repressor", model the activity of the galactosidase-forming system is inhibited in the wild type by a specific "repressor" (probably also involving a galactosidic residue) synthesized under the control of the i^+ gene. The inducer is required only in the wild-type as an *antagonist* of the repressor. In the constitutive (i^-), the repressor is not formed, or is inactive, hence the requirement for an inducer disappears. This model accounts equally well for the dominance of i^+ and for the kinetic relationships.

(3) The "repressor" hypothesis might appear strictly *ad hoc* and arbitrary were it not also suggested by other facts which should be briefly recalled. That the synthesis of certain constitutive enzyme systems may be specifically inhibited by certain products (or even substrates) of their action, was first observed in 1953 by Monod & Cohen-Bazire working with constitutive galactosidase (of *E. coli*) (1953*a*) or with tryptophan-synthetase (of *A. aerogenes*) (1953*b*), and by Wijesundera & Woods (1953), and Cohn, Cohen & Monod (1953) independently working with the methionine-synthase complex of *E. coli*. It was suggested at that time that this remarkable inhibitory effect could be due to the displacement of an internally-synthesized inducer, responsible for constitutive synthesis, and it was pointed out that such a mechanism could account, in part at least, for the proper adjustment of cellular syntheses (Cohn & Monod, 1953; Monod, 1955). During the past two or three years, several new examples of this effect have been observed and studied in some detail by Vogel (1957), Yates & Pardee (1957), Gorini & Maas (1957). It now appears to be a general rule, for bacteria, that the formation of sequential enzyme systems involved in the synthesis of essential metabolites is *inhibited* by their end product. The convenient term "repression" was coined by Vogel to distinguish this effect from another, equally general, phenomenon: the control of enzyme *activity* by end products of metabolism.

(4) The facts which demonstrate the existence and wide occurrence of repression effects justify the basic assumptions of the repressor model. They do not allow a choice between the two models. Further considerations make the repressor model appear much more adequate:

(a) The repressor model is simpler since it does not require an independent inducer-synthesizing system.

(b) It predicts that constitutive mutants should, as a rule, synthesize more enzyme than induced wild-type. This appears to be the case for such different systems as galactosidase, amylomaltase (Cohen-Bazire & Jolit, 1953), glucuronidase (Stoeber, 1959, unpublished data), galactokinase of *E. coli* and penicillinase of *B. cereus* (Kogut, Pollock & Tridgell, 1956).

(c) The inducer model, if generalized, implies that internally synthesized inducers (Buttin, unpublished) operate in all constitutive systems. This assumption, first suggested as an interpretation of repression effects, has not been vindicated in recent work on repressible biosynthetic systems (Vogel, 1957; Gorini & Maas, 1957; Yates & Pardee, 1957). In contrast, the synthesis of numerous inducible systems has been known for many years (Dienert, 1900; Stephenson & Yudkin, 1936; Monod, 1942) to be inhibited by glucose and other carbohydrates. The recent work of Neidhardt & Magasanik (1957) has shown this glucose effect to be comparable to a non-specific repression and these authors have suggested that glucose acts as a preferential metabolic source of internally synthesized repressors. If this is so, and if our repressor model is correct, the conversion of glucose into specific galactosidase-repressor should be blocked in the constitutives. Accordingly the galactosidase-forming system of the

mutant should be largely insensitive to the glucose effect while other inducible systems should retain their sensitivity. That this is precisely the case (Cohn & Monod, 1953) is a very strong argument in favor of the repressor model.

(5) If adopted and confirmed with other systems, the repressor model may lead to a generalizable picture of the regulation of protein syntheses; according to this scheme, the basic mechanism common to all protein-synthesizing systems would be inhibition by specific repressors formed under the control of particular genes, and antagonized, in some cases, by inducers. Although the wide occurrence of repression effects is certain, the situation revealed with the present system, namely a genetic "complex" comprising, besides the "structural" genes (*z*, *y*) a repressor-making gene (*i*) whose function is to block or regulate the expression of the neighboring genes is, so far, unique for enzyme systems. But the formal analogy between this situation and that which is known to exist in the control of immunity and zygotic induction of temperate bacteriophage is so complete as to suggest that the basic mechanism might be essentially the same. It should be recalled that according to Jacob & Wollman (1956), when a chromosome from a λ-lysogenic ♂ of *E. coli* is injected into a non-lysogenic ♀, the process of vegetative phage development is started, which involves as an essential, probably as a primary, step the synthesis of specific proteins. When the reverse mating (♂ non-lysogenic × ♀ λ-lysogenic) is performed, zygotic induction does not occur; nor does vegetative phage develop when such zygotes are superinfected with λ particles. The λ-lysogenic cell is therefore immune against manifestations of prophage or phage potentialities, *and the immunity is expressed in the cytoplasm* (Jacob, 1958–59). Moreover the immunity is strictly specific, since it does not extend to other, even closely related, phages. The formation, under the control of a phage gene, of a specific repressor, able to block synthesis of proteins determined by other genes of the phage, would account for these findings.

(6) Implicit in the repressor model are two critical questions, which for lack of evidence we have avoided discussing, but which should be explicitly stated in conclusion. These questions are:

(a) What is the chemical nature of the repressor? Should it be considered a primary or a secondary product of the gene?

(b) Does the repressor act at the level of the gene itself, or at the level of the cytoplasmic gene-product (enzyme-forming system)?

We are much indebted to Professor Leo Szilard for illuminating discussions during this work and to Mme M. Beljanski, Mme M. Jolit and Mr. R. Barrand for assistance in certain experiments.

REFERENCES

Anderson, T. F. & Maze, R. (1957). *Ann. Inst. Pasteur*, **93**, 194.
Benzer, S. (1957). "The elementary units of heredity", in *The Chemical Basis of Heredity*, ed. by W. McElroy& B. Glass, p. 70. Baltimore: Johns Hopkins Press.
Cohen, G. N. & Monod, J. (1957). *Bact. Rev.* **21**, 169.
Cohen-Bazire, G. & Jolit, M. (1953). *Ann. Inst. Pasteur*, **84**, 937.
Cohn, M. (1957). *Bact. Rev.* **21**, 140.
Cohn, M., Cohen, G. N. & Monod, J. (1953). *C.R. Acad. Sci., Paris*, **236**, 746.
Cohn, M. & Monod, J. (1953). In *Adaptation in Microorganisms*, p. 132. Cambridge: University Press.
Dienert, F. (1900). *Ann. Inst. Pasteur*, **14**, 139.
Fisher, K. W. (1957). *J. Gen. Microbiol.* **16**, 120.

Gorini, L. & Maas, W. K. (1957). *Biochim. biophys. Acta*, **25**, 208.
Hayes, W. (1953). *Cold Spr. Harb. Symp. Quant. Biol.* **18**, 75.
Jacob, F. (1955). *Virology*, **1**, 207.
Jacob, F. (1958–59). Harvey Lectures, Series **54**, in the press.
Jacob, F. & Wollman, E (1956). *Ann. Inst. Pasteur*, **91**, 486.
Jacob, F. & Wollman, E. (1957). *C.R. Acad. Sci., Paris*, **244**, 1840.
Jacob, F. & Wollman, E. (1958). *Symp. Soc. Exp. Biol.* **12**, 75. Cambridge: University Press.
Kogut, M., Pollock, M. R. & Tridgell, E. J. (1956). *Biochem. J.* **62**, 391.
Lederberg, J. (1947). *Genetics*, **32**, 505.
Lederberg, J., Lederberg, E. M., Zinder, N. D. & Lively, E. R. (1951). *Cold Spr. Harb. Symp. Quant. Biol.* **16**, 413.
Monod, J. (1942). *Recherches sur la croissance des cultures bactériennes*. Paris: Herman Edit.
Monod, J. (1955). *Exp. Ann. Biochim. Méd., série* **17**, 195. Paris: Masson & Cie Edit.
Monod, J. & Cohen-Bazire, G. (1953*a*). *C.R. Acad. Sci., Paris*, **236**, 417.
Monod, J. & Cohen-Bazire, G. (1953*b*). *C.R. Acad. Sci., Paris*, **236**, 530.
Neidhardt, F. C. & Magasanik, B. (1957). *J. Bact.* **73**, 253.
Pardee, A. B., Jacob, F. & Monod, J. (1958). *C.R. Acad. Sci., Paris*, **246**, 3125.
Rickenberg, H. V., Cohen, G. N., Buttin, G. & Monod, J. (1956). *Ann. Inst. Pasteur*, **91**, 829.
Stephenson, M. & Yudkin, J. (1936). *Biochem. J.* **30**, 506.
Vogel, H. J. (1957). In *The Chemical Basis of Heredity*, ed. by W. D. McElroy & B. Glass, p. 276. Baltimore: Johns Hopkins Press.
Wijesundera, S. & Woods, D. D. (1953). *Biochem. J.* **55**, viii.
Wollman, E. & Jacob, F. (1955). *C.R. Acad. Sci., Paris*, **240**, 2449.
Wollman, E., Jacob, F. & Hayes, W. (1956). *Cold Spr. Harb. Symp. Quant. Biol.* **21**, 141.
Yates, R. A. & Pardee, A. B. (1957). *J. Biol. Chem.* **227**, 677.

J. Mol. Biol. (1969) **43**, 281–298

An Analysis of "Revertants" of a Deletion Mutant in the *C* Gene of the L-Arabinose Gene Complex in *Escherichia coli* B/r: Isolation of Initiator Constitutive Mutants (I^c)

ELLIS ENGLESBERG, DAVID SHEPPARD†, CRAIG SQUIRES AND FRANK MERONK, JR.

Department of Biological Sciences
University of California, Santa Barbara, Calif., U.S.A.

(*Received 15 November 1968, and in revised form 15 February 1969*)

Nineteen independent L-arabinose-utilizing "revertants" were isolated from *Escherichia coli B/r* containing a deletion (*Δ*719) that encompasses all known mutations in the regulator gene *araC*. The revertants contain the original deletion plus a secondary initiator constitutive mutation (I^c). All produce low constitutive levels of enzymes in the L-arabinose pathway. The I^c mutant sites are all closely linked to the deletion. Nine of the I^c sites, mapped with greater resolution, are located to the left of deletion 719, in the region containing the proposed initiator controlling site (*araI*) for this operon. The I^c alleles in all revertants tested are *cis*-dominant to the wild-type allele, I^+, and have no *trans* effect. All the strains are hyperinducible to varying degrees in the presence of a functioning C^+ allele in the *trans* position.

In merodiploids, the C^+ (even in the absence of inducer) and not the C^- alleles are able to stimulate the expression of the *araA* gene *cis* but not *trans* to the I^+ *Δ*719 and to most I^c*Δ*719 mutations. These results, together with other evidence, support a modified positive control model in which P1, the initial product of the *araC* gene, is a true repressor existing in equilibrium with P2, the activator, and with P1 and P2 attached to their respective controlling sites, the operator, *araO*, and the initiator, *araI*, located as follows: *araB*, *araI*, *araO*, *araC*. Evidence indicates that the repressor–operator site function is epistatic over the activator–initiator site function. The *cis* effect of deletion 719 in the presence of a *trans* acting C^+ allele is explained on the basis that this deletion desensitizes this operon to the repressor, P1, by excising the operator, and thus allows this operon to be activated by P2 whose presence would otherwise have remained cryptic.

Among 19 constitutive revertants of deletion 719, none could be identified as mutants in a regulatory gene of the negative control type. Therefore, no evidence could be found to support a model of negative control internal induction in this system.

1. Introduction

Evidence, previously presented (Englesberg, Irr, Power & Lee, 1965; Sheppard & Englesberg, 1966,1967), clearly demonstrates that the gene *araC* in the L-arabinose system is distinct from a regulator gene of a negative control system; *e.g.* the *i* gene

† Present address: Department of Biological Sciences, University of Delaware, Newark, Del., U.S.A.

of the β-galactosidase system. Whereas a product of the *araC* gene, the activator, is required for the expression of the structural genes in the L-arabinose system, the only known functional product of the *i* gene, the repressor, prevents the otherwise free expression of the related structural genes (Jacob & Monod, 1961). Thus deletions (Sheppard & Englesberg, 1966,1967) and nonsense mutations (Irr & Englesberg, 1967) of the *araC* gene lead to a pleiotropic negative phenotype, C^-, (non-inducible for the enzymes specifically involved in L-arabinose metabolism), while deletions or nonsense mutations of the *i* gene in the β-galactosidase system (Willson, Perrin, Cohn, Jacob & Monod, 1964; Bourgeois, Cohn & Orgel, 1965; Muller-Hill, 1966) lead to a pleiotropic constitutive phenotype. While the $araC^-$ alleles are recessive (*cis* and *trans*) to the alleles of the *araC* gene for inducibility (C^+) and constitutivity (C^c), lactose non-utilizing pleiotropic negative mutants of the *i* gene (i^s), on the other hand, are *cis* and *trans* dominant to the analogous alleles of that gene (Jacob & Monod, 1961; Willson *et al.*, 1964).

The above characteristics of the *araC* gene are consistent with a model of positive control (Englesberg *et al.*, 1965; Sheppard & Englesberg, 1966,1967). According to this model, the gene *araC* (C^+ allele, wild type) produces a product (Pl), an allosteric protein, which is in equilibrium with P2, the activator. In the absence of L-arabinose, the equilibrium is in the direction of P1. L-Arabinose shifts the equilibrium to P2. P2, by reacting with a controlling site, the initiator (*araI*), located in the region between genes *araB* and *araC*, stimulates the expression of the structural genes *araB*, *araA* and *araD*. The initiator is thus the site of action of activator and the site of initiation of gene expression. (It has not been specified as to whether activator functions at the transcriptional or translational level.) According to the model, ara^- mutants in the *C* gene, C^-, fail to produce a biologically active C product. C^c mutants produce substantial amounts of an activator (P3, P4 . . . P*n*) in the absence of L-arabinose, as a result of a primary alteration in the amino acid sequence of the C product. Thus C^- mutants, producing no biologically active C product, are *cis* and *trans* recessive to C^c and C^+. Through interaction with the *araI* site, activator produced by the C^c allele in the absence of L-arabinose and activator produced by the C^+ allele in the presence of L-arabinose "turns on" the L-arabinose structural genes *cis* to *araI*.

The existence of P1 and a complication to the model of pure positive control was indicated by the finding that, in the absence of inducer, C^+ is dominant to C^c. Apparently P1, produced by the *C* gene in the absence of the inducer, antagonizes the action of the activator produced by the C^c allele. There have been three hypotheses proposed to explain the dominance of C^+ to C^c. (1) P1 is a repressor and competes with P2 for attachment at the initiator site (Sheppard & Englesberg, 1966). (2) P1 is a repressor which attaches at a separate site, the operator, and circumvents the action of P2 (Sheppard & Englesberg, 1966). (3) Interaction between the possible subunits of P1 and P3 results in the production of an inactive molecule (Sheppard & Englesberg, 1967). In another paper (Englesberg, Squires & Meronk, 1969), evidence is presented showing that P1 acts as a true repressor with a separate site of attachment, the operator (*araO*), located between *araI* and the *araC* gene (see Fig. 1). This has recently been confirmed by Kessler & Englesberg (1969).

Several lines of evidence indicate that the controlling sites for this system, including *araI*, are located in the region between genes *araB* and *araC* (see Fig. 2). (1) Gene *araC* is not part of the *B,A,D* operon (Englesberg *et al.*, 1965; Sheppard & Englesberg,

1967). (2) Polarity is in the direction *B,A,D* (unpublished data of Katz & Englesberg, 1968 and Hogg & Englesberg, 1969). (3) Deletions that excise the region between *araB* and *araC* lead to an absolute pleiotropic negative, *cis* dominant, phenotype (Sheppard & Englesberg, 1967). (4) Deletions that end within the *B* gene and the leucine operon and thus excise the region between *araB* and *araC* remove the remaining structural genes in the L-arabinose *B,A,D* operon from the control by L-arabinose and gene *araC* (as demonstrated in heterogenotes) and place them under the control of the leucine regulator gene; whereas deletions that end within the *C* gene and the leucine operon do not affect the L-arabinose-gene *araC* control of the L-arabinose *B,A,D* operon (Kessler & Englesberg, 1969).

A model that has been most frequently proposed as an alternative to positive control is one based upon negative control modified by internal induction. According to this model, the arabinose operon is proposed to be actually under the negative control of a yet undiscovered repressor-forming regulatory gene (gene *R*) comparable to the *i* gene in the β-galactosidase system. Gene *C*, according to this model, is the structural gene for an enzyme that converts L-arabinose into the real inducer. This inducer reacts with the repressor produced by gene *R* and inactivates it. (Although there are a number of related negative control models different from one another by the function assigned to gene *araC*, they are all predicated upon the existence of a typical repressor-forming regulatory gene *R*.) Based upon this model, C^c mutants would provide an altered enzyme converting some internal metabolite into the inducer. Some mutants of this proposed regulatory gene that we might have been expected to find are R^- constitutives. It is argued that we have not searched hard enough for such mutants.

In this paper, we describe the isolation and characterization of 19 Ara^+ revertants of an Ara^- mutant strain, SB1094, containing a deletion (*Δ*719) that covers all known point mutations in the *araC* gene and produces, as a result, a pleiotropic negative phenotype. The 19 revertants contain the original deletion and a closely-linked secondary mutation mapping within the *araI* region of the L-arabinose operon and producing a *cis*-dominant constitutive phenotype characteristic of initiator constitutive mutants (I^c). The fact that none of the revertants maps in a hitherto undescribed repressor gene and that none has the characteristics of R^- constitutives suggest that no such regulatory gene exists. Therefore, we find no support for the negative control internal induction model. On the other hand, this type of reversion pattern (I^c's but no R^-'s) is what we would expect on the basis of the positive control model.

2. Materials and Methods

The media, strains (see Table 1), and general procedures for matings, the preparation and verifications of the genotype of merodiploids have been previously described (Sheppard & Englesberg, 1967).

The following abbreviations are used in media designation: M, mineral base; Ara, L-arabinose; Thr, L-threonine; Leu, L-leucine; Met, L-methionine; Glu, glucose; CH, casein hydrolysate; Str, streptomycin.

Transducing lysates of phage P1bt were prepared by the method of Gross & Englesberg (1959) as modified by Boyer, Englesberg & Weinberg (1962).

Transduction experiments were carried out as previously described (Gross & Englesberg, 1959).

Deletion 719 was originally isolated in Hfr 33 *araD139 his*$^-$ as a result of a spontaneous mutation producing resistance to the L-arabinose inhibition (strain SB1122) (Sheppard &

TABLE 1

List of strains

Strain	Mating type	Genotype					Origin, source or reference
		L-Arabinose	*thr1*	*leuB1*	*his*	*str*	
UP1001	F⁻	*ara*⁺	+	+	+	s	Gross & Englesberg (1959)
UP1004	F⁻	*ara*⁺	−	−	+	s	Gross & Englesberg (1959)
UP1005	F⁻	*ara*⁺	−	−	+	r	From UP1004 spontaneous mutations
UP1002	F⁻	*ara*⁺	+	−	+	s	From UP1004 by transduction
UP1089	F⁻	*A2*	−	−	+	s	Gross & Englesberg (1959)
UP1080	F⁻	*A2*	+	−	+	s	From UP1089 by transduction
UP1027	F⁻	*B24*	+	+	+	s	Gross & Englesberg (1959)
UP1010	F⁻	*C3*	+	+	+	s	Gross & Englesberg (1959)
UP1092	F⁻	*C5*	−	−	+	s	Gross & Englesberg (1959)
UP1082	F⁻	*C5*	+	−	+	s	From UP1092 by transduction
SB1122	Hfr33	*D139Δ719*	+	+	−	s	Sheppard & Englesberg (1967)
SB1094	F⁻	*Δ719*	−	+	+	r	SB1122 × UP1005, this paper
SB2000 to SB2018	F⁻	*I°1 Δ719* to *I°19 Δ719*	−	+	+	r	From SB1094, this paper
SB3139	F′	F′ *C19/C19*	+/+	+/+	+	s	Sheppard & Englesberg (1967)
SB3116	F′	F′ *C12/C12*	+/+	+/+	+	s	Sheppard & Englesberg (1967)
SB3141	F′	F′ *C101/C101*	+/+	+/+	+	s	Sheppard & Englesberg (1967)
SB3107	F′	F′ *B24/B24*	+/+	+/+	+	s	Sheppard & Englesberg (1967)
SB3101	F′	F′ *A2/A2*	+/+	+/+	+	s	Sheppard & Englesberg (1967)
SB1509	F⁻	*Δ1109*	+	+	+	r	Kessler & Englesberg (1969)
SB3550	F′	F′ *A2/Δ719*	+/−	+/+	+	r	SB3101 × SB1094, this paper
SB3538	F′	F′ *A2/C3*	+/+	+/+	+	s	Cross SB3101 × UP1010, this paper
SB3147	F′	F′ *A2 C3/A2 C3*	+/+	+/+	+	s	Negative segregants of SB3538
SB3551 to SB3569	F′	F′ *A2/I°1 Δ719* to F′ *A2/I°19 Δ719*	+/−	+/+	+	r	SB3101 × SB2000 to SB2018, this paper
SB3571 to SB3589	F′	F′ *A2 C3/I°1 Δ719* to F′ *A2 C3/I°19 Δ719*	+/−	+/+	+	r	SB3147 × SB2000 to SB2018, this paper
SB2149	F⁻	*A2 I°1 Δ719*	+	+	+	s	Phage P1bt (SB2000) × UP1080, this paper

TABLE 1 [*continued*]

List of strains

Strain	Mating type	Genotype					Origin, source or reference
		L-Arabinose	*thr1*	*leuB1*	*his*	*str*	
SB2151	F⁻	*A2 I^c13 Δ719*	+	+	+	s	Phage P1bt (SB2012) × UP1080, this paper
SB2152	F⁻	*A2 I^c14 Δ719*	+	+	+	s	Phage P1bt (SB2013) × UP1080,
SB2153	F⁻	*A2 I^c17 Δ719*	+	+	+	s	Phage P1bt (SB2016) × UP1080, this paper
SB2154	F⁻	*A2 I^c19 Δ719*	+	+	+	s	Phage P1bt (SB2018) × UP1080, this paper
SB5312	F⁻	*A2 I⁺ Δ719*	+	+	+	s	Phage P1bt (SB1094) × UP1080, this paper
SB3617	F′	F′ *B24/A2 I^c1 Δ719*	+/+	+/+	+	s	SB3107 × SB2149, this paper
SB3619	F′	F′ *B24/A2 I^c13 Δ719*	+/+	+/+	+	s	SB3107 × SB2151, this paper
SB3620	F′	F′ *B24/A2 I^c14 Δ719*	+/+	+/+	+	s	SB3107 × SB2152, this paper
SB3621	F′	F′ *B24/A2 I^c17 Δ719*	+/+	+/+	+	s	SB3107 × SB2153, this paper
SB3622	F′	F′ *B24/A2 I^c19 Δ719*	+/+	+/+	+	s	SB3107 × SB2154, this paper
SB3636	F′	F′ *B24/A2 I⁺ Δ719*	+/+	+/+	+	s	SB3107 × SB5312, this paper

Abbreviations used: ara, L-arabinose; *A*, structural gene for L-arabinose isomerase; *B*, structural gene for L-ribulokinase; *C*, regulator gene in the L-arabinose system; *D*, structural gene for L-ribulose 5-phosphate 4-epimerase; *Δ*, deletion; his, histidine; leu, leucine; thr, threonine; str, streptomycin; r, resistant; s, sensitive; +, ability to synthesize or utilize; −, inability to synthesize or utilize.

Englesberg, 1967). This deletion has been shown to encompass the region of the *C* gene defined by the outermost C^- mutants available at that time; *C12* (left end) and *C19* (right end) (Sheppard & Englesberg, 1967). Another C^- mutant site (*C101*) has subsequently been mapped to the left of *C12*. Deletion 719 fails to recombine with *C101*. Thus deletion 719 encompasses, at least, the entire *C* gene as defined by point mutations *C101* and *C19* (see Fig. 2). To isolate deletion 719 free of the *D139* marker, strain SB1122 was crossed to strain F^- *thr1 leuB1* str^r1 and selection was carried out on M–Glu–Thr–Str–agar plates. Among the $Leu^+His^+Str^r$ recombinants, a *thr1 araD*$^+$ *Δ*719 str^r recombinant (strain SB1094) was isolated and characterized by tests on appropriate media and by progeny tests with F′*ara*$^-$/*ara*$^-$ homogenotes. To ensure the isolation of independent revertants of SB1094, 20 tubes containing L-broth were inoculated with approximately 200 cells of this strain. The cultures were incubated at 37°C with shaking overnight. 3 ml. of each independent culture was then incubated with 0·3 ml. of diethyl sulfate at 37°C for 30 min. 0·1 ml. of the treated culture was then diluted into 5·0 ml. of fresh L-broth, grown overnight, and samples were plated on M–Ara–Thr–Met–agar plates and incubated at 37°C. (Methionine was added as a precaution, in certain of these experiments, since it was found that some strains carrying the *thr1* marker develop a requirement for methionine after prolonged storage as slant cultures. Subsequent experiments, however, indicated that such a precaution was unnecessary with the cultures used in these experiments.) No revertants were detected until approximately 6 days of incubation. 19 revertants, one from each independent culture, were isolated in pure culture by restreaking twice on homologous media and the cultures were subsequently stored on nutrient agar slants and also lyophilized. No revertants were detected on plates inoculated with untreated cultures (10^{10} total bacteria).

For the preparation of cell-free extracts, all cultures were grown at 37°C with shaking in a medium containing mineral salts (Sheppard & Englesberg, 1967), 1% casein hydrolysate (Difco) and 0·4% L-arabinose when required. In early experiments, as will be indicated, cells were harvested in the exponential phase of growth. Subsequently, stationary phase cultures were employed. The latter invariably yielded cell-free extracts with higher and more reproducible L-arabinose isomerase activity than obtainable with exponential phase cultures. Extracts were prepared as previously described with slight modifications (Cribbs & Englesberg, 1964).

L-Ribulokinase assays were performed as previously described (Englesberg *et al.*, 1965).

L-Arabinose isomerase activity was performed as previously described (Englesberg *et al.*, 1965) except that the enzyme assay was determined at 30°C instead of 37°C, since the reaction was found to be linear for longer periods of time at the lower temperature.

Protein was estimated by the method of Lowry, Rosebrough, Farr & Randall (1951), using crystalline bovine serum albumin (California Corp. for Biochemical Research) as a standard.

3. Results

General properties of the revertants

Nineteen independent diethyl sulfate-induced, arabinose-utilizing revertants of strain SB1094 carrying deletion 719 were isolated. All grew poorly on M–Thr–Ara or M–Thr–Met–Ara media, producing colonies approximately 1-mm in diameter in 48 hours at 37°C as compared to colonies 3-mm in diameter for the wild type. All had low constitutive levels of L-arabinose isomerase and L-ribulokinase, ranging from 1·5 to 8·7 and 0·4 to 1·8%, respectively, of the wild-type induced levels with exponentially grown cultures (Table 2). (Enzyme levels were higher in stationary phase cultures, Table 6.) There appears to be a general lack of co-ordination between isomerase and kinase levels as compared to that of the wild type. (For an explanation, see Discussion.) None of the revertants showed any significant increase in isomerase or kinase levels when subjected to conditions of induction with L-arabinose as the inducer. This low constitutive expression of the L-arabinose gene cluster though

TABLE 2

Enzymic characterization of I^c*Δ719 revertants*

Strain	Arabinose genotype	L-Arabinose isomerase		L-Ribulokinase	
		Non-induced	Induced	Non-induced	Induced
SB2000	*I*^{c}1 *Δ*719	1·8	1·6	0·17	0·14
SB2001	*I*^{c}2 *Δ*719	1·3	1·7	0·11	0·15
SB2002	*I*^{c}3 *Δ*719	1·1	2·1	0·06	0·14
SB2003	*I*^{c}4 *Δ*719	1·0	1·0	0·11	—
SB2004	*I*^{c}5 *Δ*719	2·1	3·5	0·24	0·23
SB2005	*I*^{c}6 *Δ*719	1·1	1·2	0·08	0·09
SB2006	*I*^{c}7 *Δ*719	2·9	4·0	0·18	0·24
SB2007	*I*^{c}8 *Δ*719	0·7	0·9	0·10	—
SB2008	*I*^{c}9 *Δ*719	1·7	2·4	0·12	0·18
SB2009	*I*^{c}10 *Δ*719	2·3	2·5	0·12	0·20
SB2010	*I*^{c}11 *Δ*719	1·0	1·1	0·07	0·08
SB2011	*I*^{c}12 *Δ*719	2·3	2·0	0·11	0·09
SB2012	*I*^{c}13 *Δ*719	0·5	0·6	—	—
SB2013	*I*^{c}14 *Δ*719	1·3	1·6	—	—
SB2014	*I*^{c}15 *Δ*719	0·8	0·4	0·10	—
SB2015	*I*^{c}16 *Δ*719	0·9	1·2	0·06	0·06
SB2016	*I*^{c}17 *Δ*719	1·6	1·1	0·07	0·09
SB2017	*I*^{c}18 *Δ*719	1·4	1·6	0·11	0·10
SB2018	*I*^{c}19 *Δ*719	0·8	0·9	0·14	—
SB1094	*I*$^+$ *Δ*719	0·10	0·17	0·008	<0·01
UP1001	*I*$^+$*C*$^+$ (wild type)	0·08	33·4	0·03	13·4

Cell-free extracts were prepared from cells in the exponential phase of growth. Enzyme activity is expressed in μmoles of product formed/hr/mg protein.

refractory to induction is apparently sufficient to permit the slow growth observed with these mutants.

To avoid the confusion of a duplicate nomenclature, we shall refer to the revertant mutant sites now as *I*^{c} (initiator constitutive) and will justify this later. For convenience, the combined phenotype of *I*^{c}*Δ*719, characterized by slow growth with L-arabinose as the carbon source, will be indicated by the symbol Ara$^{+\,s1}$, in contrast to the wild-type L-arabinose phenotype, designated as Ara$^+$ and the L-arabinose non-utilizing phenotype as Ara$^-$.

(i) *Presence of the deletion*

When each of the revertants (strains SB2000 to SB2118) were crossed with F′*araC19*/*araC19* and F′*araC101*/*araC101* homogenotes, each carrying an *ara*$^-$ mutant site that mapped at either end of the *C* gene, no wild-type Ara$^+$ recombinants were detected. However, when F′*araA*$^-$/*araA*$^-$ and F′*araB*$^-$/*araB*$^-$ homogenotes were used as donors, wild-type Ara$^+$ recombinants were obtained in each case. Thus it is evident that these revertant strains still contained a deletion encompassing known point mutations in the *C* gene characteristic of the deletion in the parent strain SB1094.

(ii) *Mapping of the revertant mutant sites*

Cross I. P1bt transducing phage, prepared with each of the revertant strains SB2000 to SB2018 and the parental deletion strain SB1094, were used as donors in

TABLE 3

Cross I I^c*Δ719* (*donor*) × leuB1 (*recipient*)

Donor Strain	Donor Arabinose genotype	Selected Leu$^+$ analyzed	Unselected Ara^{+s1} among Leu$^+$†	Leu$^+$Ara^{+s1} Leu$^+$ (%)	Unselected Ara$^-$ among Leu$^+$‡	Leu$^+$Ara$^-$ Leu$^+$ (%)
SB2000	*I*^{c}1 *Δ719*	150	79	52	1	0·67
SB2001	*I*^{c}2 *Δ719*	333	171	52	3	0·90
SB2002	*I*^{c}3 *Δ719*	444	248	56	1	0·23
SB2003	*I*^{c}4 *Δ719*	270	144	53	1	0·37
SB2004	*I*^{c}5 *Δ719*	590	319	54	4	0·68
SB2005	*I*^{c}6 *Δ719*	299	154	52	1	0·34
SB2006	*I*^{c}7 *Δ719*	298	168	56	1	0·34
SB2007	*I*^{c}8 *Δ719*	323	175	54	2	0·62
SB2008	*I*^{c}9 *Δ719*	303	172	57	2	0·66
SB2009	*I*^{c}10 *Δ719*	168	95	57	1	0·60
SB2010	*I*^{c}11 *Δ719*	298	161	54	2	0·67
SB2011	*I*^{c}12 *Δ719*	407	227	56	1	0·24
SB2012	*I*^{c}13 *Δ719*	300	169	56	1	0·33
SB2013	*I*^{c}14 *Δ719*	555	311	56	5	9·90
SB2014	*I*^{c}15 *Δ719*	70	37	45	1	0·14
SB2015	*I*^{c}16 *Δ719*	300	189	63	1	0·33
SB2016	*I*^{c}17 *Δ719*	740	416	56	5	0·68
SB2017	*I*^{c}18 *Δ719*	586	311	53	5	0·85
SB2018	*I*^{c}19 *Δ719*	554	313	56	1	0·18
SB1094	*Δ719*	120	0	—	76	63

Phage P1bt was grown for two cycles on each of the revertants of strain SB1094, *Δ719*. The complete genotype of the revertants, strains SB2000 to SB2018, is *thr*$^-$*I*^{c}1–19*Δ719* *str*^{r}. The phage were used in transduction experiments with UP1002 *leuB1* as the recipient. Selection was for Leu$^+$ on M–Glu–agar. Leu$^+$ transductants were then picked to M–Glu–agar (20/plate). The plates were incubated overnight at 37°C and replica plated on to M–Ara–agar and as a control on to M–Glu–agar. In some crosses threonine and threonine and methionine were included in the medium. This had little effect on the results. (*thrI* cotransduces only about 2% of the time with *leuB1* (Gross & Englesberg, 1959).) As controls, phage were grown on the original deletion mutant, strain SB1094, and crossed to *leuB1*. Leu$^+$ transductants were selected and screened in the same manner as described above. In addition *leuB1* was plated on M–Glu–agar without phage to detect the presence of spontaneous Leu$^+$ revertants. No such revertants were observed.

† The phenotype of *I*^{c}*Δ719* mutants is Ara^{+s1} (slow growth on agar medium with L-arabinose as carbon source; in 48 hr at 37°C, colony size is approximately 1 mm in diameter as compared to 3mm for the wild type). In replica plating to M–Ara from a patch of cells on M–Glu it is a simple matter to distinguish between Ara$^+$ (wild type), Ara^{+s1} (slow grower) and Ara$^-$ (no growth). In 48 hr Ara$^+$ produces a heavy patch; Ara^{+s1}, a light but definite patch, and Ara$^-$, no patch (no growth) at all. The unselected Ara^{+s1} clones scored therefore represent the *I*^{c}*CΔ719* genotypes in this experiment. Several of these transductants from each cross were analyzed by progeny tests by crossing them to F′ *ara*$^-$ homogenotes. In each case the presence of *Δ719* was confirmed.

‡ The phenotype of *Δ719* is Ara$^-$ (no growth on agar medium with L-arabinose as carbon source). All Ara$^-$ transductants were analyzed by progeny tests against F′ Ara$^-$ homogenotes. In every case they were shown to contain the original deletion *Δ719*.

crosses with *leuB1* (strain UP1002) as the recipient (Table 3). We selected for Leu$^+$ and analyzed these recombinants to determine the L-arabinose phenotype. Among the Leu$^+$ recombinants, three different L-arabinose phenotypes were obtained; Ara$^+$, Ara^{+s1} and Ara$^-$. Several Ara^{+s1} colonies from each cross and all L-arabinose negative clones were purified and shown by progeny tests, as previously described, to contain the deletion *Δ719*. The frequency of co-transduction of *I*^{c}*Δ719* with the leucine marker, *i.e.* Ara^{+s1}Leu$^+$, is approximately 50% (the differences observed

are probably not significant) and is similar to that obtained in the control cross, *Δ719* × *leuB1*, for co-transduction of the parental deletion 719 with the leucine marker (Ara⁻Leu⁺). The low frequency of Leu⁺Ara⁻ recombinants in the experimental crosses is probably due to a rare crossover event resulting in the transfer of the original deletion 719 to the Ara⁺Leu⁻ recipient cell. (Although I^cC^+ recombinants probably occur in this population, they are probably not distinguishable from wild-type recombinants on arabinose agar plates. This conclusion is based upon the findings (see below) that merodiploids of the type $F'A2C^+/A^+I^cΔ719$ have inducible levels of L-arabinose isomerase close to fully induced wild-type cells. In addition, a strain in which I^c marker has been isolated in an otherwise wild-type L-arabinose genetic background ($A^+B^+I^c13C^+$) is indistinguishable from wild-type colonies on mineral L-arabinose agar plates.) It is clear from this evidence that the I^c mutations in strain SB1094, carrying the deletion 719 and producing the Ara^{+s1} phenotype, are separable mutational events from the deletion mutation itself. Furthermore, since the frequency of Ara⁻ among the Leu⁺ transductants is very low, the I^c mutations are closely linked to the deletion 719 or to the *leuB1* locus.

Cross II. In a second series of crosses the same donor phages prepared on each of

TABLE 4

Cross II $I^cΔ719$ (*donor*) × C5 leuB1 (*recipient*)

Donor		Total Ara^{+s1} analyzed	$Ara^{+s1}Leu^+$	$\frac{Ara^{+s1}Leu^+}{Ara^{+s1}} \times 100$
Strain	Arabinose genotype			
SB2000	I^c1 Δ719	280	172	61
SB2001	I^c2 Δ719	280	165	59
SB2002	I^c3 Δ719	350	233	67
SB2003	I^c4 Δ719	279	193	69
SB2004	I^c5 Δ719	140	86	61
SB2005	I^c6 Δ719	140	97	69
SB2006	I^c7 Δ719	349	236	68
SB2007	I^c8 Δ719	350	233	67
SB2008	I^c9 Δ719	140	102	73
SB2009	I^c10 Δ719	350	185	53
SB2010	I^c11 Δ719	350	254	73
SB2011	I^c12 Δ719	131	83	63
SB2012	I^c13 Δ719	350	246	70
SB2013	I^c14 Δ719	140	68	49
SB2014	I^c15 Δ719	350	239	68
SB2015	I^c16 Δ719	140	73	52
SB2016	I^c17 Δ719	140	78	56
SB2017	I^c18 Δ719	140	85	61
SB2018	I^c19 Δ719	140	75	54

Phage P1bt as used in Cross I was employed in transduction experiments with *araC5 leuB1* as the recipient. 0·1 ml. of a 1/10 dilution of each of the transducing mixtures and the culture of *araC5 leuB1* without phage were plated in duplicate on M–Thr–Ara–agar plates to select for arabinose utilizing colonies. After 6 days incubation at 37°C Ara^{+s1} transductants were then picked onto homologous media (20/plate) and subsequently replica plated on to M–Glu–Thr–agar and M–Glu–Leu–Thr–agar as a plating control. The number (1 to 2/plate) of Ara^+ (large colonies) appearing in 48 hr were approximately the same in all cases including that of the control cross (donor *Δ719 leuB1* × *araC5 leuB1* recipient) and on plates seeded with the *C5 leuB1* recipient strain alone and probably are revertants of *araC5*. A few small colonies, on the average 5/ml., appeared in 6 days on these control plates.

the revertant strains were crossed to *araC5 leuB1* as the recipient (Table 4). In this case selection was for arabinose-utilizing recombinants on M–Ara–Thr–Leu agar plates. The number of Ara$^+$ (wild type, large colony types) found (on the average 1/plate) were the same as on the control plates containing the cross, deletion 719 × *araC5 leuB1*, and on plates containing just the recipient and are therefore probably the result of spontaneous reversion of the *araC5* mutation. The remainder, slow growing revertant-type recombinants (Ara$^{+\,sl}$), were observed after six days of incubation in each of these crosses, at an average frequency of approximately 2×10^4/ml. plated as compared to 5/ml. with the control described above. (*I*^{c} *araC5* recombinants probably yield an Ara$^{+\,sl}$ phenotype.) These recombinants were picked to homologous medium and replica-plated onto M–Glu–Thr–Agar to determine their leucine phenotype. The percentage of the Ara$^{+\,sl}$ that were Leu$^+$ in each of the crosses varied from 49 to 73. (Because of the relatively small numbers assayed, these differences are probably not significant.) These frequencies of co-transduction of *I*^{c}*Δ*719 with *leuB1* are in close agreement with the previous observations of Gross & Englesberg (1959).

The results of the crosses, with *araC5 leuB1* as recipient, demonstrate that the revertant mutant sites must be closely linked to the deletion itself and thus to the *C* gene since there is a high frequency of segregation of the Leu character among the *I*^{c}*Δ*719 transductants.

Cross III. In a third series of crosses the same donor phages from ten of the revertants were crossed with the Ara$^-$ strain SB1509 containing the large deletion 1109 (Table 5). This deletion encompasses a genetic region extending from *araC101*, the mutant site closest to the *B* gene, to and including genes in the leucine operon (Kessler & Englesberg, 1969), and is phenotypically Ara$^-$Leu$^-$. Arabinose-utilizing recombinants were selected on M–Ara–Leu–agar medium. Only the slow growing type of arabinose-utilizing clones (Ara$^{+\,sl}$) were observed in these crosses. These

TABLE 5

Cross III I^c*Δ719 (donor)* × *Δ1109*

Strain	Donor Arabinose genotype	Selected Ara$^{+\,sl}$ analyzed	Unselected Leu$^-$ among Ara$^{+\,sl}$	$\frac{\text{Ara}^{+\,sl}\text{Leu}^-}{\text{Ara}^{+\,sl}}$ (%)	Spontaneous reversion frequency†
SB2000	*I*^{c}1 *Δ*719	1000	1	0·1	<0·02
SB2001	*I*^{c}2 *Δ*719	995	0	<0·10	<0·02
SB2002	*I*^{c}3 *Δ*719	477	2	0·42	<0·1
SB2005	*I*^{c}6 *Δ*719	914	3	0·33	<0·02
SB2008	*I*^{c}9 *Δ*719	1000	4	0·40	<0·02
SB2009	*I*^{c}10 *Δ*719	1000	6	0·60	<0·02
SB2012	*I*^{c}13 *Δ*719	1000	4	0·4	0·02
SB2013	*I*^{c}14 *Δ*719	1000	4	0·4	<0·02
SB2016	*I*^{c}17 *Δ*719	1000	5	0·5	0·08
SB2018	*I*^{c}19 *Δ*719	1000	3	0·3	0·06

Phage P1bt as used in Cross I was employed in transduction experiment with strain SB1509 containing deletion 1109, as described in Table 2. Only Ara$^{+\,sl}$ transductants were found in these crosses.

† The spontaneous reversion $= \dfrac{\text{Ara}^{+\,sl} \text{ revertants of } \Delta 1109 \text{ (control)}}{\text{Ara}^{+\,sl} \text{ (transductants)}} \times 100$

recombinants were picked to homologous medium and replica-plated to score for Leu$^-$. In nine out of the ten crosses, Leu$^-$ types were found at a very low frequency ranging from 0·1 to 0·6%. In the cross with $I^c2\Delta719$, no Leu$^-$ recombinants were isolated. The Ara$^{+\,sl}$Leu$^-$ recombinants were isolated in pure culture and verified by progeny testing with F'*araC*$^-$/*C*$^-$ homogenotes to contain a deletion that encompasses the *C* gene and were shown to be not revertible to Leu$^+$. This evidence indicates, therefore, that these recombinants contain the original deletion of the recipient used in these crosses.

Cross II has established that the revertant mutant sites were closely linked to the *araC* gene. The fact that we were able to cross, in nine out of ten cases, the revertant mutant sites into strain SB1509 containing deletion 1109, indicates that these revertant mutant sites must lie to the left of deletion 1109 and therefore to the left of the *araC* gene; that is, the side of the *C* gene closest to gene *araB*. In the cross with $I^c2\Delta719$, we presume that I^c2 lies less than 0·1% recombination units from the deletion 1109. The fact that all 19 revertants produce a similar phenotype, that the mutant sites are closely linked to *araC* and that nine have been shown to map to the left of the *C* gene (as defined by deletion 1109), forms the basis for placing all 19 mutant sites within the initiator region of the L-arabinose complex located between genes *araB* and *araC*.

(iii) *Complementation analysis of the revertants*

Cis-*dominance test.* We initially constructed merodiploids of the type F'$A2I^+C^+$/$A^+I^c\Delta719$ for each of the $I^c\Delta719$ revertants plus, as a control, merodiploid F'$A2I^+C^+$/$A^+I^+\Delta719$, and analyzed these merodiploids and their related F$^-$ haploid strains for L-arabinose isomerase activity. If the constitutive levels of isomerase produced by the revertants $I^c\Delta719$ were the result of mutation in the initiator region (*araI*), one would expect that the isomerase levels of the merodiploids would be similar to those found with the respective haploid $I^c\Delta719$ strains; *i.e.* I^c should be *cis*-dominant to I^+.

Our experiments demonstrated that in a few cases the isomerase levels of the merodiploids were the same but, in a majority of cases, the levels were higher than those of the F$^-$ haploid strains (Table 6). The control merodiploid, F'$A2I^+C^+$/$A^+I^+\Delta719$, however, also showed a significant increase in isomerase activity over that of the F$^-A^+I^+\Delta719$ haploid. Because of this increased isomerase activity in the control merodiploid, we considered the possibility that such increases occurring in the merodiploids might obscure a *trans*-dominant effect of I^+ on I^c. Therefore, before we can fully assess the significance of these complementation analyses, it is necessary for us to understand the cause of these increases in isomerase activity and attempt to eliminate this effect, so as to uncover a possible cryptic *trans* effect of I^+.

We do know that this unexpected increase in non-induced isomerase levels is not the result of recombination and segregation in the merodiploid cultures. First of all, if by a recombination event, the episome in the control merodiploid now carried $A^+I^+C^+$, even if there were several copies of the episome per nucleus, this could not explain the increase in basal level of isomerase from 0·1 to 3·4 units, since the wild-type strain has a basal level of isomerase of only 0·1 unit. Similarly, in the experimental set of merodiploids, to explain the large increases in constitutive isomerase levels, one might suppose that a recombination event might have occurred producing

TABLE 6

L-Arabinose isomerase activity of merodiploids of the type F′ A2I^+C$^+$/A$^+$I^c*Δ719 and* F′ A2I^+C3/A$^+$I^c*Δ719*

Endogenote	haploid	L-Arabinose isomerase Non-induced diploid exogenote F′ *A2I$^+$C$^+$*	F′ *A2I$^+$C3*	Induced diploid exogenote F′ *A2I$^+$C$^+$*
I^{c}1 Δ719	5·2	6·8(2)		56
I^{c}2 Δ719	3·5	3·9		44
I^{c}3 Δ719	3·6	12	4·5	54
I^{c}4 Δ719	4·0	7·4(2)	3·9	44
I^{c}5 Δ719	5·8	11	4·1(3)	47
I^{c}6 Δ719	3·6	4·6	4·1(2)	52
I^{c}7 Δ719	4·8	9·8(2)		51
I^{c}8 Δ719	2·8	16	3·6	52
I^{c}9 Δ719	2·9	4·3	3·9	39
I^{c}10 Δ719	2·0	7·4		53
I^{c}11 Δ719	5·5	17		55
I^{c}12 Δ719	2·5	2·9	2·6	62
I^{c}13 Δ719	3·6	5·6	4·9	50
I^{c}14 Δ719	3·5	7·5		57
I^{c}15 Δ719	3·2	6·6		69
I^{c}16 Δ719	3·2	4·0(3)	4·2	57
I^{c}17 Δ719	3·5	5·2	4·0	80
I^{c}18 Δ719	2·1	4·7		42
I^{c}19 Δ719	2·4	4·6	3·4	63
A$^+$I$^+$ Δ719	0·1	3·4(4)	0·1	49
WT	0·08(28·8)†			
A2 I$^+$C$^+$	0·1			

Cell-free extracts were prepared from cells in the stationary phase of growth. Enzyme activity is expressed in μmoles of product formed/hr/μg protein. The haploid F$^-$ strains were initially analyzed as two groups with *I^{c}10Δ719* present in each group. There was no significant difference in the isomerase levels of *I^{c}10Δ719* in each run. In analyzing the non-induced merodiploids, usually the corresponding F$^-$ haploid strains were run in conjunction with the corresponding merodiploid and the isomerase levels were normalized to the isomerase levels of the F$^-$ haploid strain as initially determined in the group run of F$^-$ strains. In most cases the correction was a minor one. The analysis of the induced merodiploid was performed in 4 batches (*I$^+$Δ719*, *I^{c}1Δ719* to *I^{c}4Δ719*; *I^{c}5Δ719* to *I^{c}9Δ719*; *I^{c}10Δ719* to *I^{c}14Δ719*; *I^{c}15Δ719* to *I^{c}19Δ719*) together with the corresponding haploid strains. The isomerase values listed for the former are uncorrected. The figure in parenthesis indicates the number of independent analyses performed.

† Induced L-arabinose isomerase activity.

an episome of the genotype *A$^+$I^{c}C$^+$*. This, however, is a very unlikely event since, as we have shown, *I^c* is very closely linked to deletion 719. Besides, as shown by Englesberg *et al.* (1969), the *C$^+$* allele in this case (no deletion 719 in a *cis* position) would have an epistatic effect on the function of the *I^c* allele. In any case, by an analysis of the merodiploid cultures used for the preparation of enzyme extracts for L-arabinose isomerase activity, we have been able to rule out an explanation based upon genetic recombination. Each of the ten Ara$^+$ clones, isolated on eosin–methylene blue–Ara, from each culture, was shown to segregate Ara$^-$ progeny which were *A2C$^+$*.

To explore this phenomenon further, a second series of merodiploids was constructed and analyzed in which a C^- allele was substituted for the C^+ allele in the exogenote (F′$A2I^+C3/A^+I^c\Delta719$ as well as a control merodiploid, F′$A2I^+C3/A^+I^+\Delta719$) (Table 6). The L-arabinose isomerase levels of these merodiploids were determined and compared with those of the corresponding F$^-I^c\Delta719$ strains and F$^-I^+\Delta719$. In all cases the isomerase levels of the merodiploids containing the $A^-I^+C^-$ alleles in the exogenote were similar to the isomerase levels of the respective F$^-$ haploid strains. Thus the increases in isomerase activity found with some merodiploids of the type F′$A2I^+C^+/A^+I^c\Delta719$ and with merodiploid F′$A2I^+C^+/A^+I^+\Delta719$ are due to the product of the C^+ allele produced in the absence of the inducer. (See Discussion for explanation of this effect.) By eliminating this effect of the C^+ allele we have established, unambiguously, in the 11 cases tested, that I^+ has no effect on the constitutive expression of *araA cis* to I^c. Thus I^c is dominant to I^+.

It will be noted that the merodiploid F′$A2I^+C^+/A^+I^+\Delta719$ is hyperinducible in the presence of L-arabinose; the isomerase activity is nearly twice that of the fully-induced wild type. Merodiploids of the type $A2I^+C^+/A^+I^c\Delta719$ are in all cases also hyperinducible. The induced levels found are higher than those of the fully induced wild type and in some cases higher than those found for the control merodiploid $A2I^+C^+/A^+I^+\Delta719$. An explanation of differences in basal and induced levels of isomerase, found for the various I^c in merodiploids of the type F′$A2I^+C^+/A^+I^c\Delta719$, will be presented in the Discussion.

Cis–trans *dominance test*. The *cis* dominance of I^c to I^+ is characteristic of mutations at a controlling site producing a constitutive phenotype. If this were the case, the I^c alleles should have no *trans* effect. To test for a *trans* effect, we constructed and analyzed merodiploids of the type F′$A^+B^-I^+C^+/A2B^+I^c\Delta719$ for five of the I^c mutations, and as a control, merodiploid, F′$A^+B^-I^+C^+/A^-B^+I^+\Delta719$. Non-induced and induced (in some cases) isomerase and kinase activity was determined for each of the merodiploids and the appropriate haploid strains. In all five cases, there was no demonstrable *trans* effect of the I^c alleles (Table 7). For instance, F$^-A^+B^+I^c\Delta719$ has a constitutive level of isomerase equal to 4·36. When *A2*, a mutation in the *araA* structural gene, was introduced into this strain, isomerase activity was reduced to less than 0·01 unit. This F$^-$ strain still carries the constitutive marker as evidenced by its kinase activity. F$^-A^+B24I^+C^+$ has a basal isomerase level of 0·35 unit. If the I^c allele has a *trans* effect, the *araA* gene in the episome should be activated and the merodiploid would be expected to have, at a minimum, approximately 4 units of isomerase activity. A basal level of 0.37 unit of isomerase was obtained with the control merodiploid (no I^c allele), as would be predicted. A value of 0·43 was obtained with the experimental merodiploid (F′$A^+B24I^+C^+/A^-2B^+I^c1\Delta719$), a value not significantly different from 0·37. The constitutive production of kinase in the merodiploid is evidence for the presence of a functional I^c allele. Essentially similar results were obtained for each of the I^c mutants tested. Besides the enzymic analysis, the genotype of each merodiploid culture employed in the preparation of cell-free extracts was verified. Crossover and segregation was negligible in the experiments recorded. Therefore with the five I^c alleles tested, I^c is found to be *trans* recessive to I^+.

It will be noted that kinase activity did not increase in the merodiploids containing the I^c allele in the endogenote, while there was a small but significant increase in kinase activity in the control merodiploid containing the I^+ allele in the endogenote. On the basis of the demonstrated co-ordinate expression of the structural genes

TABLE 7

L-Arabinose isomerase and L-ribulokinase
Activity of merodiploids of the type F′ A⁺B24I⁺C⁺/A2B⁺I^c*Δ719*

Strain	Arabinose genotype	Non-induced		Induced	
		Kinase	Isomerase	Kinase	Isomerase
SB3617	F′ *A⁺B24I⁺C⁺/A2 B⁺I^{c}1 Δ719*	0·52	0·43	11	88
SB2149	F⁻ *A2B⁺I^{c}1 Δ719*	0·55	0·01		
SB2000	F⁻ *A⁺B⁺I^{c}1 Δ719*	0·38	4·36		
SB3619	F⁻ *A⁺B24I⁺C⁺/A2B⁺I^{c}13 Δ719*	0·64	0·72	10	73
SB2151	F⁻ *A2B⁺I^{c}13 Δ719*	0·61	0·01		
SB2012	F⁻ *A⁺B⁺I^{c}13 Δ719*	0·35	3·52		
SB3620	F′ *A⁺B24I⁺C⁺/A2B⁺I^{c}14 Δ719*	0·58	0·36	12	94
SB2152	F⁻ *A2B⁺I^{c}14 Δ719*	0·61	0·01		
SB2013	F⁻ *A⁺B⁺I^{c}14 Δ719*	0·38	5·60		
SB3621	F′ *A⁺B24I⁺C⁺/A2B⁺I^{c}17 Δ719*	0·46	0·44	8·9	99
SB2153	F⁻ *A2B⁺I^{c}17 Δ719*	0·46	0·01		
SB2016	F⁻ *A⁺B⁺I^{c}17 Δ719*	0·61	4·63		
SB3622	F′ *A⁺B24I⁺C⁺/A2B⁺I^{c}19 Δ719*	0·38	0·35	7·6	79
SB2154	F⁻ *A2B⁺I^{c}19 Δ719*	0·49	0·02		
SB2018	F⁻ *A⁺B⁺I^{c}19 Δ719*	0·41	4·86		
SB3636	F′ *A⁺B24 I⁺C⁺/A2B⁺I⁺ Δ719*	0·17	0·37	9·7	78
SB5312	F⁻ *A2B⁺I⁺ Δ719*	0·01	0·01		
SB1094	F⁻ *A⁺B⁺I⁺ Δ719*	0·01	0·16		
UP1001	F⁻ *A⁺B⁺I⁺C⁺*	0·01	0·08	7·2	19
UP1027	F⁻ *A⁺B24C⁺*	0·01	0·35		

See Table 2 for explanations.

araB, *araA*, *araD* (Englesberg *et al.*, 1965), one would have expected similar results to those obtained with the isomerase. We do not fully understand this discrepancy. It is possible that the absence of a co-ordinate increase in kinase activity may be a reflection of increased instability of the kinase molecule at low concentrations, in the absence of substrate as a result of the I^c mutations. This possible instability of the kinase in I^c mutants is indicated by the finding that the ratio of kinase activity to isomerase activity in the F⁻ strains of the type $A^+B^+I^c\Delta719$ is about three to five times less than the comparable ratio in the induced wild type.

4. Discussion

(a) *The nature of* I^c *mutants*

From an Ara⁻ strain of *Escherichia coli* containing a deletion that excises most, if not all of gene *araC*, 19 independent revertants have been isolated. All the revertants were shown to be the result of a secondary mutation (I^c) closely linked to but separable from the deletion itself and all possess the characteristics required of mutations in the initiator region of the L-arabinose operon.

(1) They produce a *cis*-dominant constitutive phenotype, and all those tested showed no *trans* effect.

(2) All 19 revertant mutant sites map within the initiator region of the L-arabinose operon. Three-factor transduction crosses, with phage previously grown on each of the revertants and with *leuB1* and *araC5 leuB1* as recipients, have established that each of the 19 revertants map within the *ara leu* region of the chromosome and are closely linked to and separable from the *araC* segment of the chromosome. In nine out of ten cases, it was possible to cross, at low frequency, the I^c mutant sites into a strain containing deletion 1109, a deletion extending from *araC* to the leucine operon. This placed nine I^c mutant sites to the left of the deletion, and thus to the left of *araC*, in the region proposed to contain the initiator site for the L-arabinose operon. We presume that the exceptional site (I^c2) is also to the left of the *C* gene but so close to deletion 1109 that, in the number of transductants analyzed, we failed to pick up an $I^c\Delta1109$ recombinant. Because of the close linkage of all 19 mutant sites to deletion 719 and the similarity of their phenotypes, it seems likely that all 19 mutant sites reside within the same region.

(3) Several of the revertants that have the same level of L-arabinose isomerase activity in the haploid state (in the presence or absence of the inducer) have significantly different hyperinducible levels of this enzyme in merodiploids containing a C^+ allele in the *trans* position. Thus in these cases, the I^c mutation, besides permitting expression of the structural genes *cis* to it in the absence of a *C* gene product, has altered the sensitivity of the operon to the *C* gene product. Thus, by definition, these mutations must have occurred in the initiator region of the operon.

(b) *A single regulatory gene for the* ara-OIBAD *operon*

We have been unable to isolate any revertants of a *C* gene deletion which have the properties of mutations in a proposed repressor gene (*R*). In both the *lac* and *gal* systems, where negative control has been established by genetic evidence, mutations to constitutivity in the repressor gene occurs at a relatively high frequency (Willson *et al.*, 1964; Shapiro, 1967). It is expected that, at least, nonsense mutants and deletions would abolish the activity of a repressor gene. The failure to find any such mutants for the Ara system down to a *spontaneous* frequency of less than 1×10^{-10} is a strong indication that no such repressor gene exists. The possibility, however, is not completely ruled out by these studies, since it is conceivable that such a gene does exist but either is essential to bacterial growth or is present in duplicate copies.

Other properties of the *ara* operon and the *araC* gene, as indicated in the discussion below, suggest strongly that the *C* gene product directly interacts with the *ara* operon-controlling elements. Thus it is unlikely that the *C* gene functions as proposed in the negative control internal induction model.

(c) *Modified positive control model*

The increase in L-arabinose isomerase activity in merodiploid $A^-I^+C^+/A^+I^+\Delta719$ and with most $I^c\Delta719$ mutants in merodiploids of the type $A^-I^+C^+/A^+I^c\Delta719$, in the absence of the inducer, over that of the corresponding F^- haploid strains has been examined in greater detail by Englesberg *et al.* (1969). We have shown in this article that this increase is the result of a *trans*-acting C^+ allele on the activity of the *araA* gene *cis* to the deletion, since the substitution of a C^- allele in the merodiploids described above does away with this increase in isomerase activity. Thus a product

formed by the C^+ allele (in the absence of inducer) is responsible for this phenomenon. The fact that the isomerase level in merodiploid $A^-I^+C^+/A^+I^+\Delta 719$ is approximately 34 times higher than the basal level of the wild type ($A^+I^+C^+$) and that there is no *trans* effect (compare isomerase and kinase levels of UP1027, SB5312 and SB3636 in Table 7 and corresponding data in Table 6) suggests that deletion 719 must have a significant role in this increased expression of the *cis araA* gene. This is in fact shown to be the case, since no such increase in basal or constitutive levels of isomerase is found when deletion 766 (a deletion whose left end terminates within the *C* gene between two C^- point mutants) is used in place of deletion 719 (Englesberg *et al.*, 1969). In fact, with merodiploids of the type $A^-I^+C^+/A^+I^c\Delta 766$, the C^+ allele (but not the C^- allele) on the episome severely depresses the constitutive expression of the *araA* gene *cis* to the I^c mutations.

To explain these results, Englesberg *et al.* (1969) have proposed a modification of the simplified model for positive control, as described in the Introduction. They propose that (1) P1, the initial product of the *araC* gene, is a repressor with a site of attachment, the operator, located between *araI* and *araC*; (2) P1 is in equilibrium with P2, the activator, and with P1 and P2 attached to their respective controlling sites *araO*, the operator, and *araI*, the initiator; (3) L-arabinose, the inducer, removes P1 from *araO* and shifts the equilibrium to P2. P2 acts at *araI* and thereby stimulates

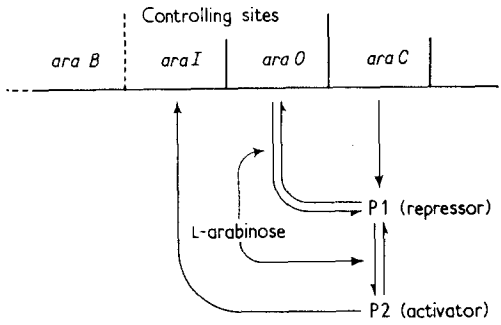


FIG. 1. Positive control model.

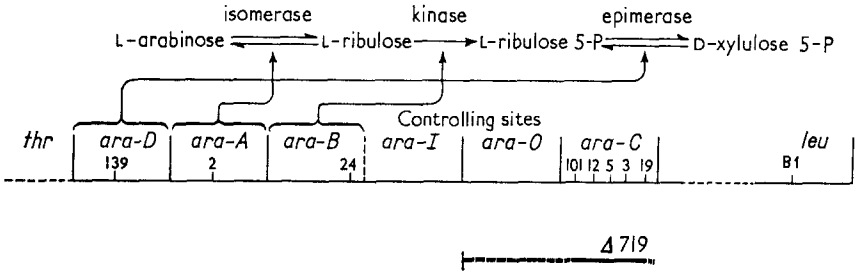


FIG. 2. The L-arabinose gene–enzyme complex.

Structural genes: *araB*, L-ribulokinase; *araA*, L-arabinose isomerase; *araD*, L-ribulose 5-phosphate 4-epimerase.

Controlling sites: *araI*, initiator site—this is the position of I^c mutations and the site for activator (P2) function. *araO*, operator site—this is the site for repressor (P1) function.

The numbers indicate the mutants used in this study.

Deletions—solid lines indicate the portion of the genome excised by the deletion as determined by genetic mapping with F′ Ara⁻ homogenotes. The dashed portion of the lines extending the deletion are based upon complementation, and enzymic analysis.

the expression of the operon. It is necessary both for P2 to be present at *araI* and for P1 to be removed from *araO* for full expression of the operon to occur (Fig. 1). With the exception of a more precise positioning of the operator, this model is essentially the one previously proposed as a possibility to explain the dominance of C^+ to C^c (Sheppard & Englesberg, 1967).

According to this model, the increase in isomerase levels in merodiploids of the type $A^-I^+C^+/A^+I^+\Delta 719$ over that of $F^-A^+I^+\Delta 719$ is explained on the basis that deletion 719 excises all or part of the operator site (Fig. 2). In the absence of a functional operator site, the amount of P2 existing in equilibrium with P1 is able to partially turn on the expression of the structural genes *cis* to the deleted operator. Depending upon how the I^c mutation has modified the initiator site, the amount of P2 present in the absence of inducer may or may not further activate the structural genes *cis* to $I^c\Delta 719$. However, in the presence of inducer, there is sufficient P2 to hyperinduce the operon in all the I^c mutants analyzed, although the I^c mutations appear to govern the extent of hyperinducibility.

(d) *Elimination of other explanations for the properties of the* C *gene*

The properties of regulatory mutants in the maltose (Schwartz, 1967) and rhamnose (Power, 1967) systems indicate that the structural genes for these pathways are also under positive control. Schwartz (1967) has proposed other hypotheses as alternatives to positive control to explain these properties. In terms of the *ara* operon these would be (1) that the supposed positive control gene *araC* actually produces an enzyme that converts L-arabinose into a true inducer (this we have eliminated: see Discussion above); (2) that the *araC* gene produces a protein subunit required for the activity of the other enzymes in the *ara* operon; and (3) that *araC* is a structural gene for a component of the L-arabinose permease system.

Both of these alternatives have been eliminated in the L-arabinose system. First of all it has been shown that it is possible to obtain high levels of L-arabinose isomerase activity, similar to that obtainable in the wild type, in mutants containing an *araB-leu* deletion; *i.e.* a deletion that cuts out *araI*, *araO*, and *araC* and fuses the L-arabinose isomerase structural gene (*araA*) to the leucine operon. Thus it is unlikely that gene *araC* provides a required polypeptide chain for the L-arabinose isomerase (Kessler & Englesberg, 1969). (There are also many other cogent arguments against this model.) Second, it has been recently shown that C^- mutants do have L-arabinose permease activity (Englesberg, unpublished results), although lower than that of a reference *araA* strain. Thus although there is evidence that the *C* gene controls permease activity (see also Englesberg *et al.*, 1965), certain C^- mutants that are recessive to C^+ and C^c alleles are still able to concentrate L-arabinose internally.

This investigation was supported in part by National Science Foundation grant GB5392, a U.S. Public Health Service grant GM13607 and a contract between the University of California, Santa Barbara and the Office of Naval Research. One of us (D.S.) was a recipient of a U.S. Public Health Service Fellowship GM11, 751-02.

We would like to thank Mr R. Calsen for his assistance in the isolation and initial characterization of the revertants and Dr J. Beckwith for helpful suggestions in preparing the manuscript. A preliminary report of this work has appeared (Englesberg, E. & Squires, C., 1968 *Proc. 12th Int. Congr. Genetics*, section 2, p. 47).

REFERENCES

Bourgeois, S., Cohn, M. & Orgel, L. E. (1965). *J. Mol. Biol.* **14**, 300.
Boyer, H., Englesberg, E. & Weinberg, R. (1962). *Genetics*, **47**, 417.
Cribbs, R. & Englesberg, E. (1964). *Genetics*, **49**, 95.
Englesberg, E., Irr, J., Power, J. & Lee, N. (1965). *J. Bact.* **90**, 946.
Englesberg, E., Squires, C. & Meronk, F., Jr. (1969). *Proc. Nat. Acad. Sci., Wash.* in the press.
Gross, J. & Englesberg, E. (1959). *Virology*, **9**, 314.
Irr, J. & Englesberg, E. (1967). *Bact. Proc.* p. 54.
Jacob, F. & Monod, J. (1961). *J. Mol. Biol.* **3**, 318.
Kessler, D. & Englesberg, E. (1969). *J. Bact.* in the press.
Lowry, O., Rosebrough, N., Farr, A. & Randall, R. (1951). *J. Biol. Chem.* **193**, 265.
Muller-Hill (1966). *J. Mol. Biol.* **15**, 374.
Power, J. (1967). *Genetics,* **55**, 557.
Schwartz, M. (1967). *Ann. Inst. Pasteur,* **112**, 673.
Shapiro, J. A. (1967). Ph.D. thesis, University of Cambridge.
Sheppard, D. & Englesberg, E. (1966). *Cold Spr. Harb. Symp. Quant. Biol.* **31**, 345.
Sheppard, D. & Englesberg, E. (1967). *J. Mol. Biol.* **25**, 443.
Willson, C., Perrin, D., Cohn, M., Jacob, F. & Monod, J. (1964). *J. Mol. Biol.* 8, 582.

Week 4

Recombination and Transposition

Horizontal Gene Transfer

J. Mol. Biol. (1968) **31**, 487–505

Integration-negative Mutants of Bacteriophage Lambda

MAX E. GOTTESMAN AND MICHAEL B. YARMOLINSKY

Section on Microbial Genetics, Laboratory of Molecular Biology
National Institute of Arthritis and Metabolic Diseases
National Institutes of Health
Bethesda, Maryland 20014, U.S.A.

(*Received 25 May 1967, and in revised form 29 September 1967*)

Mutants of bacteriophage lambda unable to form stable lysogens have been obtained by a simple screening procedure. These mutants, designated *int*, form turbid plaques. The procedure used for their isolation also allows the titration and selection of rare lysogens in a predominantly non-lysogenic population. The recombinational behavior of *int* mutants in wild-type and recombination-deficient (*rec*) bacteria suggests that the *int*$^+$ gene determines a site-specific function involved in normal prophage integration and detachment. The Int function can be provided in *trans* by an *int*$^+$ phage. The expression of the *int*$^+$ gene is subject to repression by the phage immunity substance.

Single *int* lysogens give a low yield of active phage upon induction, whereas double lysogens, when tandem, give nearly normal phage yields. Both single and double *int* lysogens give large yields of transducing particles. Recombination between vegetative *int* phages and productive induction of *int* lysogens, both of which occur in recombination-deficient bacteria, may be attributed to a separate function (Red) operative in phage-infected or induced cells. The Red function is not available for use in integration of a superinfecting phage, a function which the product of the bacterial *rec*$^+$ gene can accomplish.

1. Introduction

Bacteriophage λ engages in two kinds of genetic recombination: (1) mating between λ genomes and (2) localized recombinations between λ and the bacterial chromosome. Recombinations of the first kind occur during vegetative growth of phage and involve the substitution of genetic regions of one phage genome by homologous regions of another (Meselson & Weigle, 1961; Kellenberger, Zichichi & Weigle, 1961*a*). Thus, if a sensitive bacterium is mixedly infected with genetically marked λ, the burst will contain recombinant particles in which the genotypes of the input phage have recombined.

Recombinations of the second kind occur in the formation of stable lysogens following infection and in the transition to the vegetative state following lysogenic induction. Lysogenization by λ entails the linear integration of λ DNA into the bacterial chromosome (Calef & Licciardello, 1960; Campbell, 1963; Franklin, Dove & Yanofsky, 1965; Rothman, 1965; Signer, 1966; Hoffman & Rubenstein, manuscript in preparation). The integration process is visualized as recombination between closed forms of the two genomes at the b2$^+$ region of λ and a specific, perhaps homo-

logous, region of the *Escherichia coli* chromosome (Campbell, 1962; Zichichi & Kellenberger, 1963). The reciprocal event, occurring early in the process of phage induction, can account for prophage detachment (Campbell, 1962; Weisberg & Gallant, 1967).

Bacteria deficient in chromosomal recombination have been described (Clark & Margulies, 1965). When these *rec* bacteria† are mixedly infected with genetically marked bacteriophage λ, the mating of phage genomes is unimpaired (Brooks & Clark, 1967; Zissler, 1967). Likewise, the defect in *rec* bacteria does not interfere with lysogenization and prophage detachment (Brooks & Clark, 1967; Hertman & Luria, 1967). Thus, λ can provide recombination-deficient bacteria with one or more phage-determined functions enabling its own genome to engage in recombination of both kinds.

The existence of a site-specific phage function involved in the process of integration has been adduced from the behavior of deletion mutants of φ80 unable to lysogenize normally unless "helped" by wild-type phage with the appropriate attachment specificity (Signer & Beckwith, 1966). Such mutants differ from λb2 deletion mutants, which have been reported not to form stable single lysogens in *E. coli* K12 under any conditions (Kellenberger, Zichichi & Weigle, 1961*b*; Zichichi & Kellenberger, 1963). Whereas the defect in λb2 appears to be structural (Campbell, 1965), resulting in an inability to pair appropriately for lysogenization (Fischer-Fantuzzi, 1967), the φ80 mutants of Signer & Beckwith (1966) appear deficient in a diffusible enzyme.

We describe here the isolation and properties of non-integrating mutants of λ the characteristics of which suggest that they are point mutants deficient in a site-specific integration function. These mutants also exhibit defects in prophage detachment upon induction and in superinfection curing. Thus, *int* mutants appear defective in performing recombinations of the second kind. However, they do recombine with one another in a recombination-deficient host, i.e. the *int* defect does not markedly impair recombinations of the first kind. The two kinds of genetic recombination, therefore, are resolved genetically. The existence of recombinations between *int* phages in a *rec* host requires that there be an additional recombination function. Mutants of λ defective in this function have been isolated recently (Signer; Echols, personal communications) and they are designated *red* for recombination-deficient. Deletion mutants defective in the same function have been isolated as well in φ80–λ hybrid phage by Franklin (1967). In what follows we distinguish the influence which each of the three recombination functions, Int, Rec and Red, exerts upon recombinational events in which λ participates.

The isolation of phage mutants with similar properties has been achieved independently in several laboratories (phage P22: Smith & Levine, 1967; phage λ: Zissler, 1967; Gingery & Echols, 1967). Deletion mutants of φ80 and φ80–λ hybrid phage, which appear similar in some respects, have been studied by Signer & Beckwith (1966), and Franklin (1967). In addition, deletion mutants of φ80–λ hybrid prophage, defective in prophage detachment, have been studied by Franklin *et al.* (1965), Gratia (1966), Dove (1967) and Franklin (1967).

† In conformity with the nomenclature proposed by Demerec, Adelberg, Clark & Hartman (1966), the following abbreviations are used: *rec*, *rec*$^+$ refer to mutant and wild-type alleles of a bacterial gene the Rec function of which is required for recombination. Similarly, *int*, *int*$^+$ and Int are used in reference to a phage-determined integration function and *red*, *red*$^+$ and Red in reference to a function involved in generalized recombination, which is found in phage-infected bacteria. The allele of supII conferring non-permissiveness towards amber mutants is designated supII (Signer, Beckwith & Brenner, 1965).

2. Materials and Methods

(a) *Media*

Tryptone broth (TB): 1% Tryptone, 0·5% NaCl, 10^{-3} M-$MgSO_4$ and 1·0 μg thiamine hydrochloride/ml. Tryptone agar: 1% Tryptone, 0·25% NaCl, 1·0 μg thiamine hydrochloride/ml. and 1·1% agar. Tryptone top agar: 1% Tryptone, 0·5% NaCl and 0·7% agar. Tryptone soft agar: 1% Tryptone, 0·5% NaCl and 0·35% agar. Luria Broth (LB) (Luria, Adams & Ting, 1960): 1% Tryptone, 0·5% NaCl and 0·5% yeast extract (Difco). MacConkey–galactose agar is the MacConkey agar of the *Difco Manual* (Difco Laboratory, Detroit, Mich.) except that galactose was substituted for lactose. All of the above media were adjusted to pH 7·0 with NaOH. EMBO agar is the EMB agar of Campbell (Campbell, 1957) with sugar omitted. The minimal medium, M56 (Monod, Cohen-Bazire & Cohn, 1951) was used diluted and appropriately supplemented with amino acids, thiamine and 0·2% glucose and solidified with 2% agar. Phosphate buffer: 0·7% Na_2HPO_4, $7H_2O$, 0·3% KH_2PO_4, 0·4% NaCl and 10^{-3} M-$MgSO_4$.

(b) *Sensitive bacteria*

The following *E. coli* K12 derivatives were used: *rec* strain 152 (Meselson, unpublished work) and its *rec*$^+$ parent, strain 28, Sm^r derivatives of W3102 galK2 (Lederberg, 1960) non-permissive towards *sus* phage mutants, from Dr M. Meselson; *E. coli* C600 (Appleyard, 1954) used in titrating *sus* phage mutants, from Dr F. Jacob; M5073, a supII homomerozygote (non-permissive) used as donor of F′1*gal*$^+$ (an F′ isolated by E. L. Wollman and described in Liedke-Kulke & Kaiser, 1967), from Dr E. Signer; PL2, defective in uridine diphosphoglucose 4-epimerase (epimerase) from Dr G. Buttin (Buttin, 1963*b*) used as recipient in *gal* transduction; W602, defective in biotin synthesis (Wollman, 1963), from Dr J. Rothman, used as recipient in *bio* transduction.

Bacterial counts were done by plating in soft agar.

(c) *Phages*

	Phage	Derivation or reference	Source
(1)	λ*susA11*	a	F. Jacob
(2)	λ*susN7,53*	a	J. Weil
(3)	λ*susP3*	a	D. Korn
(4)	λcI*72*	b	V. Bode
(5)	λcI*857*	c	F. Jacob
(6)	λcI*857susA11*	(1) & (5)	This work
(7)	λ*int6c*	λcI*857int6* spontaneous mutant	This work
(8)	λb2	d	C. A. Thomas, Jr.
(9)	λb2*c*	d	M. Gellert
(10)	λb2cI*857susA32susJ27*	(5), (8) & (15)	M. Yarmolinsky
(11)	434	e	F. Jacob
(12)	434hy	f	F. Jacob
(13)	434hycI*1*	(12) by ultraviolet mutagenesis	M. Yarmolinsky
(14)	434hycI*2*	(12) by ultraviolet mutagenesis	M. Yarmolinsky
(15)	434hy*susA32susJ27*	g	D. Korn
(16)	434hy*int6*	(12) & λ*int6*	This work
(17)	434hy*int6c*	(16) spontaneous mutant	This work
(18)	434hycI*1susP3*	(3) & (13)	M. Yarmolinsky
(19)	434hycI*1susP3int41*	(18) & λ*int41*	This work
(20)	434hy*susN7,53*	a & f	L. Pereira da Silva
(21)	21	e	H. Wiesmeyer
(22)	21gp	h	A. D. Kaiser
(23)	21*c*	(21) by ultraviolet mutagenesis	M. Yarmolinsky
(24)	21hy5	h	M. Liedke-Kulke
(25)	φ80	i	A. Weissbach

References: (a) Campbell, 1961; (b) Kaiser, 1957; (c) Sussman & Jacob, 1962; (d) Kellenberger, Zichichi & Weigle, 1961*a*; (e) Jacob & Wollman, 1956,1961; (f) Kaiser & Jacob, 1957; (g) Radding & Kaiser, 1963; (h) Liedke-Kulke & Kaiser, 1967; (i) Matsushiro, 1963.

Phage (13) was selected for the capacity to make a turbid plaque at 30°C and a clear plaque at 40°C. Lysogens of (13) are induced in Tryptone broth at temperatures above 36°C. The phage is not inducible by ultraviolet irradiation or by mitomycin C. Its properties will be described elsewhere.

Phage (14) forms a clear plaque at all normal incubation temperatures. It is complemented by phage (13) at low temperatures.

Lambda *cI857* was derived from λInd and is presumed to have retained this character in the several derivatives made from it.

(d) *Phage assays and phage stocks*

Assays of free phage and infective centers: pour-plates of indicator bacteria and phage on Tryptone agar were incubated at 39 to 40°C to distinguish the plaques of phages thermally induced at this temperature from those which are not. Higher temperatures were avoided because certain *sus* mutants do not plate efficiently on C600 above 40°C. Mixtures of 434 and λ were resolved by plating on C600(λ) and C600(434). Stocks of phage were obtained by mitomycin C (2 μg/ml.) or heat induction of appropriate lysogens grown in TB, by infection of bacteria in TB with 0·01 M-$MgSO_4$ added, or by the confluent lysis method using Tryptone soft agar.

Infection with phage was performed as follows: saturated overnight bacterial cultures were harvested, washed with phosphate buffer, suspended in 0·5 vol. of 0·01 M-$MgSO_4$, and aerated for 1 hr at 33°C. Bacteria thus treated did not lose viability or show altered sensitivity to phage for at least 4 weeks when stored at 5°C. The starved cells were mixed with phage suspended in TB, and adsorption was allowed to proceed for 20 min at 33°C. In all experiments, adsorption, measured as loss of free phage, was adequate to ensure infection of at least 80% of the cells.

(e) *Mutagenesis and preparation of* λint *mutants*

Mutagenesis of λ was performed essentially according to the method of Adelberg, Mandel & Chen (1965). C600(λcI*857*) was grown in 5 ml. TB at 33°C to about 10^8 cells per ml., and then aerated at 39°C for 15 min. The culture was chilled, harvested by centrifugation, and resuspended in 25 ml. of 0·05 M-Tris maleate buffer (pH 6·1) containing 10^{-3} M-$MgSO_4$ and 16 mg of *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (Aldrich Chemical Co., Milwaukee, Wisc.). After incubation for 30 min at room temperature, the suspension was diluted 100-fold into TB, distributed among separate tubes, and aerated for 120 min at 38°C. Chloroform was then added and the phage plated at 33°C on strain 152 as indicator. The centers of normal-appearing, turbid plaques were picked to Tryptone agar and to EMBO seeded with about 10^9 co-immune clear-plaque phage. After overnight growth at 33°C, *int* mutants were identified by the dark mottled appearance of the colony arising on EMBO. The corresponding colony on Tryptone agar was transferred to TB, aerated for 2 hr at 39°C, and chloroformed. The *int* phage in the lysate was purified by repeated single-plaque isolations. Phage stocks were obtained by the confluent lysis method using Tryptone soft agar, or by induction of a tandem homogenotic lysogen constructed with the aid of helper phage.

Lambda *cI857susA11int21* was prepared by a similar technique from C600(λcI*857susA11*) using C600 indicator bacteria.

(f) *Construction of* int$^+$ *lysogens*

N103: 152(λcI*857susA11*), N173: 28(λcI*857susA11*), and N192: M5073 bearing an episomal 434hy prophage, were prepared by infection of *E. coli* strains 152, 28 and M5073, respectively.

(g) *Construction of* int *lysogens*

Sensitive cells were infected with λ*int* and hetero-immune *int*$^+$ helper, each at a multiplicity of about 10, and the λ lysogens were recovered from an EMBO plate seeded with about 10^9 λ*intc*. Alternatively, rare *int* lysogens could be obtained from EMBO plates spread with *int*$^+$ clear-plaque phage when cells were infected with an *int* mutant alone. The high concentrations of infected bacteria which must be applied to the EMBO plates

when helper is omitted results in the appearance of λ-resistant clones on the plate. Such clones may be detected by their slimy character at 33°C and/or their ability to give a pink colony on EMBO seeded with a hetero-immune clear-plaque phage, e.g. 434hyc.

Lysogens prepared without added helper were: N111: 152(λcI*857int6*), a low yielder; and N123: 152(434hy*int6*).

N214: 28(λcI*857int6*), a low yielder, was prepared with helper 21hy5. N121: 152(λcI*857int6*), a normal yielder; N122: 152(λcI*857int6*, 434hy); N171: 152(λcI*857susA11int6*) and N172: 28(λcI*857susA11int6*) were constructed with helper 434hy.

N201: 152(434hy*int6*, λcI*857int6*) was constructed by superinfecting N123 with λcI*857int6*; N178: 152(λcI*857int6*, 434hy*int6*) was obtained by superinfection of N111 with 434hy*int6*. Both lysogens were recovered from EMBO plates seeded with *int*$^+$ clear-plaque phage.

N202: 152(λcI*857int6*)/F′ (434hy) was constructed as follows: N111 was infected with 434hy*int6* and then mated with N192 grown in LB to about 2×10^8 cells/ml. Double lysogens were detected by spreading the mating mixture on an EMBO plate seeded with 434hy*int6c* and λ*int6c*. The double lysogens were then streaked on MacConkey–galactose plates and galactose-positive clones which gave frequent galactose-negative segregants, owing to loss of the *gal*$^+$ episome, were selected. The galactose-negative segregants were found to have lost 434 but not λ immunity.

N229: M5073 bearing an episomal λcI*857int6* was prepared by infecting M5073 with λcI*857int6* and helper 434hy. Lambda lysogens in which the λ had integrated into the episome were selected by their behavior on MacConkey–galactose plates. At 33°C such clones gave galactose-positive colonies whereas at 39°C, only galactose-negative segregants, cured of the episome and its prophage, survived.

N237: 152(434hy*int6*)/F′(λcI*857int6*) was constructed by infecting N123 with λcI*857int6* and mating with N229. Double lysogens, obtained as described for N202, were transferred to MacConkey-galactose plates. Those which gave galactose–positive colonies at 33°C and galactose-negative, λ-sensitive segregants at 39°C, were selected.

(h) *Complementation between* int *mutants*

Cells infected with one or more *int* phages were transferred by means of a platinum wire to EMBO seeded with clear-plaque λ. The wire was first dipped into the cell suspension and then stabbed into the EMBO plate. The presence of 10 lysogens (N103) among about 10^4 non-lysogens transferred with each dip can be detected.

(i) *Measurement of lysogenization frequency*

Infected bacteria were spread on EMBO plates seeded with co-immune clear-plaque phage. Lysogens appeared as pink-staining colonies after about 20 hr of incubation at 33°C. To eliminate phage-resistant clones and to detect lysogens carying a second hetero-immune prophage, the putative lysogens were transferred to EMBO plates seeded with clear-plaque phage of the appropriate immunity and scored as before according to the appearance of the resulting colonies.

Lysogenization frequencies were also determined by plating infected bacteria on Tryptone agar, where no selection pressure favoring lysogens is exerted, and subsequently transferring the colonies to EMBO-indicator plates. Equivalent results were obtained with both methods.

(j) *Titration of* gal *and* bio *transducing particles*

The epimerase-negative strain PL2 or the biotin-negative strain W602 was grown to saturation at 34°C and starved in 0·01 M-$MgSO_4$ for 1 hr. Strain PL2 at 2×10^9 cells/ml. or W602 at 5×10^8 cells/ml. was mixed with helper phage at a multiplicity of exposure of 4. The helper phage used in these experiments was λcI*857* induced from the corresponding lysogen. The phage to be titrated was added, and the mixture incubated for 20 min at 34°C. For assays of *gal* transduction, the mixture was diluted at least fivefold, and 0·05 ml. spread on MacConkey–galactose agar plates. These were incubated 22 to 24 hr at 34°C, and the red colonies counted. For assays of *bio* transduction, the mixture was poured in minimal top agar onto supplemented minimal plates and the colonies counted after 40 hr at 32°C. Assays were shown to be linear with respect to quantity of tranducing lysate added.

(k) *Measurement of* gal *escape synthesis*

The activity of UDP glucose: α-D-galactose-1-phosphate uridylyl transferase was assayed in centrifuged samples of cell suspensions essentially by the method of Buttin (1963*a*).

(l) *Superinfection curing*

A thermo-inducible *rec* lysogen was superinfected with a hetero-immune phage, diluted at least 50-fold in TB and aerated for 2 hr at 33°C to allow segregation of non-lysogenic nuclei. Prophage bearing *sus* mutations were used to minimize the possibility of re-infection of cured cells. The culture was then incubated at 41°C for 15 min to induce and kill lysogens, and plated for survivors at 39°C. The frequency with which these survivors had acquired the immunity of the superinfecting phage was determined as described above.

3. Results

(a) *Isolation of non-integrating* (int) *mutants and the selection of rare lysogens*

Plaques of wild-type λ are turbid because lysogenic immune cells arise and grow during formation of the plaque. Phage mutants which are unable to establish true lysogeny may form turbid plaques, the turbid centers of which are comprised of abortive lysogens. These continually give rise to non-immune progeny, which may become re-infected. If such abortive lysogens are picked from the center of a mutant plaque to a plate previously seeded with about 10^9 clear-plaque type co-immune phage, infection of the non-immune progeny by the clear phage kills these bacteria as they are formed. If the indicator phage are seeded on EMBO agar, the resultant partially lysed colonies are easily scored by their small dark and irregular appearance (Lederberg & Lederberg, 1953). True lysogens, which transfer immunity to all their progeny, invariably form a large pink colony. Plaques of the deletion mutant λb2, which cannot lysogenize, may be distinguished readily from plaques of λb2$^+$ by this test (Plate I, upper two rows).

The test may also be used to screen for *int* mutants in stocks of heavily mutagenized phage λ. As many as 5% of the turbid plaques score as *int*. Fifteen independent mutants have been isolated by this test (e.g. Plate I, rows 3 and 4).

In the majority of experiments described below, EMBO plates which had been spread with clear-plaque *int* phage were used in selecting or titrating infrequent true lysogens in a predominantly non-lysogenic population. Lysogenization frequencies have been corrected for the contribution of λ-resistant bacteria which appear on the plates with a frequency of about 3×10^{-4}.

(b) *Lysogenization by* int *phages on single and mixed infection*

Of fifteen independent *int* mutants obtained by *N*-methyl-*N'*-nitro-*N*-nitroso-guanidine mutagenesis, 12 have been examined for complementation with λ*cI857int6* and all 15 for complementation with λb2. Typical results are presented in Plate I. Whereas all *int* mutants were complemented by λb2, no demonstrable complementation occurred among the pairs of *int* mutants tested. We have also tested our *int* mutants (6, 21 and 41) for complementation with *int4* isolated by Zissler (1967). No complementation was observed. This complementation behavior is consistent with the interpretation that all the mutations tested belong to a single cistron lying outside the region deleted in λb2.

When lysogens prepared by complementation between λb2 and 434hy*int* are analyzed, it is found that the complementation is non-reciprocal (Table I). When only

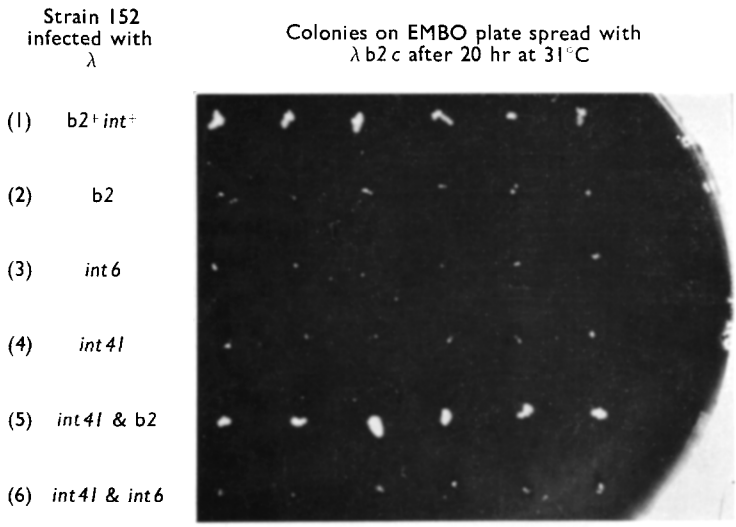


PLATE I. Spot test for the Int character.

Phages of each type were adsorbed to samples of strain 152 (10^8 viable cells/ml.) at a multiplicity of exposure of 10. After adsorption, a platinum wire was repeatedly dipped into the suspension of infected cells and stabbed into the EMBO plate, previously spread with 10^9 λb2*c*, to form a row. Incubation was as indicated. The relevant genotypes of the phages, which are all derivatives of λc*I857*, are also presented. Additional, and irrelevant, mutations present are *susA11* in λb2$^+$*int*$^+$, and *susA32 susJ27* in λb2.

The large spots are uniformly pink; the dots are surrounded by irregular darkened areas of cell lysis.

[*facing p. 492*

TABLE 1

*Non-reciprocal complementation between λb2 and 434hy*int41 *for lysogenization*

Infecting phage	Percentage of input cells having acquired immunity of		
	λ	434	λ and 434
λb2	0·01	—	—
434hy*int41*	—	<0·001	—
λb2 and 434hy*int41*	<0·02	5·4	3·0

Strain 152, at 5×10^8 cells/ml., was infected with the indicated phages at multiplicities of exposure of 4·4 for λ and 2·4 for 434hy. Dilutions of the infected cell suspensions were spread on EMBO plates previously seeded with 10^9 λb2*c* and EMBO plates seeded with 434hy*int6c*. After 24-hr incubation at 31°C, pink colonies were picked from these plates and tested for both immunities. Additional, and irrelevant, mutations present are *cI857susA32susJ27* in λb2 and *cI1susP3* in 434hy*int41*.

one phage has integrated (identified here by its immunity), invariably the phage is 434hy*int*. This non-reciprocality is interpreted to mean that λb2 lacks a structural element, acting *cis*, which is involved in integration (cf. Campbell, 1965), but that this mutant can provide in *trans* the function missing in an *int* mutant.

When either a *rec*$^+$ or *rec* host is used, rare colonies, stably lysogenic for λb2 alone, can be isolated. At even lower frequencies, *rec*$^+$ lysogens of λ*int*, obtained without exposure to *int*$^+$ phage, may be recovered. The location of the prophage is normal in the b2 lysogens, abnormal in the *int* lysogens. The properties of these λb2 and λ*int* lysogens will be the subject of separate communications.

Evidence concerning the nature of the Int function has been obtained from a comparison of the effect of various *int*$^+$ phages in helping to form stable lysogens of *int* phages (Table 2). The genomes of most temperate phages are attached to the bacterial chromosome at a specific site (Jacob & Wollman, 1961). It is possible that the prophages of λ, 434 and 434hy are inserted at identical sites (Liedke-Kulke & Kaiser, 1967), between bacterial markers *gal* and *bio* (Rothman, 1965). The hybrid between 21 and λ used here (21hy5) (Liedke-Kulke & Kaiser, 1967) is inserted at the λ site. Prophages 21 and φ80 are located in the vicinity of the *trp* region (Jacob & Wollman, 1958; Matsushiro, 1961; Signer, Beckwith & Brenner, 1965). Of these λ-related phages, only those which are located at (or adjacent to) the λ site serve as helpers for λ*int* integration. These results imply that the Int function is site-specific. We propose that the *int*$^+$ locus of λ supplies an enzyme or some other diffusible gene product, the existence of which was inferred by Signer & Beckwith (1966), which promotes recombination between the bacterial λ attachment site and a locus within or adjacent to the b2$^+$ region of λ.

The preceding experiments were performed with a *rec* strain of *E. coli* to avoid complications which the bacterial Rec function might introduce. However, the presence of a *rec*$^+$ gene in the host does not raise the lysogenization frequency of an *int* mutant to a value above 10^{-3} per infected cell (Table 3). The host Rec function does not appear capable of promoting recombination between the genetic elements involved in normal phage integration.

TABLE 2

Helping effect of int$^+$ *phages in lysogenization by* λint6

No.	Bacterial host	Helper phage	Lambda lysogens per 100 infected cells
1	*rec*	none	<0·05
2	*rec*	φ80	<0·05
3	*rec*	21gp	<0·05
4	*rec*	434	5
5	*rec*	434hy	23
6	*rec*	21hy5	21
7	*rec*(21gp)	none	<0·05
8	*rec*(21gp)	21hy5	<0·05
9	*rec*(21gp)	434hy	0·80†

Strain 152 was infected with *λint6* at a multiplicity of exposure of 15 and simultaneously with the indicated helper phage at the following multiplicities of exposure: φ80, 24; 21gp, 15; 434, 20; 434hy, 16; 21hy5, 15. Similarly, the lysogenic derivative of strain 152 was simultaneously infected with *λint6* at a multiplicity of exposure of 12, and helper phages 21hy5 and 434hy at multiplicities of 10 and 18, respectively. The starved cultures were infected at a final cell density of at least 2×10^8/ml. After adsorption, the infected cell suspensions were diluted appropriately and spread on plates of EMBO–*λint6c* for determination of λ lysogens. In the experiments of lines 4, 5 and 6, the λ lysogens obtained were tested for whether they were also lysogenic for the helper phage. The proportion of lysogens harboring λ alone was 0·83, 0·74 and 0·85, respectively.

† The helping effect of 434hy is considerably reduced by the presence of 21 prophage. However, the reduction in helping effect is not complete and may be attributed to some partial exclusion of 434hy by 21. A comparable reduction by 21 prophage in the efficiency with which 434hy alone lysogenizes (from 64 to 2%) was observed in this same experiment.

TABLE 3

Helping effect of bacterial Rec function in lysogenization by a superinfecting hetero-immune int *phage*

No.	Genotype		Lysogens with 434 immunity per 100 cells superinfected with	
	host	λ prophage	434hy*int6*	434hy
1	*rec*	none	<0·1	41
2	*rec*$^+$	none	<0·1	92
3	*rec*	*int6*	<0·1	32
4	*rec*	*int*$^+$	<0·1	55
5	*rec*$^+$	*int6*	1·4	81
6	*rec*$^+$	*int*$^+$	1·4	67

The *rec*$^+$ strain 28, the *rec* derivative (strain 152), and the indicated λcI*857* lysogens of these strains, at a cell density no less than $1{\cdot}2 \times 10^8$/ml., were infected with 434hy*int6* or with 434hy at the following respective multiplicities of exposure: (1) 4, 7; (2) 4, 6; (3) 3, 2; (4) 5, 9; (5) 8, 5; (6) 5, 8. After adsorption, appropriate dilutions of the infected cell suspensions were spread on EMBO–434hy*int6c* and incubated at 33°C. Cells were also plated at 39°C to determine if prophage substitution rather than double lysogenization had occurred. It was found that of *rec*$^+$ cells which acquire 434 immunity by superinfection with 434hy*int6*, less than 10% lost the immunity of the prophage.

(c) *Control of expression of the Int function*

Evidence is also presented in Table 2 that the Int function, like other phage functions, is under control of the immunity repressor. Whereas 21hy5 is an effective helper for integration of *λint* into a non-lysogen, the helping effect of 21hy5 is entirely eliminated when the host is lysogenic for 21.

A comparison of lines 4 and 5 of Plate I also demonstrates that the Int function is under immunity control. When cells are infected with *λint41* first, and immunity is established, the Int function of the superinfecting λb2c on the plate is not expressed, and λ lysogens are not produced (line 4). When infection is simultaneous (line 5), complementation occurs because the helping phage is not repressed initially.

Early defective mutants of lambda bearing amber (*sus*) mutations in the N cistron are deficient in a variety of lambda functions for which the N product may act as an inducer (Protass & Korn, 1966; Thomas, 1966; Dambly, Couturier & Thomas, 1968). These mutants are capable of efficient lysogenization at a high multiplicity of infection (Brooks, 1965). To test the possibility that this lysogenization might derive from "leakiness" of the N mutation, accentuated at high multiplicity, or might be due to the action of the bacterial recombination system, a phage bearing two mutations in the N cistron was assayed for lysogenization of a *rec* host. Twenty-five per cent of the infected cells of the *rec*, non-permissive strain 152 were lysogenized by 434hy*susN7,53* at a multiplicity of infection of five. Similar results were obtained with *λsusN7,53*. To eliminate possible phenotypic suppression by high concentrations of magnesium ions (Dahl, 1967), the standard lysogenization assay was altered. Cells were not starved in $MgSO_4$, and the concentration of Mg^{2+} throughout was reduced to $< 10^{-4}$ M. No significant difference in the lysogenization frequency of *λsusN7,53* was observed under these conditions. These results indicate that the *int*$^+$ gene is expressed under conditions where N cistron function is eliminated.

(d) *Lysogenization by superinfecting* int *phages*

Unlike the formation of *int* single lysogens, the formation of *int* polylysogens by superinfection is affected appreciably by the bacterial Rec function (Table 3). The lysogenization frequency of an *int* superinfecting phage is increased more than tenfold by the presence of hybrid prophage of differing immunity. This increase is entirely dependent upon the presence of a *rec*$^+$ gene in the host. It is independent of the prophage *int* allele. Thus, the bacterial Rec function, while not promoting single lysogenization, appears to permit recombination between a prophage and a superinfecting hetero-immune phage genome, as illustrated in Fig. 1.

(e) *Recombination among vegetative* int *phages*

Zissler (1967) has reported that *int* phages recombine in a *rec* host. The results of Table 4 confirm Zissler's findings. The extent of recombination between two vegetative phages in a *rec* host does not appear to be affected by the absence of the Int function (Table 4, line 4) when distant *sus* markers spanning the b2$^+$ region are used.

(f) *Induction of* int *lysogens*

Lysogens bearing only *int* prophages may be constructed by using *int*$^+$ helper phage (see Materials and Methods). Two distinct classes of lysogens are observed (Table 5). The most frequently obtained *int6* lysogen is one which, upon induction, yields

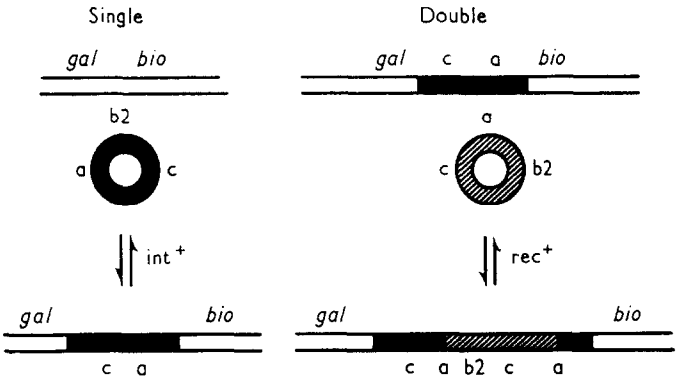


FIG. 1. Lysogenization.

TABLE 4

Recombination between int sus *mutants in a* rec *host*

Cross	*sus*+ recombinants per 100 progeny phage
λ*susA* × 434hy*susP*	14·0
λ*susAint21* × 434hy*susP*	14·6
λ*susA* × 434hy*susPint41*	12·6
λ*susAint21* × 434hy*susPint41*	16·5

The recombination-deficient non-permissive strain 152 was infected at 2×10^8 cells/ml. with phage at a multiplicity of exposure of 2·4 for each phage. After adsorption, the cultures were aerated at the inducing temperature of 39°C for 2 hr. The lysates were then chloroformed and the phage titrated on the non-permissive host, strain 28, and the permissive host, C600. The markers whose recombination is being assayed are *susA11* and *susP3*.

about 4×10^{-3} infective centers per bacterium and the same proportion of free phage, i.e. a burst size of about one. Although the phage yield is subnormal, heat-induction of these lysogens results in death and lysis of every cell.

The inability of certain *int* lysogens to produce normal phage yields is believed to be due to a defect in prophage excision. This interpretation is consistent with the finding that the lysates of these induced lysogens contain large numbers of transducing particles (Table 5). It is also supported by the observed inability of *int* mutants to promote superinfection curing (see below), which is known to involve prophage excision (Ptashne, 1965). It appears therefore that the same defect which impairs phage integration impairs prophage detachment; both processes rely at some stage on the same function.

Induction of a λ lysogen, but not infection of a sensitive cell, leads to derepression of the adjacent *gal* operon. This phenomenon, discovered by Buttin and Yarmolinsky & Wiesmeyer (see Buttin, 1963*c*; Yarmolinsky, 1963), is referred to as "escape

synthesis". Since escape synthesis is found only on induction of a prophage, we expected it might be influenced by a defect in prophage excision. Accordingly *λint* lysogens were tested for their ability to exhibit escape synthesis on lysogenic induction. The experiments of Fig. 2, performed by Dr Elke Jordan, demonstrate that a *λint* single lysogen shows an even greater escape synthesis than the *λint*$^+$ control. On the other hand, cells doubly lysogenic for *λint* (described below), in which active phage yields are normal, show no more escape synthesis than the *int*$^+$ control (data not shown). Whatever the cause of escape synthesis, the Int function is clearly dispensable.

TABLE 5

Phage yields of λcI857 lysogens thermally induced

Genotype		Infectious centers per input cell	Active phage per infectious center	Yield per input cell ($\times 10^4$) of phage transducing	
λ prophage	Host			*gal*	*bio*
int$^+$	*rec*$^+$	1·0	130	0·22	—
int$^+$	*rec*	1·0	71	0·15	0·24
int6	*rec*$^+$ (3/3)	0·004	*ca.* 1	1·7	—
int6	*rec* (13/17)	0·004	*ca.* 1	9	0·68
	rec (4/17)	1·0	60	3·3	—

The indicated lysogens of strains 28 (*rec*$^+$) and 152 (*rec*) were grown at 33°C in TB to 10^8 cells/ml., washed by centrifugation, and portions were plated at 33°C for determination of viable counts. Infectious centers were determined by plating at 39°C with strain 28 as indicator bacteria. For determination of phage yield, lysogens were grown at the inducing temperature of 39°C for 2 hr, chloroformed and plated with strain 28 at 33°C. Transducing particle titration is described in Materials and Methods.

Int lysogens were prepared by mixed infection with *int*$^+$ hetero-immune helper. The lambda lysogens so obtained were selected for the absence of helper immunity and *int*$^+$ phage in the burst. Results typical of individual clones of each class are reported. The figures in parentheses represent clonal frequencies.

The presence or absence of the host Rec function does not affect the phage yield. This is consistent with the finding, reported above, that the bacterial Rec function does not permit, unless very rarely, single lysogenization by *int* mutants.

In addition to the low-yielding clones of *int* lysogens, clones are obtained in which every cell can yield an infective center and the average burst size is nearly normal. We propose that these normal yielders bypass the impaired Int function as a consequence of homogenotic polylysogeny. A tandem-double lysogen, in principle, can detach one complete phage genome by a recombinational event anywhere along the length of the two prophages (see Fig. 1). The presence of the prophage genome in duplicate provides large areas of homology. Crossing over in these areas, resulting in the excision of one prophage, could be promoted by any generalized recombination system. By this means, the absence of the site-specific Int function could be circumvented. In the case of *int* lysogens of a *rec* host which are normal yielders, the recombination system involved in prophage excision might be the same non-specific

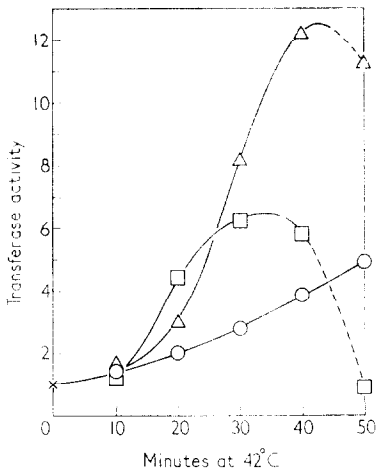


FIG. 2. Effect of *int* mutation on *gal* escape synthesis.

Derivatives of a *gal*$^+$ *rec* revertant of the *gal rec* strain 152 were grown at 34°C in TB and shifted to 43°C to induce the lysogens. The strains carry: no λ prophage (○); λc*I857int*$^+$ (□), λc*I857int6* low yielder (△). The activity of UDP glucose: α-D-galactose-1-phosphate uridylyl transferase per ml. of culture is normalized to unity at the time of the temperature change. Cell lysis is occurring where the lines are dashed. Results shown are the average values from two experiments performed by Dr Elke Jordan.

function (Red) which allows recombination between vegetative phage in a *rec* bacterium.

The hypothesis that homogenotic *int* polylysogens would give nearly normal phage yields upon induction was tested by constructing double lysogens, heterogenotic for immunity, in which the two prophages were located adjacent to, or separate from, each other. The results presented in Table 6 show that when the prophages are adjacent, every bacterium on thermal induction yields a burst of λ. When each prophage enjoys a separate location, e.g. one chromosomal, the other on an F′ episome, the yield of λ is low, generally no higher than from a single *int* lysogen.

The immunity marker of the second prophage is also recovered in the lysate (Table 6). The yield of 434hy*int6* is seen to be high in heat-induced lysates of N178 and low in lysates of the isogenic strain N201. This difference may reflect the sequence of the two prophages, and can be explained if we assume the order in N178 to be *gal*-λ-434hy-*bio* and in N201 to be *gal*-434hy-λ-*bio*. Random recombination between the two prophages would result in preferential excision of the centrally located immunity region of the prophage to the right.

The high yield of 434hy in the burst of N178 indicates that replication of this phage must occur. This result is unexpected, since an excised prophage should not replicate unless induced, despite multiplication of hetero-immune phage in the same cell (Thomas & Bertani, 1964). The mechanism of cross-induction of 434hy by induction of λ is under investigation.

(g) *Induction of superinfected* int *single lysogens*

We have seen earlier that an appropriate superinfecting *int*$^+$ phage allows integration of an *int* phage. Superinfecting phage can also aid in the excision of an induced *int* prophage. The data of Table 7 show that the yield of λ from an induced λ lysogen is reduced slightly by superinfection with the competing phage 434hy. However,

TABLE 6

Helping effect of adjacent prophage in thermal induction of λcI857int6 *heterogenotic* rec *polylysogens*

Strain	Second prophage		Total infectious centers per input cell	Lambda *cI857* per infectious center	Second phage per infectious center
	Name	Location			
N111	None	—	0·004	1·1	—
N158	21hy5	λ near *gal*	1	72	0·7
N159	21gp	21 near *trp*	0·004	1·1	<0·01
N157	434	434 near *gal*	1	25	2·5
N122	434hy	λ near *gal*	1	29	5
N202	434hy	on F'_1 *gal*$^+$	1	0·07	3
N178	434hy*int6*	λ near *gal*	1	7·7	86
N201	434hy*int6*	λ near *gal*	1	90	0·5
N237	434hy*int6*	on F'_1 *gal*$^+$	0·006	0·4†	0·02

Lysogens were treated as described in the legend of the previous Table. Phage immunities were distinguished by plaque type at 39°C, i.e. clear for lambda, turbid for all other phages used.

The second prophage was introduced into the *rec* strain 152 lysogenized with λcI*857int6* previously (most strains), simultaneously (N122), or in the reverse order (N201). Preparation and identification of *int* double lysogens and of F′ heterogenotes are described under Materials and Methods.

† The observation that λ per infectious center is less than one is due to some replication of the lysogens prior to lysis in the infectious center assay. This leads to an overestimate of the number of infectious centers per input cell.

when the prophage is λ*int*, the same superinfecting phage increases the λ yield about four orders of magnitude. A comparable increase in λ yield is accomplished by superinfection with 21hy5. The yield of λ is also increased, although not to the same extent, when the superinfecting phage does not supply the appropriate *int*$^+$ function (e.g. 21gp, 434hy*int6*). This is probably a consequence of recombination between the two phages, providing an alternative mechanism by which superinfection can increase the yield of phage carrying prophage immunity. As seen in Table 7, this effect of marker rescue can account for only a small fraction of the helping by superinfecting *int*$^+$ phage.

(h) *Superinfection curing and prophage substitution*

Superinfection of a lysogen with a related hetero-immune phage can result in loss of the prophage immunity from the cells surviving superinfection. Some of these survivors have acquired the immunity of the superinfecting phage (prophage substitution), whereas others have not (superinfection curing) (Cohen, 1959; Six, 1960, 1961; Eisen, Siminovitch & Mohide, 1967; Liedke-Kulke & Kaiser, 1967). Homoimmune superinfection does not result in curing, suggesting that a repressible phage function is required for this process.

We have exploited the heat sensitivity of thermo-inducible lysogens to measure prophage loss after hetero-immune superinfection. Cells which have lost their thermo-inducible prophage will form colonies at 39°C. These colonies may then be tested for acquisition of the immunity of the superinfecting phage.

As seen in Table 8, prophage loss can be accomplished efficiently by superinfection with an *int*$^+$ phage having the same locus of attachment as λ. The curing function has the same site-specificity as the Int function. In these experiments, *rec* lysogens were used in order to minimize the frequency of recombinations between prophage and superinfecting phage resulting in the exchange of immunity regions.

TABLE 7

Yield of lambda from induced lambda single lysogens following hetero-immune superinfection

Induced lysogen	Superinfecting phage	Lambda per induced cell
rec(λ)	none	61
rec(λ)	434hy	17
rec(λ*int6*)	none	0·0002
rec(λ*int6*)	21gp	0·02
rec(λ*int6*)	434hy*int6*	0·1
rec(λ*int6*)	434	2·8
rec(λ*int6*)	434hy	5·7
rec(λ*int6*)	21hy5	21

Lysogens of strain 152, at no less than 3×10^8 cells/ml., were incubated with phage at the following multiplicities of exposure: 434hy, 5; 21gp, 2·4; 434hy*int6*, 2·7; 434, 9; 21hy5, 6. The infected cell suspensions were diluted in TB and shaken at 39°C for 2 hr. The lysates were chloroformed and the lambda titer determined with strain 152 as indicator at 39°C.

TABLE 8

Correlation of superinfection curing specificity with site of chromosomal attachment

Superinfecting phage		Prophage loss: survivors at 39°C per 100 survivors at 33°C
Name	Location	
none	—	<0·001
φ80	φ80 near *trp*	<0·001
21gp	21 near *trp*	<0·001
434	434 near *gal*	18
434hy	λ near *gal*	82
21hy5	λ near *gal*	70
434hy*susN7,53*	λ near *gal*	1·0

Strain 152 lysogenic for λ*cI857susA11int21* was infected at a density of 7×10^8 cells/ml. with phage at the following multiplicities of exposure: φ80, 8; 21gp, 2; 434, 30; 434hy, 7; 21hy5, 10; 434hy*susN7,53*, 2·5. The high multiplicity of exposure of 434 was used to obtain a multiplicity of infection of 3, as adsorption of this phage is poor. After adsorption, the infected cell suspensions were diluted at least 2×10^{-2} in TB, samples were plated for determination of the proportion of cells surviving infection and the remainder was grown 2 to 3 hr at 33°C to allow segregation of nuclei no longer lysogenic for λ. Samples were again plated for survivors at 33°C and the remainder incubated 15 min at 41°C to induce and kill λ lysogens, and then plated for survivors at 39°C. From 23 to 77% of the cells survived infection, except for φ80-infected cells, of which 15% survived.

TABLE 9

Roles of Int function in superinfection curing and of bacterial Rec function in prophage substitution

No.	Genotype: Host	Genotype: Prophage λcI857	Super-infecting phage†	Survivors at 39°C per 100 survivors at 33°C following superinfection with: none	434hy *int*	434hy *int*$^+$	Figures measure
1	*rec*	*susA11int21*	cI$^+$	<0·001	0·006	21	Prophage loss = curing and substitution
2	*rec*	*susA11int*$^+$	cI$^+$	0·009	0·041	24	
3	*rec*$^+$	*susA11int*$^+$	cI$^+$	0·042	0·38	28	
4	*rec*$^+$	*sus*$^+$*int6*	cI$^+$	<0·001	0·12	92	
5	*rec*$^+$	*sus*$^+$*int6*	cI*1*	<0·001	0·003	28	Curing alone

Lysogens of strain 28 or its *rec* derivative, strain 152, at cell densities of at least 3×10^7/ml., were superinfected with the indicated phages at multiplicities of exposure of 5 to 9. Superinfecting 434hy*int* phages were *int6* in experiments 1 through 4, and *int41susP3* in experiment 5. Following adsorption, the cell suspensions were diluted in TB and segregation of non-lysogenic nuclei was allowed to take place during incubation for 120 min (experiments 1 through 3), or on pour-plates incubated for 140 min (experiments 4 and 5). Survivors at 33°C were determined by sampling the broth cultures and plating at 33°C or, in experiments 4 and 5, by continuing incubation at 33°C of certain of the previously poured plates. Lysogens carrying thermo-inducible prophages were then killed by plating at 39°C directly (experiment 1), or after a preliminary incubation at 41°C for 15 min (experiments 2 and 3), or by transferring the remaining pre-incubated plates to a 39°C incubator (experiments 4 and 5). Surviving colony formers were scored after overnight incubation at 39°C.

† cI$^+$ phage are not thermo-inducible; cI*1* phage are induced at temperatures above 36°C.

The early defective 434hy mutant *susN7,53* unable to synthesize phage DNA (Brooks, 1965) was assayed for ability to cause prophage loss. The defect in N reduces the extent of prophage loss about 40-fold. Similar results have been obtained by Dr Naomi Franklin (personal communication).

Table 9 shows the different roles of the Int and the Rec functions in prophage loss. When superinfection is by 434hy*int* in place of 434hy*int*$^+$, the number of cells which lose their thermo-inducible prophage is reduced almost two orders of magnitude (line 3). The residual prophage loss seen after superinfection with *int* mutants can virtually be eliminated in two ways: (1) by the use of a *rec* lysogen (lines 1 and 2), or (2) by making the superinfecting phage thermo-inducible as well (lines 4 and 5). When the superinfecting phage is also thermo-inducible, exchange of immunity markers with a thermo-inducible prophage will not result in the formation of a heat-stable lysogen. Lysogens in which an exchange of immunity markers has occurred, as well as those in which it has not, will both be induced at 39°C; only cured cells, bearing neither phage, will survive plating at this temperature. The results of Table 9 indicate that, of the two modes of prophage loss, only prophage substitution, utilizing the bacterial Rec function, occurs in the absence of the phage Int function.

4. Discussion

Three recombinational systems participate in genetic exchanges involving λ bacteriophage. The Rec function is bacterial; the other two functions, Int and Red, are phage-determined. The Int function (defective in *int* mutants) appears specialized for normal integration and detachment of the prophage genome.

We conclude that the *rec*$^+$ gene product participates in recombination between an *int* superinfecting phage and hetero-immune prophage, since stable lysogenization by the superinfecting phage requires a *rec*$^+$ allele. Presumably the recombination found to occur between prophages (Calef & Licciardello, 1960) is also catalyzed by enzymes of the bacterial host.

The existence of a second non-specific recombination system (Red), operative in phage-infected cells, is inferred from the finding that *int* mutants recombine efficiently during vegetative growth in a *rec* host. This same system functions also in recombination between vegetative phage and prophage (marker rescue) and in recombination between tandem prophages (induction from a double lysogen); both processes occur in the absence of either *int*$^+$ or *rec*$^+$ alleles. Cell survival is not required in any of these situations. Processes such as lysogenization of a hetero-immune lysogen or prophage substitution, involving similar recombinational events but requiring also that the infected cell survive, are not catalyzed by the Red function. Thus, the Red function is not efficiently expressed in the lysogenic pathway.

The erratic prophage excision which occurs following induction of an *int* single lysogen is probably due to expression of the Red function. The appearance of large numbers of transducing particles in the lysates of these induced lysogens and the low yield of viable phage can be accounted for by generalized recombination in the region of the induced prophage. Large areas of homology would not be involved and excision of a complete phage genome would be a rare event.

The phage-determined site-specific Int function participates in integration, prophage detachment upon induction, and superinfection curing. All of these functions are impaired by *int* mutations. In superinfection curing, the *int*$^+$ product of the incoming phage detaches the prophage in lieu of integrating the infecting phage. In subsequent cell divisions, both phages, each repressed, are diluted out. This model is supported by our observations and those of Signer & Beckwith (1966), and of Signer (cited in Signer & Beckwith, 1966), which indicate that the curing specificity of the superinfecting phage is identical to its attachment specificity. Further support is presented in a note added in proof to the article by Liedke-Kulke & Kaiser (1967).

The chromosomal location of phage 434 may be identical with the location of λ (Liedke-Kulke & Kaiser, 1967). The complementation which occurs between 434*int*$^+$ and λ*int*, and the ability of superinfecting 434*int*$^+$ to cure λ lysogens, is consistent with identity of the chromosomal attachment sites of 434 and λ.

The Int function is repressible. This is in agreement with the observation that integration of an *int*$^+$ superinfecting phage is known to be blocked by co-immune prophage (Dahl & Calef, 1966), even at an unoccupied site (Campbell & Zissler, 1966; Taylor & Yanofsky, 1966) or simply by repressor remaining in a non-lysogenic segregant of an abortive lysogen (Ogawa & Tomizawa, 1967). For normal integration to take place, therefore, the *int*$^+$ gene must be expressed prior to the expression of immunity, although integration itself may be delayed for as long as several generations (Smith & Levine, 1967).

There is no clear indication that Int function is under control of the N cistron. N-defective mutants efficiently lysogenize a *rec* host. The low efficiency of curing in the absence of N function, either upon transient induction as described by Eisen *et al.* (1966), or upon hetero-immune superinfection as described here, need not imply that Int is under N control. Instead it may reflect an additional N-dependent function in curing, e.g. host chromosome repair. Various functions affected by the N defect are expressed following hetero-immune superinfection with an N^+ phage. Thomas (1966) and Dambly *et al.* (1968) have proposed that these functions are under the positive control, *in trans*, of the N cistron product. Since integration of an *int* superinfecting phage is not complemented by an *int*$^+$ prophage, *trans* induction of the Int function does not appear to occur.

The function of the *int*$^+$ gene product is to promote, with high efficiency, integration of phage DNA within a specific region of bacterial DNA. It is generally assumed that this integration requires pairing of homologous regions. The fraction of the λ genome involved in such pairing should correspond to the frequency of detachment of λ prophage from a single lysogen relative to the frequency of detachment of a prophage from a tandem double lysogen, provided the recombination systems involved in each case do not exhibit site-specificity. This requirement is perhaps met in the case of induction of *int* phage from *rec* lysogens, where prophage excision involves only the Red function. Under inducing conditions, the ratio obtained is 4×10^{-3}/l. This ratio represents an upper limit for the relative size of the pairing region, first, because the denominator cannot exceed unity (no more than all induced lysogens can yield phage); second, because the numerator includes excisions which may lie outside the normal pairing regions yet result in the detachment of a viable phage. Calef, Marchelli & Guerrini (1965) have calculated a ratio of 10^{-2} using *int*$^+$ prophage. This value is probably overestimated, due to the bias introduced by the occasional derepression of the site-specific Int function. The Int product, appearing as a result of temporary derepression, may be largely responsible for the spontaneous curing of single lysogens; single *int* lysogens are very stable. Our calculations, as well as those of Calef *et al.* (1965), are based on the assumption that phage and bacterial copies of the attachment site are homologous.

Integration of λ by means of generalized recombination would require broad regions of homology and would result in unstable integration. The Int function, which allows specific recombination, obviates this requirement. Thus, the Int function, when expressed, permits efficient integration and excision. When repressed (in the lysogenic state), the absence of Int function permits prophage to be stable.

A number of workers, cited in the text, have generously furnished bacteria and phage. We thank Drs William Dove, Harrison Echols, Harvey Eisen, Naomi Franklin, Lià Fischer-Fantuzzi, Jonathan Gallant, A. D. Kaiser, Ethan Signer, René Thomas, Robert Weisberg and James Zissler for making their manuscripts available to us prior to publication and for communicating unpublished data. We are indebted to Dr J. L. Rosner for helpful discussions during the course of this work, and to Drs Martin Gellert and John Little for their critical reading of the manuscript. Dr Elke Jordan has kindly permitted us to publish her data on escape synthesis.

REFERENCES

Adelberg, E. A., Mandel, M. & Chen, G. C. C. (1965). *Biochem. Biophys. Res. Comm.* **18**, 788.
Appleyard, R. K. (1954). *Genetics*, **38**, 440.
Brooks, K. (1965). *Virology*, **26**, 489.

Brooks, K. & Clark, A. J. (1967). *J. Virology*, **1**, 283.
Buttin, G. (1963*a*). *J. Mol. Biol.* **7**, 164.
Buttin, G. (1963*b*). *J. Mol. Biol.* **7**, 183.
Buttin, G. (1963*c*). *J. Mol. Biol.* **7**, 610.
Calef, E. & Licciardello, G. (1960). *Virology*, **12**, 81.
Calef, E., Marchelli, C. & Guerrini, F. (1965). *Virology*, **27**, 1.
Campbell, A. (1957). *Virology*, **4**, 366.
Campbell, A. (1961). *Virology*, **14**, 22.
Campbell, A. (1962). *Advanc. Genetics*, **11**, 101.
Campbell, A. (1963). *Virology*, **20**, 344.
Campbell, A. (1965). *Virology*, **27**, 340.
Campbell, A. & Zissler, J. (1966). *Virology*, **28**, 659.
Clark, A. J. & Margulies, A. D. (1965). *Proc. Nat. Acad. Sci., Wash.* **53**, 451.
Cohen, D. (1959). *Virology*, **7**, 112.
Dahl, D. (1967). *Atti. Assoc. Genet. Ital.* **12**, 496.
Dahl, D. & Calef, E. (1966). *Virology*, **30**, 572.
Dambly, C., Couturier, M. & Thomas, R. (1968). *J. Mol. Biol.*, in the press.
Demerec, M., Adelberg, E. A., Clark, A. J. & Hartman, P. E. (1966). *Genetics*, **54**, 61.
Dove, W. F. (1967). In *The Molecular Biology of Viruses*, ed. by J. S. Colter & W. Paranchych, p. 111. New York: Academic Press.
Eisen, H. A., Fuerst, C. R., Siminovitch, L., Thomas, R., Lambert, L., Pereira da Silva, L. & Jacob, F. (1966). *Virology*, **30**, 224.
Eisen, H. A., Siminovitch, L. & Mohide, P. T. (1967). *Virology*, in the press.
Fischer-Fantuzzi, L. (1967). *Virology*, **32**, 18.
Franklin, N. C. (1967). *Genetics*, **57**, 301.
Franklin, N. C., Dove, W. F. & Yanofsky, C. (1965). *Biochem. Biophys. Res. Comm.* **18**, 910.
Gingery, R. & Echols, H. (1967). *Proc. Nat. Acad. Sci., Wash.* **58**, 1507.
Gratia, J. P. (1966). *Biken J.* **9**, 77.
Hertman, I. & Luria, S. E. (1967). *J. Mol. Biol.* **23**, 117.
Jacob, F. & Wollman, E. L. (1956). *Ann. Inst. Pasteur*, **91**, 486.
Jacob, F. & Wollman, E. L. (1958). *Ann. Inst. Pasteur*, **95**, 497.
Jacob, F. & Wollman, E. L. (1961). *Sexuality and the Genetics of Bacteria*. New York: Academic Press.
Kaiser, A. D. (1957). *Virology*, **3**, 42.
Kaiser, A. D. & Jacob, F. (1957). *Virology*, **4**, 509.
Kellenberger, G., Zichichi, M. L. & Weigle, J. (1961*a*). *Proc. Nat. Acad. Sci., Wash.* **47**, 869.
Kellenberger, G., Zichichi, M. L. & Weigle, J. (1961*b*). *J. Mol. Biol.* **3**, 399.
Lederberg, E. M. (1960). *Microbial Genetics*, 10th Symp. Soc. Gen. Microbiol. p. 115. Cambridge: University Press.
Lederberg, E. M. & Lederberg, J. (1953). *Genetics*, **38**, 51.
Liedke-Kulke, M. & Kaiser, A. D. (1967). *Virology*, **32**, 465.
Luria, S. E., Adams, J. N. & Ting, R. C. (1960). *Virology*, **12**, 348.
Matsushiro, A. (1961). *Biken's J.* **4**, 139.
Matsushiro, A. (1963). *Virology*, **19**, 475.
Meselson, M. & Weigle, J. (1961). *Proc. Nat. Acad. Sci., Wash.* **47**, 857.
Monod, J., Cohen-Bazire, G. & Cohn, M. (1951). *Biochim. biophys. Acta*, **7**, 585.
Ogawa, T. & Tomizawa, J. I. (1967). *J. Mol. Biol.* **23**, 225.
Protass, J. J. & Korn, D. (1966). *Proc. Nat. Acad. Sci., Wash.* **55**, 1089.
Ptashne, M. (1965). *J. Mol. Biol.* **11**, 90.
Radding, C. M. & Kaiser, A. D. (1963). *J. Mol. Biol.* **7**, 225.
Rothman, J. (1965). *J. Mol. Biol.* **12**, 892.
Signer, E. R. (1966). *J. Mol. Biol.* **15**, 243.
Signer, E. R. & Beckwith, J. R. (1966). *J. Mol. Biol.* **22**, 33.
Signer, E. R., Beckwith, J. R. & Brenner, S. (1965). *J. Mol. Biol.* **14**, 153.
Six, E. (1960). *J. Bact.* **80**, 728.

Six, E. (1961). *Virology*, **14**, 220.
Smith, H. O. & Levine, M. (1967). *Virology*, **31**, 207.
Sussman, R. & Jacob, F. (1962). *C. R. Acad. Sci. Paris*, **254**, 1517.
Taylor, M. & Yanofsky, C. (1966). *Virology*, **29**, 502.
Thomas, R. (1966). *J. Mol. Biol.* **22**, 79.
Thomas, R. & Bertani, L. E. (1964). *Virology*, **24**, 241.
Weisberg, R. A. & Gallant, J. A. (1967). *J. Mol. Biol.* **25**, 537.
Wollman, E. L. (1963). *C. R. Acad. Sci. Paris*, **257**, 4225.
Yarmolinsky, M. (1963). In *Viruses, Nucleic Acids, and Cancer*, p. 151. Baltimore: The Williams and Wilkins Co.
Zichichi, M. L. & Kellenberger, G. (1963). *Virology*, **19**, 450.
Zissler, J. (1967). *Virology*, **31**, 189.

Note added in proof: The inference that Red function, in the absence of Int or Rec functions, permits occasional prophage excision from single lysogens and relatively efficient prophage excision from tandem double lysogens, has been tested. Strain 152 (*rec*) was lysogenized with λcI*857int6red3*, using 21hy5 helper. The *int6* and *red3* mutations eliminate Int and Red functions, respectively. The *red3* derivative of λcI*857int6* was kindly supplied by Drs J. Weil and E. Signer who isolated and characterized this mutation (manuscript in preparation). Low yielders and normal yielders of λ*int red* were obtained; the phage yields did not differ by more than 50% from *int red*$^+$ controls. Accordingly, barring complementation between *red3* and the *rec* mutation in strain 152, excision may require very little expression of the non-specific recombination systems or alternative mechanisms of phage release may be responsible for the excision observed.

Cell, Vol. 19, 845–854, April 1980, Copyright © 1980 by MIT

Phase Variation: Genetic Analysis of Switching Mutants

Michael Silverman and Melvin Simon
Department of Biology, B-022
University of California, San Diego
La Jolla, California 92093

Summary

Site-specific inversion of a controlling element is responsible for flagellar phase transition in Salmonella. When a 900 bp DNA sequence is in one configuration, it allows the expression of the *H2* gene, a structural gene which codes for the flagellar antigen. When it is in the opposite configuration, the *H2* gene is not expressed. A hybrid λ phage containing the invertible control region and the adjacent *H2* gene was constructed, and expression of the *H2* gene was shown to be regulated by the orientation of the inversion region. Transposon Tn5 insertion derivatives of this hybrid phage were isolated and λ*H2::Tn5* mutants defective for inversion (*H2* switching) were selected and characterized. Two classes of switching phenotypes were observed among the mutants—those which had slightly reduced frequencies of transition from expression of the *H2* gene (*H2* on) to nonexpression (*H2* off) (intermediate class) and those in which the frequency of transition was reduced at least three orders of magnitude (null class). Physical mapping of the Tn5 insertion sites revealed that in all mutants the insertion was located inside the inversion region. Tn5 insertion sites in the null class of mutants defined a region of DNA including approximately 500 bp which was necessary for inversion. Genetic complementation tests showed that these λ*H2::Tn5* mutants could invert the *H2* gene control element if the 500 bp region was introduced in the trans configuration. It is concluded that a gene is located inside the inversion segment and codes for a protein which is required for the inversion event. Furthermore, the two sites at which the crossover event occurred functioned in a cis configuration and were required for inversion. The presence of a gene which is involved in controlling site-specific recombination events may be a general feature of transposon-like elements.

Introduction

Regulation of gene expression by mechanisms involving site-specific recombination has been suggested by recent work with a variety of experimental systems. In eucaryotes, genetic experiments with the yeast mating-type system (Hicks, Strathern and Herskowitz, 1977; Kushner, Blair and Herskowitz, 1979) and molecular cloning experiments involving the immunoglobulin genes in mice (Sakano et al., 1979) support the idea that specific genetic rearrangements determine the nature of gene expression. In procaryotes, the inversion of the G loop region of bacteriophage Mu DNA has been correlated with the infectivity of the phage particle (Bukhari and Ambrosio, 1978; Kamp, et al., 1978). We have shown that the mechanism of phase variation in Salmonella involves a specific rearrangement of DNA structure (Zieg et al., 1977; Zieg, Hilmen and Simon, 1978; Silverman et al., 1979b).

In Salmonella, two genes code for the major flagellar structural protein, flagellin. These two genes, the *H1* and *H2* genes, map in different regions of the Salmonella genome (Lederberg and Edwards, 1953). The phenomenon of phase variation refers to the ability of the cell to alternate or switch between expression of the two flagellin structural genes. This variation of antigenicity presumably allows Salmonella to evade the host immune response. The frequency with which cells undergo phase transition varies with different Salmonella strains from 10^{-3} to 10^{-5} per bacterium per generation (Stocker, 1949). The alternative expression of the *H1* and *H2* genes is controlled by the state of a genetic element linked to the *H2* gene (Lederberg and Iino, 1956). Another gene, *rhl*, linked to and coordinately expressed with *H2*, codes for a repressor substance that prevents expression of the *H1* gene (Fujita, Yamaguchi and Iino, 1973; Silverman, Zieg and Simon, 1979a). Thus when a cell is expressing the *H2* gene it also expresses *rhl*. This results in the repression of the *H1* gene, and only *H2*-type flagella are formed. When a cell is in phase 1, neither the *H2* nor *rhl* gene products are synthesized, and the *H1* gene can be expressed, leading to the formation of *H1*-type flagella.

To understand the mechanism of phase variation at the molecular level, recombinant molecules which carried the *H1* and *H2* gene regions were constructed and cloned in E. coli, where the phase variation effect could be reproduced. Genetic and physical analysis of the recombinant DNA molecules showed that inversion of a 900 bp region adjacent to the *H2* gene controlled the expression of this gene: in one orientation the *H2* operon was "on" and in the opposite orientation the *H2* operon was "off" (Zieg et al., 1977). The inversion of this control region which contains the *H2* operon promoter was found to be site-specific and independent of the RecA recombination system of E. coli (Zieg et al., 1978; Silverman et al., 1979b). These observations are summarized in Figure 1.

Inversion of a controlling element explains the oscillatory nature of *H2* operon expression, but to describe the mechanism which controls the frequency of the phase variation phenomenon precisely we have attempted to define by genetic techniques functions which are necessary for the inversion process. Two genetic alterations which affect the frequency of tran-

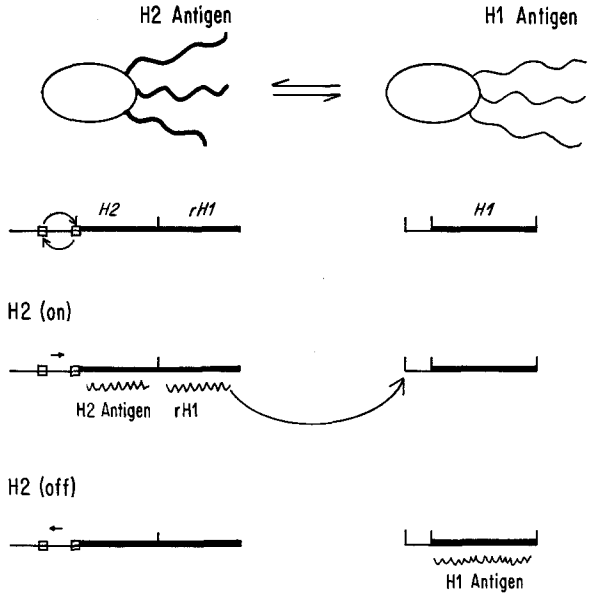


Figure 1. Model for the Alternation of Expression of the *H1* and *H2* Genes in Salmonella

Expression of the flagellar serotypes is regulated by the orientation of an invertible DNA sequence adjacent to the *H2* operon. When the *H2* operon is transcribed, the *H2* and *rhl* gene products are synthesized and the *rhl* (repressor of *H1*) gene produce prevents expression of the *H1* gene (H2 phase). When the *H2* operon is not transcribed, the *H1* gene product is formed (*H1* phase).

sition are known. One is a variant termed vH2$^-$ which was found in natural populations of Salmonella (Iino, 1961). It restricts phase transitions and was shown to map adjacent to the *H2* gene. The other is a deletion which removes about 50% of the DNA sequences on one side of the invertible region and fixes the *H2* gene in the *H2* (on) state (Silverman et al., 1979b). To further define the functions involved, a variety of mutants are required. To this end, a hybrid λ phage was constructed which contained the *H2* gene with its invertible control region. Convenient techniques were devised to measure switching of the state of expression of the *H2* gene on this hybrid λ, and transposon Tn5 was used to introduce insertion mutations into the hybrid. This report describes the isolation of mutants defective in the phase transition process. These mutants define a region of DNA inside the inversion region which codes for a gene whose product is necessary for the inversion process.

Results

E. coli is monophasic and has only one flagellin-specifying gene (*hag*), which is analogous to the *H1* gene of Salmonella. A hybrid λ phage (λ*fla 157*) was constructed by inserting a 3.75 bp Eco RI endonuclease restriction fragment that carried the *H2* gene derived from Salmonella onto a λ cloning vehicle. A Hag$^-$ E. coli strain lysogenized with the hybrid λ phage alternates between the nonflagellate and the flagellate (H2 serotype) phenotype. Cells with these two phenotypes could be conveniently distinguished by their susceptibility to the flagellotropic phage χ (Silverman et al., 1979b). Cells lysogenized with *H2* in the "on" configuration were sensitive to this phage, while lysogens with *H2* in the "off" configuration were resistant. The proportion of cells in a population in either state could be measured as a function of the number of generations of growth, and thus the frequency of phase transition could be determined (see Experimental Procedures).

To isolate λ*H2* derivatives which were mutagenized by transposon Tn5 insertion (λ*H2::Tn5*), the hybrid phage was grown in an E. coli strain with the Tn5 transposon inserted in the *ilv* gene (see Berg, 1977; Kleckner, Roth and Botstein, 1977). The resulting population of phage was then used to lysogenize a Hag$^-$ E. coli strain, and selection was applied for phage-mediated transduction of the kanamycin determinant residing on the Tn5 transposon (see Figure 2). In this way, a large number of lysogens which contained λ*H2::Tn5* insertions were collected. As a qualitative method to screen λ*H2::Tn5* phage with switching defects, clones of these lysogens were inoculated onto motility agar plates containing the flagellotropic phage χ. A lysogen clone in the *H2* (off) configuration or a lysogen in the *H2* (on) configuration which was capable of transition to the *H2* (off) phase would have been resistant to the flagellotropic phage. Only those lysogens harboring λ*H2::Tn5* which contained the *H2* gene in the "on" configuration and had a marked decrease in the frequency of transition to the "off" phase would be sensitive to the flagellotropic phage (see Figure 2). Using this screening method, putative switching mutants were chosen for further analysis from among 3000 λ*H2::Tn5* lysogens. The frequency of phase transition of these candidates was then measured. The *H2* (on) to *H2* (off) frequency for λ*H2::Tn5* lysogens with normal phase transition was approximately 10^{-2} per cell per generation. Among the mutant candidates, thirteen independently isolated clones showed reduced *H2* switching frequencies (see Table 1).

Two classes of mutant phenotypes were apparent: one (intermediate class) showed approximately one fifth the frequency of *H2* switching, while a second (null class) showed an approximately 1000 fold reduction in the frequency of phase transition. These phenotypes were characteristic of the hybrid λ, since the isolated phage could be used to prepare new lysogens which had the same altered frequency of phase transition. Furthermore, the effect of the insertion was symmetrical; that is, *H2* (off) derivatives of mutants which switched from *H2* (on) to *H2* (off) at reduced frequencies showed similar reductions in transition in the opposite direction from *H2* (off) to *H2* (on).

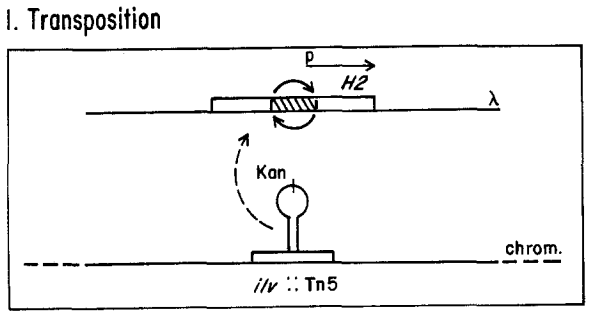


Figure 2. Isolation of Tn5-Induced *H2* Switching Mutants

Tn5 transposition to a hybrid λ phase containing the *H2* gene region resulted when this phage was grown in an E. coli strain containing Tn5 inserted in the chromosomal *ilv* gene (step 1). λ*H2::Tn5* insertion derivatives were isolated from the phage population by selection for lysogen cells which carried λ hybrids with the Tn5 kanamycin resistance determinant (step 2). Lysogens of λ*H2::Tn5* switching mutants were recognized by sensitivity to the flagellotropic phage χ (step 3).

Table 1. Switching Phenotypes of λ*H2* Mutants

λ Lysogen[a]	Mutant	Switching Frequency (per Generation)[b]
λ*fla157*	λ*H2*(wt)	1×10^{-2}
λ*fla250*	λ*H2*::Tn5	2×10^{-3}
λ*fla252*		
λ*fla255*		
λ*fla257*		
λ*fla242*	λ*H2*::Tn5	1×10^{-5}
λ*fla243*		
λ*fla244*		
λ*fla245*		
λ*fla247*		
λ*fla248*		
λ*fla251*		
λ*fla253*		
λ*fla254*		
λ*fla380*	λ*H2*(Δ)	1×10^{-2}
λ*fla385*		
λ*fla378*	λ*H2*(Δ)	2×10^{-5}
λ*fla381*		
λ*fla364*	λ*H2*(Δ)	$<1 \times 10^{-6}$
Mutant Class		wt / Intermediate / Null

[a] λ*H2* mutants were used to lysogenize E. coli Hag⁻ strain MS6302. See Figures 3 and 4 for description of hybrid λ.

[b] *H2* switching of lysogens was measured in the *H2* (on) to *H2* (off) direction as described in Experimental Procedures. Actual values varied ±50% of those shown above.

Localization of the Tn5 Insertions

Restriction fragment analysis was used to determine the location of the Tn5 insertions in the DNA of the thirteen λ*H2::Tn5* hybrid phage with switching defects, one hybrid phage with a H2⁻ phenotype and 19 hybrid phage with nondefective *H2* switching phenotypes. Initially, the phage DNA was restricted with Eco RI, which cleaves the λ*H2::Tn5* genome into five discrete restriction fragments. Insertion of transposon Tn5, which contains no Eco RI site in 5200 bp of length, into a particular Eco RI fragment markedly altered the size of that restriction fragment. Transposon Tn5 was localized in the 3.75 bp *H2* gene Eco RI insert in the case of all λ*H2::Tn5* mutants and the H2⁻ λ*H2::Tn5* hybrid, but with all hybrid phage with a nondefective switching phenotype Tn5 insertion was located in other regions of the hybrid λ genome. Further restriction analysis gave the approximate position of insertion within the Eco RI fragment and the orientation of insertion of the Tn5 transposon (data not shown). To more precisely locate the points of insertions, the mutant λ*H2::Tn5* hybrids were restricted with Hpa II, which cleaved the *H2* gene insert into several well characterized fragments. The fragments were identified using the Southern blotting technique. Fragments with homology to the *H2* gene insert were detected by hybridization with a probe containing DNA from the central region of the *H2* gene insert (see Figure 3). From an analysis of the nature of the Hpa II restriction fragments from the region of Tn5 insertions, the location of each Tn5 insertion was unambiguously determined. For example, with the hybrid phage λ*fla243*, Tn5 insertion was in the 120 bp Hpa II fragment. Thus the 120 bp fragment disappears and two new fragments, resulting from the fusion of the 120 bp sequences to the arms of the Tn5, appear. Furthermore, the restriction pattern of the wild-type phage λ*fla157* contains fragments that are

characteristic of the *H2* (on) orientation (700 and 400 bp) and of the *H2* (off) orientation (560 and 540 bp). In the restriction pattern of λ*fla243* the two restriction fragments characteristic of the *H2* (off) orientation (540 and 560 bp) were missing. This demonstrates that there is little or no switching at the molecular level. The phenotype of a lysogen containing this phage—that is, the null level of transition—is consistent with this observation. Transposon Tn5 insertion in hybrid phage λ*fla250* was in the 700 bp fragment, but fragments characteristic of the "off" orientation were also present. Thus, again in agreement with the phenotype of this phage, inversion occurred at intermediate frequencies. From these results and other restriction analysis, the location of Tn5 insertions in the mutant λ*H2::Tn5* hybrids shown in Figure 3 was determined. With all the switching mutants (either intermediate or null class) the location of Tn5 insertion was clearly within the invertible control region. With hybrid λ*fla249*, which has an H2$^-$ phenotype, Tn5 insertion had occurred in a region known to contain the *H2* structural gene (Silverman et al., 1979b).

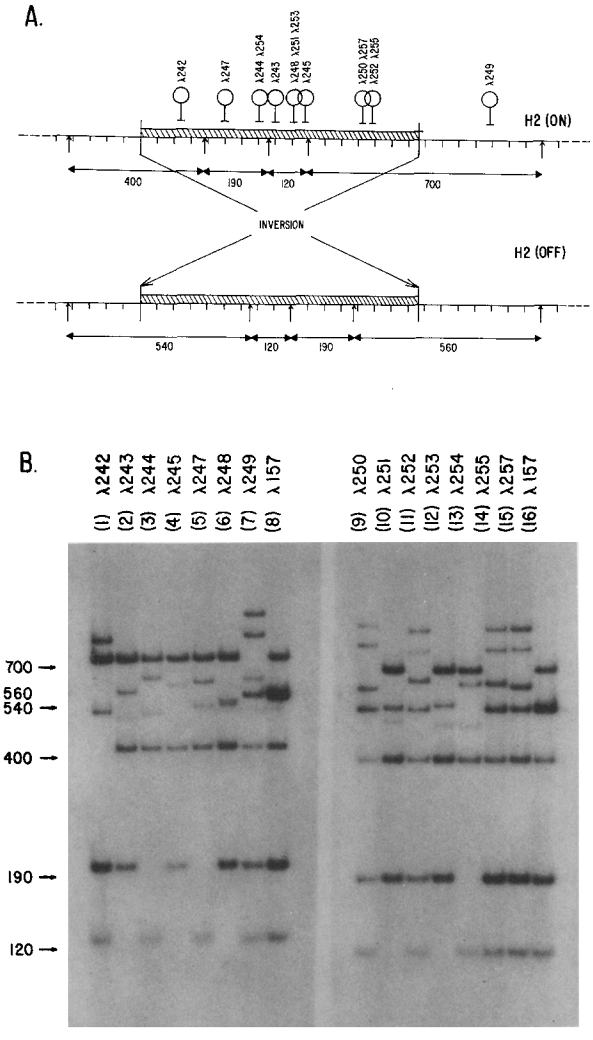


Figure 3. Location of Tn5 Insertions in *H2* Switching Mutants

The location of Tn5 insertions in mutant phage described in Table 1 is shown in (A). The inversion region (cross-hatched) and adjacent DNA sequences (*H2* gene on right) are shown in both *H2* (on) and *H2* (off) configurations. Restriction of DNA at Hpa II sites (vertical arrows) resulted in six DNA fragments. The 700 and 400 bp fragments are characteristic of the *H2* (on) phase and the 560 and 540 bp fragments are characteristic of the *H2* (off) phase. Length of DNA is marked in 50 bp intervals (vertical lines) and positions of Tn5 insertion are accurate to approximately 25 bp. Tn5 positions were determined in part from analysis of Hpa II restriction fragments of λ*H2::Tn5* DNA shown in (B). Hpa II fragments from mutant phage were transferred to nitrocellulose paper from acrylamide gels and hybridized to a ^{32}P-labeled probe (PJZ121) containing *H2* region sequences. Fragments characteristic of wild-type *H2* phage DNA (λ*fla157*) are shown at left. Interpretation of pattern of Hpa II fragments from Tn5 insertion mutants is described in the text.

Transposition-Generated Deletions

Tn5 stimulates the formation of deletions adjacent to its point of insertion (Berg, 1977; Kleckner, 1977). The deletions usually have one endpoint within the transposable element and extend in either direction from that point into adjacent sequences. Deletion mutants of hybrid λ*H2::Tn5* phage generated by Tn5 transposition were obtained by chelating agent (Na pyrophosphate) selection (Parkinson and Huskey, 1971). Deletions originating at Tn5 insertion points inside the inversion region in λ*fla250*, λ*fla252*, λ*fla255* and λ*fla257* are particularly interesting, since they retain most of the phage transition function, and the loss of remaining function can be correlated with deletions. 72 deletions were selected. All of these deletions had lost the kanamycin determinant. Some of these were the result of apparently precise excision events, and phase transition and the integrity of the restriction fragments was completely restored. Others resulted in a variety of deletions. These were analyzed genetically and by Southern transfer hybridization (Southern, 1975). The orientation of the deletion could be easily determined, since one fragment resulting from the insertion remained unchanged while the other fragment was either shortened or eliminated. The size

The location of Tn5 insertion could be correlated with the mutant phenotype. The four λ*fla* phages that showed an intermediate switching phenotype all carried the Tn5 insertion within a 100 bp sequence inside the inversion region (see Figure 3). All mutant phage with the null phenotype contained Tn5 insertions within the 120, 190 and 400 bp fragments, including a target of about 500 bp. It is of interest to note that elongation of the 900 bp inversion region by insertion of the 5200 bp Tn5 transposon apparently had little effect (approximately 5 fold) on the frequency of phase transition (for example, λ*fla250*, λ*fla252*, λ*fla255*, λ*fla257*). On the other hand, Tn5 insertion into the 500 bp region defined by hybrids such as λ*fla245* drastically reduced the ability of the region to invert.

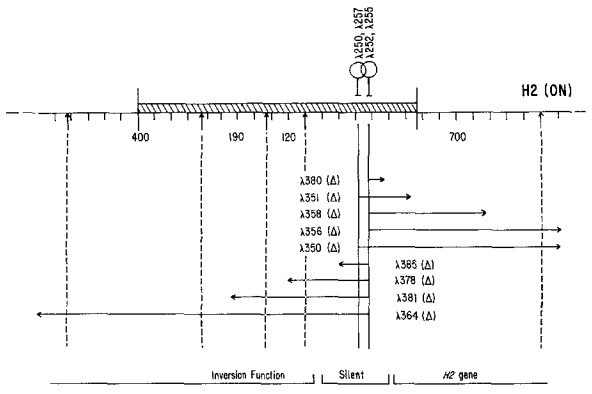


Figure 4. Deletion Mutants Derived from λ*H2::Tn5* Phage

Deletion mutants which arose by Tn5-mediated transposition were derived from four λ*H2::Tn5* mutants (intermediate switching class) shown above the Hpa II restriction map of the *H2* region in the *H2* (on) phase. One terminus of the deletion is within an arm of the Tn5 element (solid vertical lines), and deletion extends through Tn5 DNA (kanamycin determinant is deleted) to another terminus in *H2* DNA (horizontal arrows). Deletion endpoints shown in *H2* DNA are imprecise, but were accurate relative to Hpa II sites (vertical dotted lines), to crossover points for inversion (vertical line at ends of inversion sequence), and to other deletion endpoints. Mapping of deletions was by restriction analysis such as that shown in Figure 3B. Functions defined by the phenotype of the mutant phage are summarized at the bottom of the figure. Other details of the map are the same as Figure 3A.

of deletions shown Figure 4 is approximate, since it is not clear exactly how much of the Tn5 transposon remains fused to the *H2* sequence. An estimate of the minimum size of the *H2* region that was removed could be made, however, since the deletion endpoints were accurate relative to Hpa II restriction sites, crossover points for inversion and other deletion endpoints. The ability of the remaining material to invert could be determined from the presence or absence of the restriction fragments characteristic of inversion. Figure 4 summarizes the results of the deletion mapping experiments.

On the basis of our examination of the physical and genetic properties of these phages, the following conclusions were drawn. First, deletion past the crossover points for the inversion event always abolished inversion. Thus inversion did not occur with hybrids λ*fla358*, λ*fla350*, λ*fla356* and λ*fla364*. Second, inversion did not take place if the deletion extended into the region previously defined as necessary for inversion. Thus deletion into the 120 or 190 bp fragments, for example, λ*fla378* and λ*fla381*, markedly reduced inversion, and extensive deletion, as in λ*fla364*, completely eliminated inversion. On the other hand, λ*fla385* showed inversion. In fact, the *H2* switching phenotype of hybrid λ*fla385* was similar to that of the wild-type *H2* phage (λ*fla157*) rather than that of the parent λ*H2::Tn5* hybrid (λ*fla255*). The reduction in switching frequency observed with the intermediate class switching mutants (that is, λ*fla255*) could therefore be attributed to elongation of the inversion region by insertion of the Tn5 element and not to the interruption of a particular switching function located at the point of Tn5 insertion. Third, λ*fla380* and λ*fla351* showed wild-type switching frequencies as measured by restriction analysis; however, hybrid λ*fla351* had lost the ability to express the *H2* gene. As expected, hybrids λ*fla350*, λ*fla356* and λ*fla358* also had the H2$^-$ phenotype. These data suggest that the location of the *H2* gene promotor is probably within the first 100 bp of the invertible control region. We have not determined whether any of the coding sequences for the *H2* gene product are also located within this segment. In addition to the sites where inversion takes place, these regions, shown in Figure 4, have been defined: a region containing the *H2* gene and its promotor, a region inside the invertible segment which has little apparent effect on switching frequency (described as "silent" in Figure 4), and a region of approximately 500 bp inside the invertible segment which is necessary for *H2* gene switching.

Complementation Analysis

Tn5 insertions in a region of approximately 500 bp inside the inversion sequence (that is, λ*fla251* and so on) resulted in loss of a function necessary for normal *H2* gene switching. This region of DNA is well separated from the sites where crossover takes place during inversion—indeed, some Tn5 insertions which drastically reduce inversion are located approximately 500 bp from the crossover points. It is conceivable that this region functions as a site which acts in a cis manner to activate inversion; for example, this region might act as a recognition site for a protein which catalyzes the inversion process. Alternatively, this region might be a gene and code for a protein factor which functions in trans to catalyze switching. To test the latter explanation, a genetic arrangement was devised to determine whether the defect in the λ*H2::Tn5* switching mutants could be complemented in trans. Figure 5 summarizes the strategy for a test by hybrid phage or plasmids containing the putative gene necessary for switching. E. coli cells were co-infected with a λ*H2::Tn5* mutant (*H2* fixed on) and a λ*H2* deletion mutant which contained the region of DNA necessary for inversion, the mixture of phage was isolated, Kan^r lysogens of a Hag$^-$ strain were prepared and the H2 phenotype of the lysogen was measured. The trans complementing phage was either deletion mutant λ*fla350* or λ*fla356* (see Figure 5), both of which were H2$^-$ and incapable of inversion but contained the distal portion of the inversion region. To eliminate the production of recombinant phage which might mimic the phenotype of a mutant λ*H2::Tn5*, infection was done with Red$^-$ λ derivatives in a RecBC$^-$ strain. The results of this complementation analysis are shown in Table 2. It is apparent that when the region containing the putative inversion controlling gene was provided in trans, a significant proportion of

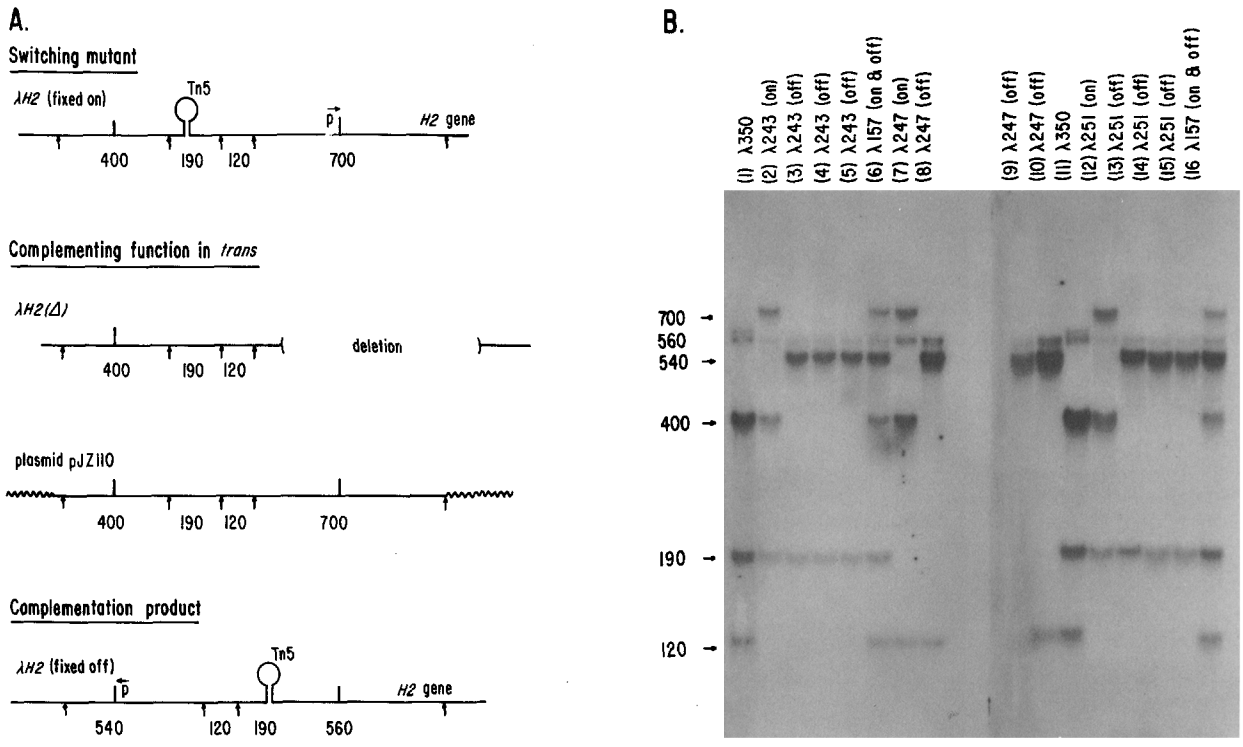


Figure 5. Complementation Analysis of Switching Mutants

The genetic test to determine whether λ*H2::Tn5* mutants could be complemented to switch, *H2* (on) to *H2* (off), by providing a function in trans is shown in (A). Null class λ*H2::Tn5* switching mutants in the *H2* (on) phase were used to infect cells which were either co-infected with a λ*H2* deletion mutant (trans function donor) or which already carried a hybrid *H2* plasmid (trans function donor). The resulting phage lysate was used to form λ*H2::Tn5* lysogens, and the H2 phenotype of the lysogen was then measured by the χ phage resistance test. Table 2 summarizes the results of this test. DNA purified from λ*H2::Tn5* mutants in the *H2* (on) configuration, a λ*H2* deletion derivative used for trans complementation, and λ*H2::Tn5* mutants which had been complemented to switch to the $H2^-$ phenotype (χ phage-resistant lysogen) were examined by hybridization analysis of Hpa II restriction fragments (see Figure 3B and Experimental Procedures). Figure 5B shows the result of this analysis. The source of the DNA was: wild-type phage λ*fla241* (lanes 6, 16); trans function donor λ*fla350* (lanes 1, 11); and λ*H2::Tn5* mutants λ*fla243* (lanes 2, 3, 4, 5), λ*fla247* (lanes 7, 8, 9, 10) and λ*fla251* (lanes 12, 13, 14, 15). DNA from λ*H2::Tn5* phage in the *H2* (on) phase was used in lanes 2, 7, 12 and DNA from λ*H2::Tn5* phage which had been complemented to switch to *H2* (off) was used in lanes 3, 8, 13 (trans donor was λ*fla350*), lanes 4, 9, 14 (trans donor was pJZ110) and lanes 5, 10, 15 (trans donor was pJZ143). The Hpa II fusion fragments generated by Tn5 insertion (see Figure 3B) were apparent only upon long exposure of the DNA blot. Phage λ*fla241* is identical to λ*fla157* (Figure 3B and Table 1) except that it contains a Tn5 insertion in a dispensable part of the λ genome.

the mutant population switched to the H2 (off) phenotype. In fact, mutant λ*H2::Tn5* could be complemented to switch to *H2* (off) at almost the same frequency as λ*H2::Tn5* mutants with the intermediate switching phenotype grown without a complementing phage. Phage DNA was purified from the *H2* (off) (χ phage-resistant) lysogens carrying the λ*H2::Tn5* mutants and subjected to restriction analysis. It was clear that the DNA from these lysogens was indeed in the *H2* (off) configuration (see Figure 5B). For example, DNA from λ*fla251* grown without complementing phage showed only the 700 and 400 bp fragments characteristics of the *H2* (on) configuration, but DNA from an $H2^-$ lysogen (χ phage-resistant) containing λ*fla251* previously grown with a complementing phage showed only the 540 and 560 bp fragments characteristic of the *H2* (off) configuration. Another complementation scheme was used in which the trans complementing region was provided by hybrid plasmids that had in common only the 900 bp *H2* inversion region (Zieg et al., 1978). In this case also (see Figure 5 and Table 2) the λ*H2::Tn5* switching mutants (null class) could be complemented to switch from *H2* (on) to *H2* (off).

These results suggest that there is a gene of approximately 500 bp in length in the inversion region which encodes a product that functions to cause inversions of the *H2* control region. Measurement of complementation of switching was also carried out with λ*H2::Tn5* mutants set initially in *H2* (off) configuration, and these mutants could be complemented to switch to *H2* (on) by providing the same DNA region in trans used for *H2* (on) to *H2* (off) complementation. Thus this factor acts to induce inversion in both directions. Even though *H2* switching was reduced by three orders of magnitude in the mutants, the *H2* phenotype was not absolutely fixed. The low residual level of H2 switching [both *H2* (on) to *H2* (off) and *H2* (off) to *H2* (on)] was observed with the λ*H2::Tn5* mutants only when a Rec^+Hag^- E. coli lysogen was examined.

Table 2. Complementation of λ*H2::Tn5* Switching Mutants

λ*H2::Tn5* Mutant[a]	λ or Plasmid in Trans[b]	Switching Frequency (per Generation)[c]	*hin* Function
λ*fla241*		8×10^{-3}	+
λ*fla243*, λ*fla244* λ*fla247*, λ*fla251*		5×10^{-6}	–
λ*fla243*, λ*fla244* λ*fla247*, λ*fla251*	λ*fla350*, λ*fla356*	1×10^{-3}	+
λfla243, λfla244 λfla247, λfla251	λ*fla364*, λ*fla381*	5×10^{-6}	–
λ*fla243*, λ*fla244* λ*fla247*, λ*fla251*	pBR322	5×10^{-6}	–
λ*fla243*, λ*fla244* λ*fla247*, λ*fla251*	pJZ110, pJZ121, pJZ143	2×10^{-3}	+

[a] All λ*H2::Tn5* are Red⁻ derivatives. Hybrid λ*fla241* has wild-type switching function and is used for comparison with mutant phage.
[b] See Figure 3 for description of hybrid phage used for trans complementation and Zieg et al. (1978) for description of plasmids.
[c] See Experimental Procedures for detailed explanation of complementation test. Switching frequencies are averages, with actual values varying ±50%.

When RecA⁻Hag⁻ lysogens were examined, no "spontaneous" switching was observed. Apparently the host recombination system (RecA) can mediate a low but measurable level of inversion of the *H2* control region. This "spontaneous" switching was observed with deletion derivatives of the λ*H2::Tn5*, but only if the inversion crossover regions were intact, and deletion derivatives of λ*H2::Tn5* phage could also be complemented to switch, but only if the inversion crossover regions were present (data not shown). This indicates again that cis-acting sites located at the cross-over points are absolutely essential for *H2* control region inversion.

Discussion

The 900 bp invertible region that controls phase variation includes a gene which encodes a product required for its own inversion. This *hin* gene is defined by deletion and insertion mutations that are localized within a DNA sequence of approximately 500 bp. An intact *hin* gene is required for phase transition, and the *hin* gene product promotes site-specific inversion in both directions. Genetic complementation tests show that the *hin* gene product acts in the trans configuration. Its presence increases the frequency of inversion by at least three orders of magnitude. Furthermore, the *hin* function is not sensitive to the size of the region that is inverted. Thus insertion of Tn5 into the silent region of the invertible segment, which increases its size by 5200 bp, and deletions which decrease its size by approximately 100 bp have only small effects on the frequency of phase transition. On the other hand, deletions that remove the sequences which contain the crossover points completely eliminate inversion. Genetic complementation cannot overcome the effects of these deletions. The crossover sites therefore behave as cis-acting elements whose participation is required for site-specific inversion.

The simplest interpretation of our data is that the *hin* gene product is a protein that catalyzes site-specific inversion. We have identified a 19,000 dalton polypeptide that is encoded by sequences within the inversion region (M. Silverman and M. Simon, manuscript in preparation) and may therefore be the product of the *hin* gene. It would function as an enzyme which is able to recognize sequences at the crossover points and catalyze the inversion event. Alternatively, it may participate in the inversion process by conferring site specificity to some more general recombination system which is endogenous to E. coli and Salmonella.

The *hin*-mediated inversion occurs independently of the RecA or RecBC systems (Zieg et al., 1978). Even in the absence of the *hin* gene, however, there is still a low residual level of phase variation. This is seen clearly when the frequency of transition is compared in deletions that are missing the *hin* function and those that have lost the crossover sequences. The loss of the crossover sequence completely eliminates all of the residual phase transition events. Furthermore, the low level residual switching in the Hin⁻ mutants disappears if the genes are put in a RecA host. These results, taken together, suggest that phase transition can occur through a "legitimate" recombinational mechanism, albeit at very low frequencies (approximately 10^{-6} per cell per generation). The crossover points may contain homologous sequences that can be recognized by the RecA system and lead to inversion (Anderson and Roth, 1977). The *hin*-mediated system provides a site-specific, RecA-independent pathway which allows phase transition to occur at much higher frequencies (10^{-2} per cell per generation).

In addition to the *hin* and the crossover functions, deletions define a region inside the invertible segment which is necessary for *H2* gene expression. We suggest that this deletion defines a promotor sequence which is located close to the crossover point. In previous work, we have shown that transcription of the *H2* gene is controlled by a sequence inside the inversion region, and phase transition operates by connecting or disconnecting this promotor from the adjacent *H2* structural gene sequence (Silverman et al., 1979b). The insertions and deletions also define a short sequence (100–200 bp) between the promotor and the *hin* gene sequences which does not appear to be required for phase variation. These sequences could possibly code for function necessary for the regulation of *H2* promotor activity. The scheme pre-

sented in Figure 6 summarizes our conclusions about the functions encoded by the DNA involved in phase transition.

The sequences that code for phase transition are similar to those involved in the G loop inversion in bacteriophage Mu. The *hin* gene function appears to be analogous to the *gin* gene function, which has been shown to be required for G loop inversion in bacteriophage Mu (Chow, Kahmann and Kamp, 1977). In fact, it appears that trans-acting functions required for site-specific events are generally found to be associated with the region of DNA in which the recombination event occurs. For example, the *int* gene is λ maps adjacent to the *att* site, where it mediates integrative recombination (Gottesman and Weisberg, 1971; Nash, 1977). Furthermore, the Tn3 transposon appears to code for polypeptides that function to catalyze and regulate transposition (Chou et al., 1979; Heffron et al., 1979). The possibility has been raised that even short sequences such as the IS-1 sequence may code for polypeptides that could be involved in IS transposition (MacHattie and Shapiro, 1978; Ohtsubo, Ohmori and Ohtsubo, 1978). The ability to code for factors required for site-specific recombination may be a general feature of transposon-like sequences and controlling elements.

In the case of Tn3 there are factors which control the frequency of transposition that are also part of the transposable sequence. One could imagine that the frequency of phase transition could be regulated by changes in the *hin* gene or in the promotor region that controls its expression. Furthermore, the specific sequences at the crossover points may also be important in regulating the frequency of inversion. The differences in the frequency of phase transition in naturally occurring Salmonella strains may be the result of such variations. One possible scenario for the origin of the phase system is that it may have arisen from the association of a primordial *H1* gene with a transposable element that also carried functions that could mediate site-specific inversion. Subsequent mutation could have provided the optimum frequency of phase transition for different Salmonella strains.

The DNA sequence of the *H2* region is currently being determined. It will allow a more precise definition of *H2* switching function and closer comparison with other systems that show site-specific recombination. The invertible sequence represents a new class of regulatory elements that effect gene expression via site-specific recombination. It is possible that such mechanisms operate in both eucaryotic and procaryotic cells to allow the cell to express alternate genes for a specific function. This type of system could provide diversity, particularly for functions that involve surface properties and are required for the physical interaction of cells with each other or with the environment. It could also be the prototype for regulatory events involved in cell differentiation and development.

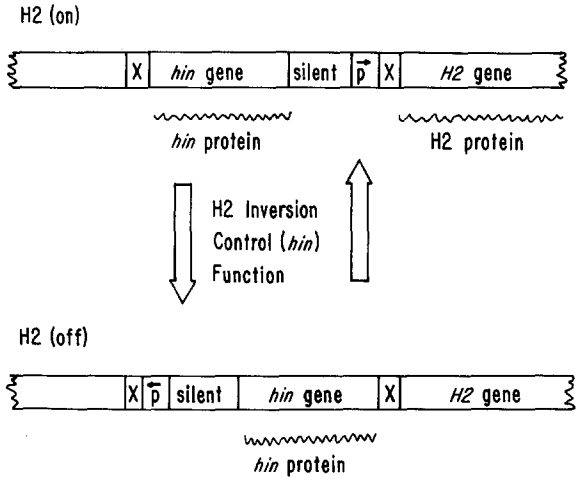


Figure 6. Model for Inversion of *H2* Control Region

Switching of *H2* gene expression is controlled by the orientation of an invertible region of DNA which contains a promotor for the H2 operon. Inversion of the control region requires specific sites (X) at the crossover points and is catalyzed by the product of a gene (*hin*) residing within the invertible region.

Experimental Procedures

Isolation of *H2* Switching Mutants

The construction of hybrid λ*fla157* (λ*H2*) which contains the 3.75 kb *H2* gene insert is described in Silverman et al. (1979b). Hybrid λ*fla157* has λ functions necessary for lysogenization, and its genome is small enough to accommodate Tn5 insertions and remain an acceptable size for DNA packaging. To obtain Tn5 insertion, λ*fla157* is grown lytically in E. coli strain DB1358 obtained from D. Berg (Washington University, St. Louis, Missouri). The resultant phage population should contain λ*H2::Tn5* phage at a frequency of approximately 10^{-4} (Berg, 1977; Kleckner, 1977; Kleckner, Roth and Botstein, 1977). Tn5 insertion phage were recovered by infection (multiplicity ≃1) of Hag^- kanamycin-sensitive strain MS6302 (Silverman et al., 1979b) with selection for Tn5-linked kanamycin resistance. Hybrid λ*H2::Tn5* lysogenic colonies on L-agar plates containing 40 μg/ml kanamycin (Sigma Chemical, St. Louis, Missouri) were recloned and screened for *H2* gene switching defects. Hybrid λ*H2::Tn5* switching mutants isolated in the *H2* fixed on orientation as lysogens of Hag^- E. coli strain MS6302 had a stable $H2^+$ phenotype and could be distinguished from lysogens of hybrid phage with functional *H2* switching which expressed both $H2^+$ and $H2^-$ phenotypes. Since the flagellotropic phage χ infects only $H2^+$ lysogens, lysogens with mutant λ*H2::Tn5* phage fixed in the *H2* (on) phase would be sensitive to χ, while lysogens with λ*H2::Tn5* phage which can switch to *H2* (off) would be resistant. Approximately 3000 kanamycin-resistant λ*H2::Tn5* lysogens were tested for χ sensitivity by inoculation onto motility agar plates with soft agar overlays containing 2×10^9 phage. (See Komeda, Silverman and Simon, 1978, for composition of motility agar plates and other media used for growth of bacteria and phage.) 100 candidate λ*H2::Tn5* phage with mutant *H2* switching phenotypes were saved as lysogens and analyzed quantitatively for the frequency of phase transition. Deletion mutants derived from λ*H2::Tn5* phage by Tn5-mediated transposition were selected by virtue of their resistance to chelating agents (Parkinson and Huskey, 1971; Ross, Swan and Kleckner, 1979). Selection of deletion mutants of various λ*fla* phage has been described (Silverman and Simon, 1977). Only those λ deletion mutants which lost the kanamycin determinant were saved for further analysis.

Measurement of *H2* Gene Expression

A quantitative method for measuring H2 gene switching, H2 (on) to H2 (off), has been described previously (Silverman et al., 1979b). As mentioned above, Hag$^-$ E. coli strains lysogenized with *H2* phage are either sensitive or resistant to the flagellotropic phage, depending upon the state of expression of the *H2* gene. To measure the *H2* switching frequencies, lysogen populations were initially enriched for the *H2* (on) phase by inoculating into motility agar where only *H2* (on) bacteria migrate and then growing cells harvested from the edge of a migrating swarm of bacteria. These *H2* (on) cells were grown for approximately 10 generations in L broth and diluted, and appropriate amounts were plated in overlay agar with and without 10^9 χ phage. The ratio of χ phage-resistant colonies to the total number of cells plated gave the fraction of cells which had switched to the *H2* (off) phase. This ratio divided by the number of generations the lysogen population had grown was used as the frequency of *H2* switching (see Stocker, 1949). Kanamycin was present in the bottom agar (motility plates) to exclude any cells not lysogenized with λ*H2::Tn5* phage. Measurement of *H2* (off) to *H2* (on) switching with the χ phage selection method was subject to large error. Because the switching products, *H2* (on) lysogens, were sensitive to χ phage, their presence was measured indirectly by calculating the difference between the total number of lysogens and the number of χ-resistant lysogens. Since the switching frequency in the mutants is low, the difference in the number of chi-resistant lysogens was small and therefore difficult to assess accurately. *H2* switching in this direction could be determined in a qualitative manner, however, by inoculating a particular lysogen in a zone on motility agar plates, and, after incubation at 37°C for 8 hr, estimating the number of motile swarms, *H2* (on), emanating from the region of inoculation. Only lysogens with very large differences in *H2* (off) to *H2* (on) switching frequencies could be differentiated by this method, as opposed to the former method, where differences in *H2* (on) to *H2* (off) frequency of ±25% were detectable. E. coli strain MS6302, which is Hag$^-$ RecA$^+$, was used for most switching measurements, but a RecA$^-$Hag$^-$ derivative of Cold Spring Harbor strain CSH4 was used to measure switching in a RecA$^-$ environment.

Complementation Analysis

Complementation of the *H2* switching defects in λ*H2::Tn5* mutants (null class) was measured by growing the λ*H2::Tn5* mutants in cells which were co-infected with another hybrid or which contained a hybrid plasmid, the latter containing DNA sequences being tested for trans-acting function. To eliminate the possibility of recombination between the λ*H2::Tn5* mutant and homologous regions on the λ or plasmid DNA in the trans configuration, Red$^-$ phage and a RecBC$^-$ host were used. RecBC$^-$ hosts supported growth of Red$^-$ phage, whereas RecA$^-$ hosts would not. Red$^-$ λ*H2::Tn5* derivatives were isolated by selection for their ability to grow on P2 lysogens (Spi$^-$ phenotype), and RecBC$^-$ strain JC5519 obtained from J. Clark (University of California, Berkeley) was used as host. Hybrid λ*H2::Tn5* synchronized in the *H2* (on) phase were obtained by ultraviolet induction of lysogens in the *H2* (on) phase. Host strain JC5519 was co-infected with a λ*H2::Tn5* mutant and a second λ (function donor) or host strain JC5519 harboring a hybrid plasma (function donor) was infected with the λ*H2::Tn5* mutant. Multiplicity of infection was approximately two, growth was at 37°C and one infection cycle was completed. Cells were infected with the resultant lysate until three cycles of phage growth were completed. This lysate was used to lysogenize Hag$^-$ strain MS6302. Infection of MS6302 was at a multiplicity of one at 30°C, and cells were allowed to grow for 16 hr, at which time they were diluted 1/100 into L broth containing 20 μg/ml kanamycin to select for λ*H2::Tn5* lysogens. After 8 hr of growth, the fraction of lysogens with λ*H2::Tn5* phage in the *H2* (off) phase was determined by the χ test (see above). The results in Table 2 are presented as the frequency of change of the *H2* phenotype [proportion *H2* (off) per generation], as are those in Table 1. However, since the λ*H2::Tn5* phage were grown not as lysogens, but vegetatively during the course of the complementation test, the term "generation" has a different meaning. Thus the values in Table 1 and Table 2 should not be compared directly. We have estimated the number doubling of the λ phage population during the course of the complementation test to be approximately 30, and use this value for "generation" in computing the switching frequencies shown for the complementation test.

The *H2* DNA sequences contained on the hybrid plasmids used in the complementation test are described in detail in Zieg et al. (1978). The *H2* DNA contained in plasmid pJZ110 is shown in Figure 5A. Plasmids pJZ121 and pJZ143 are recombinational variants derived from plasmid PJZ110. Plasmid pJZ121 contains the inversion sequence flanked on either side by part of the *H2* gene sequence (right arm in Figure 5A) and plasmid pJZ143 contains the inversion sequence flanked by the non-H2 gene sequence (left arm in Figure 5A). All three plasmids have in common the invertible region.

Restriction Analysis

Methods for restriction analysis of the *H2* gene region of hybrid λ phage have been described by Silverman et al. (1979b). Hpa II restriction fragments were separated electrophoretically on a 9% acrylamide gel. Where the method of Southern (1975) was used to make DNA transfer to nitocellulose paper, it was necessary to allow transfer to proceed for 24 hr to ensure complete transfer. ^{32}P-labeled probes were prepared from *H2* plasmids pJZ110 and pJZ121 (Silverman et al., 1979). When pJZ110 DNA was used as a probe for DNA bands transferred to nitrocellulose paper, one additional fragment between the 560 and 700 bp fragments was detected. This fragment is located outside the *H2* inversion region and is extraneous to this analysis. Isolation of hybrid λ DNA for restriction analysis was performed directly from phage lysates as described by Blattner et al. (1978).

Acknowledgments

We thank Dr. Janine Zieg for providing plasmid strains, Dr. Douglas Berg for providing Tn5 insertion strains, Dr. John Clark for Rec$^-$ strains, Dr. David Botstein for a restriction map of the Tn5 element, and Dr. Elizabeth Szekely for advice concerning the Southern transfer technique. This work was supported by grants from the NSF and the American Cancer Society.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 26, 1979

References

Anderson, R. P. and Roth, J. R. (1977). Tandem genetic duplications in phage and bacteria. Ann. Rev. Microbiol. *31*, 473–505.

Berg, D. E. (1977). Insertion and excision of the transposable kanamycin resistance determinant Tn5. In DNA Insertion Elements, Plasmids, and Episomes, A. Bukhari, J. Shapiro and S. Adhya, eds. (New York: Cold Spring Harbor Laboratory), pp. 205–212.

Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W. and Smithies, O. (1978). Cloning human fetal γ-globulin and mouse α-type globulin DNA: preparation and screening of shotgun collections. Science *202*, 1279–1284.

Bukhari, A. I. and Ambrosio, L. (1978). The invertible segment of bacteriophage Mu DNA determines the adsorption properties of Mu particles. Nature *271*, 575–577.

Chou, J., Casadaban, M. J., Lemaux, P. G. and Cohen, S. N. (1979). Identification and characterization of a self-regulated repressor of translocation of the Tn3 element. Proc. Nat. Acad. Sci. USA *76*, 4020–4024.

Chow, L. T., Kahmann, R. and Kamp, D. (1977). Electron microscopic characterization of DNAs of non-defective deletion mutants of bacteriophage Mu. J. Mol. Biol. *133*, 591–609.

Fujita, H., Yamaguchi, S. and Iino, T. (1973). Studies of H-O variants in *Salmonella* in relation to phase variation. J. Gen. Microbiol. *76,* 127–134.

Gottesman, M. E. and Weisberg, R. A. (1971). Prophage inserion and excision. In The Bacteriphage Lambda, A. D. Hershey, ed. (New York: Cold Spring Harbor Laboratory), pp. 113–138.

Heffron, F., McCarthy, B. J., Ohtsubo, H. and Ohtsubo, E. (1979). DNA sequence analysis of the transposon Tn3: three genes and three sites involved in transposition of Tn3. Cell *18,* 1153–1163.

Hicks, J. B., Strathern, J. N. and Herskowitz, I. (1977). The cassette model of mating type interconversion. In DNA Insertion Elements, Plasmids, and Episomes, A. Bukhari, J. Shapiro and S. Adhya, eds. (New York: Cold Spring Harbor Laboratory), pp. 457–462.

Iino, T. (1961). A stabilizer of antigenic phase in *Salmonella abortus-equi.* Genetics *46,* 1465–1469.

Kamp, D., Kahmann, R., Zipser, D., Broker, T. R. and Chow, L. T. (1978). Inversion of the G DNA segment of phage Mu controls phage infectivity. Nature *271,* 577–580.

Kleckner, N. (1977). Translocatable elements in procaryotes. Cell *11,* 11–23.

Kleckner, N., Roth, J. and Botstein, D. (1977). Genetic engineering *in vivo* using translocatable drug-resistance elements. J. Mol. Biol. *166,* 125–159.

Komeda, Y., Silverman, M. and Simon, M. (1978). Identification of the structural gene for the hook subunit protein of *Escherichia coli* flagella. J. Bacteriol. *133,* 364–371.

Kushner, P. J., Blair, L. C. and Herskowitz, I. (1979). Control of yeast cell types by mobile genes: a test. Proc. Nat. Acad. Sci. USA *76,* 5264–5268.

Lederberg, J. and Edwards, P. (1953). Serotypic recombination in *Salmonella.* J. Immunol. *71,* 323–340.

Lederberg, J., and Iino, T. (1956). Phase variation in *Salmonella.* Genetics *41,* 743–757.

MacHattie, L. A. and Shapiro, J. A. (1978). Chromosomal integration of phage λ by means of a DNA insertion element. Proc. Nat. Acad. Sci. USA *75,* 1490–1494.

Nash, H. (1977). Integration and excision of bacteriophage λ. Cur. Topics Microbiol. Immunol. *78,* 171–199.

Ohtsubo, H., Ohmori, H. and Ohtsubo, E. (1978). Nucleotide-sequence analysis of Tn3 (Ap): implications for insertion and deletion. Cold Spring Harbor Symp. Quant. Biol. *43,* 1269–1277.

Parkinson, J. S. and Huskey, R. J. (1971). Deletion mutants of bacteriophage lambda. J. Mol. Biol. *56,* 369–384.

Ross, D. G., Swan, J. and Kleckner, N. (1979). Nearly precise excision: a new type of DNA alteration associated with the translocatable element Tn10. Cell *16,* 733–738.

Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979). Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature *280,* 288–294.

Silverman, M. and Simon, M. (1977). Identification of polypeptides necessary for chemotaxis in *Escherichia coli.* J. Bacteriol. *130,* 1317–1325.

Silverman, M., Zieg, J. and Simon, M. (1979a). Flagellar-phase variation: isolation of the *rhl* gene. J. Bacteriol. *137,* 517–523.

Silverman, M., Zieg, J., Hilmen, M. and Simon, M. (1979b). Phase variation in *Salmonella*: genetic analysis of a recombinational switch. Proc. Nat. Acad. Sci. USA *76,* 391–395.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98,* 503–517.

Stocker, B. A. D. (1949). Measurements of rate of mutation of flagellar antigenic phase in *Salmonella typhimurium.* J. Hygiene *47,* 398–413.

Zieg, J., Hilmen, M. and Simon, M. (1978). Regulation of gene expression by site-specific inversion. Cell *15,* 237–244.

Zieg, J., Silverman, M., Hilmen, M. and Simon, M. (1977). Recombinational switch for gene expression. Science *196,* 170–172.

J. Mol. Biol. (1975) **97**, 561–575

Mutagenesis by Insertion of a Drug-resistance Element Carrying an Inverted Repetition

NANCY KLECKNER, RUSSELL K. CHAN†, BIK-KWOON TYE‡ AND DAVID BOTSTEIN

Department of Biology
Massachusetts Institute of Technology
Cambridge, Mass. 02139, U.S.A.

(Received 19 June 1975)

A novel genetic element, which carries genes conferring tetracycline resistance (flanked by a 1400 base-pair inverted repetition), is capable of translocation as a unit from one DNA molecule to another. The tet^R element, which is found in nature on a variety of R-factors, was acquired by bacteriophage P22 (producing P22Tc-10 and P22Tc-106) and has now been observed to insert into a large number of different sites on the *Salmonella* chromosome. Insertion of the tet^R element is mutagenic when it occurs within a structural gene, and polar when it occurs within an operon. Insertion of the element is usually precise, occurring without loss of information on the recipient DNA molecule. Excision, on the other hand, is usually *not* precise, although excisions precise enough to restore a gene function can always be detected at low frequencies. Both insertion and excision processes are independent of the *recA* function.

1. Introduction

Bacteriophage P22 is a temperate phage whose normal host is *Salmonella typhimurium*. We described previously an unusual variant of P22 (called P22Tc-10) which transduces resistance to tetracycline at high frequency (Watanabe *et al.*, 1972; Chan *et al.*, 1972). Examination of P22Tc-10 DNA in the electron microscope showed that this specialized transducing variant contains a large (8·3 kilobase) insertion which has an unusual structure: in heteroduplex DNA molecules, the insertion forms a lariat-like structure with a double-stranded stem (about 1·4 kilobases long) and a single-stranded loop (Tye *et al.*, 1974). This is interpreted to mean that the insertion consists of an inverted duplication separated by non-repeated DNA sequences.

The tet^R insertion was acquired by P22 during a lytic cycle of growth in a *Salmonella* strain harboring a drug-resistance plasmid (R-factor) in whose DNA a similar non-tandem reverse duplication was found to be associated with the genetic determinant for tetracycline resistance (Watanabe *et al.*, 1972; Sharp *et al.*, 1973). Two independently-arising tetracycline-transducing P22 phages were examined in our previous studies and both had identical insertions (Watanabe *et al.*, 1972; Tye *et al.*,

† Present address: Department of Genetics, SK-50, University of Washington, Seattle, Wash. 98195, U.S.A.

‡ Present address: Department of Biochemistry, Stanford University School of Medicine, Stanford, Calif. 94305, U.S.A.

1974). In neither case was there any detectable (less than 100 bases) loss of P22 DNA accompanying the insertion.

More recently, we found that the site of the *tet*^{R} insertion in P22Tc-10 is not the phage attachment site: i.e. the insertion is not at the point at which the prophage is integrated by site-specific recombination into the *Salmonella* chromosome (Chan & Botstein, 1975). This made it unlikely that P22Tc-10 was formed by the mechanism normally associated with the production of specialized-transducing genomes of temperate phages (Campbell, 1962).

From these observations, and from other aspects of the genetic behavior of P22Tc-10, we formed the hypothesis that the *tet*^{R} element from the R-factor (with its inverted repeat) is capable of translocation, as a discrete unit, from one DNA molecule into any one of many different places on other DNA molecules. This paper describes both physical and genetic evidence in support of this idea.

Berg *et al.* (1975) have recently isolated coliphage λ-transducing variants carrying a kanamycin-resistant determinant which is associated with an inverted repetition. The structures of these phages is analogous to that of P22Tc-10. Heffron *et al.* (1975*a*,*b*) have recently shown that an ampicillin-resistance determinant, associated with a very small inverted repetition, is capable of translocation from one DNA molecule to another.

We suggest that genetic elements like the *tet*^{R}, *amp*^{R} and *kan*^{R} insertions play an important role in reassortment of drug-resistance determinants among resistance factors. The apparent ubiquity of inverted repetitions in both prokaryotic (Sharp *et al.*, 1972, 1973; Daniell *et al.*, 1973; Berg *et al.*, 1975; Heffron *et al.*, 1975*b*; Hsu & Davidson, 1975) and eukaryotic DNA (Garon *et al.*, 1972; Wolfson & Dressler, 1972; Locker *et al.*, 1974; Manning *et al.*, 1975; Wadsworth *et al.*, 1975) further suggests that such elements could play an important role in the mobilization and translocation of genetic information in many different biological systems.

2. Materials and Methods

(a) *Bacteriophage*

The P22Tc-10 and P22Tc-106 genomes are too long to fit into a single P22 phage head. The strains are therefore maintained as lysogens which, upon induction, give rise to particles which are defective on single infection (Chan *et al.*, 1972; Tye *et al.*, 1974). Tc-10 recombinants carrying the additional mutations c_2ts29 (heat-inducible repressor (Levine & Smith, 1964)); *erf-am*12B (recombination-deficient (Botstein & Matz, 1970)); or int_3 (integration-deficient (Smith & Levine, 1967)) were constructed by crosses with induced lysates (Chan *et al.*, 1972) or rescue of *tet*^{R} from prophage deletions (Chan, 1974).

P22int_3 HT 12/4 was constructed in a standard phage cross between P22c^+int_3 and P22c_2^- HT 12/4 (Schmieger, 1972), a high-frequency generalized transducing derivative of P22.

(b) *Bacteria*

Strains of *S. typhimurium* used:

(i) In isolation of *tet*^{R} auxotrophs: DB7000 = LT2 *leu*A414 (Susskind *et al.*, 1974); NK80 = LT2 *edd*$^-$ constructed for this work from two *his*$^-$*edd*$^-$ strains obtained from J. Roth); and DB143 = LT7 *pro*AB47 (deleted for P22 attachment site) *recA*$^-$ (Miyake & Demerec, 1960).

(ii) In complementation studies: the following strains were all the gift of John Roth: (1) derivatives of LT2 *trpA*8 *purE*801 *his*612 (BHAFIE deletion) carrying *Escherichia coli* F'*his*bG2377, F'*his*abD2381, and F'*his*abD2382; (2) derivatives of LT2 *ser*821 *arg*501

*his*712 (DCBHAFIE deletion) carrying *E. coli* F'*his*$^+$, F'*his*C2383, F'*his*C2385, F'*his* a(b)cdB2405, F'*his*abI2312 a(b)E, F'*his* bI2413, F'*his* a(b)E2414, and F'*his* a(b)cd B245. (The small letters denote deficiencies of the mutant with respect to intragenic complementation groups.)

(iii) In deletion mapping: for *his* deletion strains listed in Fig. 2, endpoints of deletions with respect to known *his* point mutations are shown in Hartman *et al.* (1971) and Scott & Roth (1975). Deletion strains were kindly supplied by J. Roth and P. Hartman.

(c) *Media*

Complete liquid medium: LB (Chan & Botstein, 1972); solid minimal medium: M9 + 1·5% agar (Smith & Levine, 1964); complete solid media: green plates (Chan & Botstein, 1972), trypticase plates (Stahl & Stahl, 1971). P22 phage stocks are stored and diluted in buffered saline (Chan & Botstein, 1972).

(d) *Visualization of DNA heteroduplexes in the electron microscope*

Procedures used for the isolation of phage DNA and for preparation and visualization of DNA in heteroduplex structures were as described by Tye *et al.* (1974).

(e) *Isolation of* tet^R *auxotrophs*

Stocks of P22Tc-10 derivatives were made by induction of corresponding lysogens (Chan *et al.*, 1972); lysates were purified once through discontinuous CsCl gradients and the concentration of particles determined from the A_{260} (Chan *et al.*, 1972). Exponentially growing recipient bacteria were mixed directly with phage at a multiplicity of 2 to 13 particles per cell, and after 10 to 30 min, were spread directly on green plates supplemented with 25 μg tetracycline hydrochloride (Calbiochem)/ml and 0·01 M-EGTA (Eastman). Plates were incubated at 37°C (experiment A) or 41°C (experiments B and C) and then replica plated onto M9 plates supplemented with glucose and any other nutrient required by the recipient strain (DB143 is *pro*$^-$ and DB7000 is *leu*$^-$). In accordance with the notation introduced by Bukhari & Metlay (1973) for mutations made by insertion of phage Mu, the symbol :: will be used here to indicate that the *tet*^{R} element is inserted at the locus preceding the symbol. For example, *his* : : *tet*^{R} means that the *tet*^{R} element is inserted into the histidine operon.

(f) *Phage P22-mediated generalized transduction*

P22 is capable of mediating generalized transduction of bacterial DNA from one host strain to another (Zinder & Lederberg, 1952). For all generalized transduction experiments described here, the P22 derivative int_3 HT 12/4 was used. The HT 12/4 mutation, isolated and characterized by Schmieger (1972) and Raj *et al.* (1974), greatly increases the proportion of generalized transducing particles in a phage lysate. The int_3 mutation (Smith & Levine, 1967) prevents integration of normal P22 genomes into the host chromosome in the course of the transduction experiment.

Transducing lysates were made by putting a seed stock through a single cycle of growth on appropriate donor strains. For transducing lysates used in deletion mapping, the seed stock used to make single-cycle stocks was grown on a strain deleted for the entire histidine operon in order to prevent carry-over of *his*$^+$ transducing particles from the seed stock to the transducing lysates.

For co-transduction experiments (Table 2), exponential cultures of the recipient strain were infected directly with the various transducing lysates at a multiplicity of 10. After 15 min adsorption at room temperature, the mixtures were diluted and spread on green plates +25 μg tetracycline/ml, and the plates were incubated at 37°C. Resulting *tet*^{R} colonies were then tested for growth in M9–glucose plates, and on either M9 glucose and leucine, M9–glucose and histidine, or M9–gluconate plates in order to detect the donor auxotrophy.

For deletion mapping (Fig. 2), recipient *his*-deletion strains were grown to 2×10^9 cells/ml in LB broth. 0·15 ml bacteria, 0·1 ml transducing lysate (at 1 to 3×10^{10} phage/ml), and 0·15 ml LB broth were then spread together directly on an M9–glucose plate (without

prior pre-adsorption of phage and bacteria). Plates were incubated at 37°C for 2 days before scoring for the appearance of *his*$^+$ recombinants. With this protocol, a positive result meant that *at least* 50 (and as many as 10^4) *his*$^+$ recombinants appeared on a single plate. A negative result meant that *no* (0) *his*$^+$ recombinants appeared. As expected for deletion mutants, no *his*$^+$ revertants of the recipient strains were ever observed.

(g) *Complementation of F'*his *strains by* his*::*tet^R *auxotrophs*

The histidine operons of *S. typhimurium* and *E. coli* can complement each other for all of the functions in the histidine biosynthesis pathway; however, recombinants between the two operons are almost never observed (Atkins & Loper, 1970). We asked which *his* functions can be provided by *his*::*tet*^{R} auxotrophs by replica plating patches of *Salmonella* strains carrying F'*his* episomes (which themselves carried *his*$^-$ mutations in various genes) onto minimal plates spread with cultures of the various *his*::*tet*^{R} auxotrophs. The donor strain was counter-selected by omission of required nutrients (serine and arginine or tryptophan and purines). The F'*his* episome should mate into the *his*::*tet*^{R} strain, but growth in the region of a patch will be seen only if the *his*::*tet*^{R} recipient can supply the *his* function(s) not made by the F'*his*.

Cultures of donor F'*his* strains were spotted onto trypticase plates, and grown for 16 h at 37°C prior to replica plating. *His*::*tet*^{R} auxotrophs were grown to late log phase and 0·15 ml was spread directly on M9 plates. After replica plating, plates were incubated for 48 h at 37°C prior to scoring.

Various F$^-$*his*-deletion strains were included as control recipients to verify that the internal promoters detected by Atkins & Loper (1970) in such experiments were also detected here. A strain carrying an F'*his*$^+$ episome was included among donors as a control for the ability of *his*::*tet*^{R} auxotrophs to function as recipients and for the absence of negative complementation.

3. Results

(a) *DNA heteroduplexes between two independent P22* tet^R *transducing phage*

When a mixture of DNA from the two independent P22 *tet*^{R}-transducing phages (P22Tc-10 and P22Tc-106) was denatured, slowly reannealed, and examined by electron microscopy, a number of unusual structures were seen. The structure shown in Plate I(a) is apparently a heteroduplex molecule containing one strand of Tc-10 and one strand of Tc-106, since the two complementary strands are not the same. From this structure, it is clear that Tc-10 and Tc-106 each contain a single insertion which is unaccompanied by any detectable loss of P22 DNA (less than 50 to 100 bases). The Tc-10 and Tc-106 insertions are indistinguishable; both are of the same total length and are bounded by inverted repetitions of similar size. The only difference between the two transducing phages appears to be the location of the insertion into the P22 genome. This indicates that there is more than one possible site for insertion of the *tet*^{R} element.

Another structure which is seen in Tc-10/Tc-106 DNA heteroduplex preparations is shown in Plate I(b). This structure differs from that in Plate I(a) in that there has been partial pairing between the two single-strand "loops" of the inserted material. The pattern of paired and unpaired regions is reproducible among the many such structures we have seen. We interpret these structures to mean that the insertions carried by Tc-10 and Tc-106 are actually totally homologous, and that such structures represent instances in which intra-molecular pairing between the self-complementary portions of each strand preceded inter-molecular pairing between the complementary strands of the loop. Complete pairing in the loop region is apparently obstructed by by steric constraints imposed by the prior pairing of the self-complementary regions.

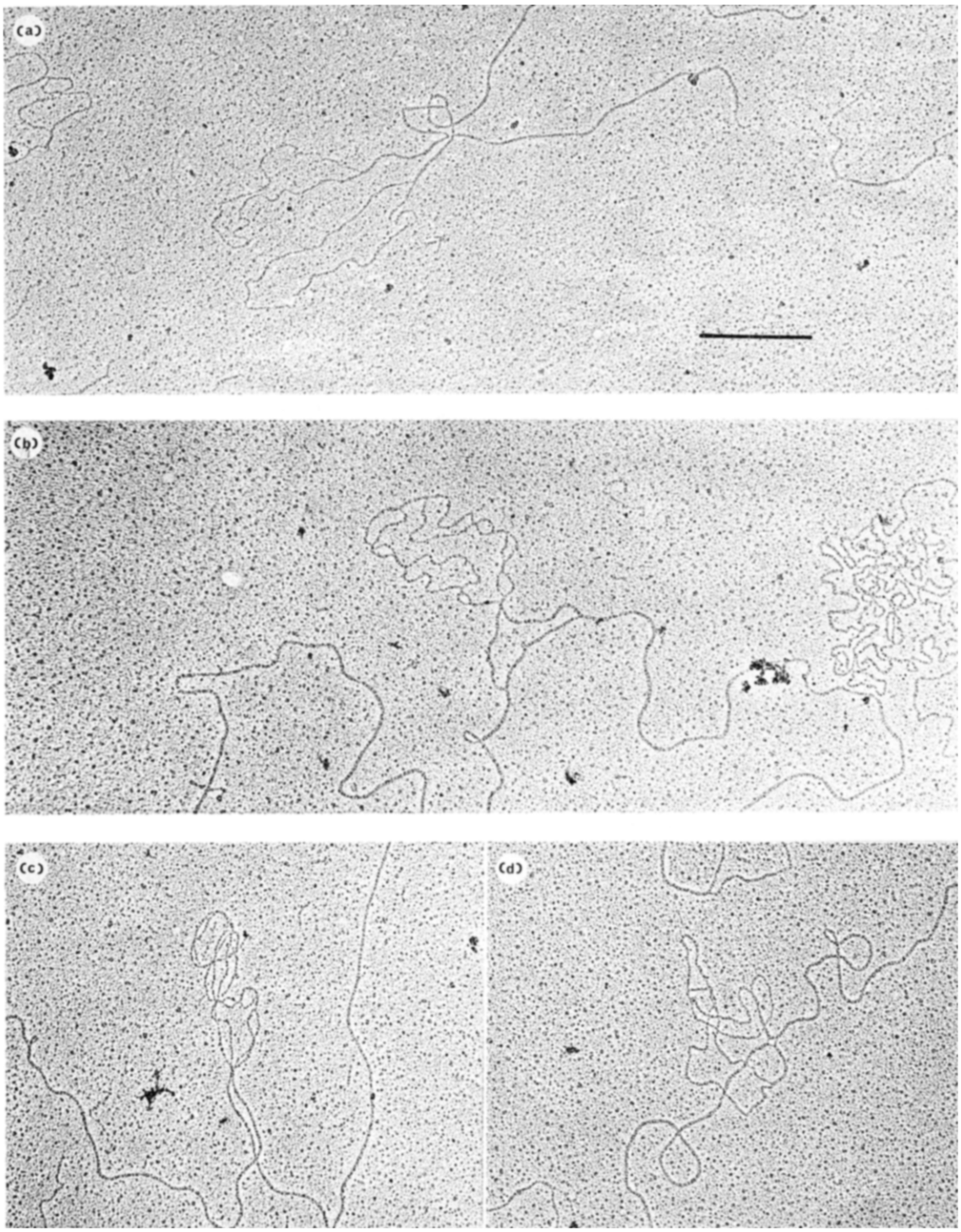


PLATE I. Heteroduplexes of Tc-10 and Tc-106.

(a) Tc-10/Tc-106 heteroduplex in which loops have not interacted.

(b) Tc-10/Tc-106 heteroduplex in which loops have interacted.

(c) Homoduplex (Tc-10/Tc-10 or Tc-106/Tc-106) in which loops have interacted.

(d) Homoduplex (Tc-10/Tc-10 or Tc-106/Tc-106) in which loops have not interacted but branch migration has occurred around the point of the insertions.

The bar represents 0·5 μm.

[*facing p. 564*

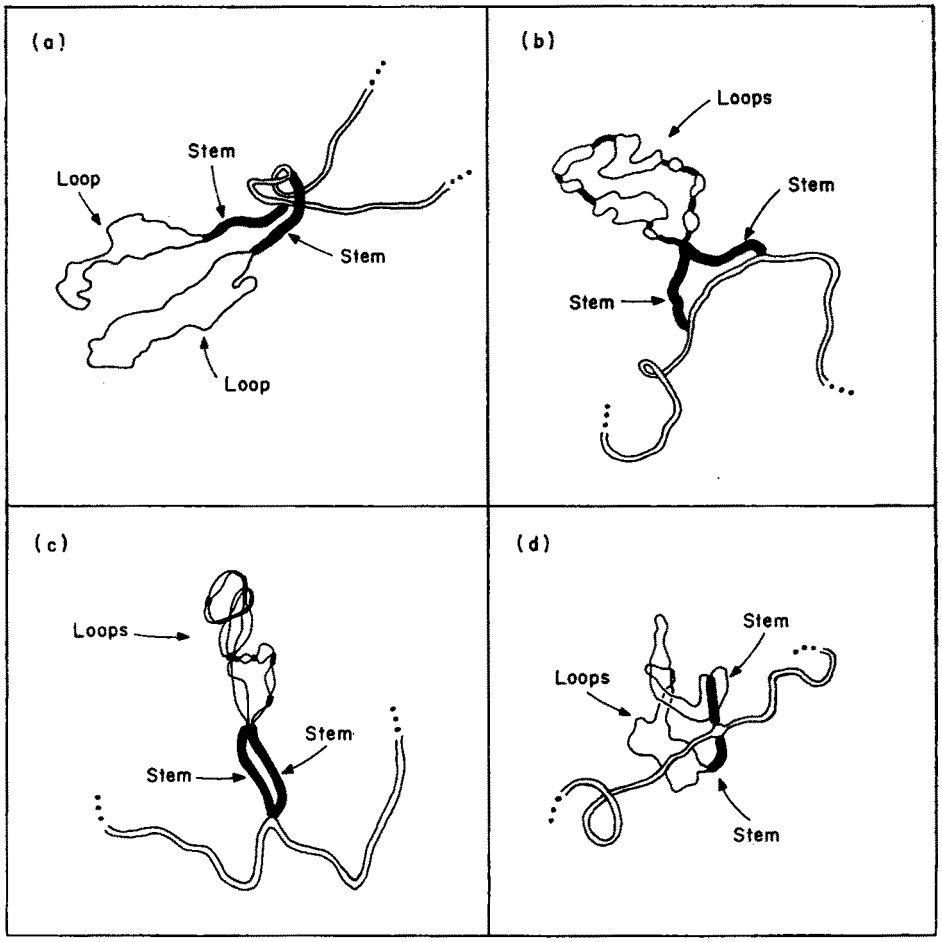


FIG. 1. Interpretation of heteroduplexes in Plate I.
(a) to (d) Correspond to micrographs (a) to (d) in Plate I. ▬▬, Paired stems (double-stranded); ▬▬, paired regions within the loop (double-stranded); ———, unpaired regions within the loop (single-stranded); ═══, P22 DNA (double-stranded).

Implicit in this interpretation is the idea that the insertions carried by P22Tc-10 and P22Tc-106 are not only homologous in base sequence but are also oriented in the same direction with respect to the P22 genome.

Alternative explanations for the structures in Plate I(b) involve partial homology within the loops. These alternative explanations are made less likely by the appearance of homoduplex molecules in the same experiment (Plate I(c)). In these structures, the stem and loop is seen emanating from the same point on the P22 genome; and here also the loops have interacted to give a partially paired structure whose pattern of paired and unpaired regions is very similar to that seen in heteroduplex molecules like the one shown in Plate I(b). Measurement of the lengths of the stems in these molecules indicates that "branch migration" (Lee *et al.*, 1970; Broker & Lehman, 1971) has occurred at the junction between the inserted DNA and the P22 DNA. For example, the lengths of both stems in Plate I(d) are about half the normal stem

length measured for other stems in the same field. Observation of branch migration (in about half of the twenty homoduplexes examined) verifies that the two insertions are actually located at the same point. The slightly separated structure (resembling a square), at the point of the insertion in Plate I(d), is typical of double-stranded branch migrations of this sort (Broker, personal communication).

In summary, P22Tc-10 and P22Tc-106 apparently contain identical insertions (unaccompanied by detectable deletion of P22 DNA) of foreign material at different locations on the P22 genome.

(b) *Mutagenesis of* Salmonella *by insertion of the* tet^R *element carried by Tc-10*

The above physical evidence suggests that the integrity of the *tet*^{R} insertion (bounded by its inverted repetition) has been maintained during translocation of the material from the R-factor onto the P22 genome. The genetic evidence which follows shows that this *tet*^{R} element is capable of further translocation *out* of the P22 genome and *into* many different locations on the chromosome of *S. typhimurium*.

When P22 Tc-10 infects *Salmonella* under conditions where the phage DNA is unable to maintain itself in the host cell, transductants carrying the *tet*^{R} determinant are still obtained. As shown in Table 1, *tet*^{R} transductants are obtained at low frequencies after infection even in the absence of phage (*erf*$^-$) and bacterial (*recA*$^-$) recombination functions, and even though the infecting phage is unable to integrate (*int*$^-$/*ataA*$^-$), or repress (*c2*^{ts}) (Watanabe *et al.*, 1972; Chan *et al.*, 1972; Tye *et al.*, 1974).

Since it seemed possible that these *tet*^{R} transductants had arisen simply by translocation of the *tet*^{R} element out of the P22 genome and into the *Salmonella* chromosome we sought to identify instances in which the insertion into the host chromosome had resulted in an identifiable mutation. We screened *tet*^{R} transductants (isolated on rich medium) for ones which had simultaneously acquired an auxotrophic mutation making them unable to grow on minimal medium. As shown in Table 1, roughly 1% of all *tet*^{R} transductants were auxotrophs. A wide variety of nutritional requirements is represented: out of 142 independent auxotrophs, 27 required proline, 18 methionine, 17 histidine, 12 purines, 8 isoleucine, 8 arginine, 7 tryptophan, 5 cysteine, 3 leucine,

TABLE 1

Isolation of tet^R *auxotrophs*

Expt	Phage	Bacterium	Frequency *tet*^{R} transductants per infecting phage particle	Percentage of auxotrophs among *tet*^{R} transductants
A	P22Tc-10*erf*$^-$	*recA*$^-$*ata*$^v_{P22}$	1×10^{-7}	~1 (14/~1500)
B	P22Tc-10*int*$^-$*c2ts*	*rec*$^+$*ata*$^+_{P22}$	not measured	2 (7/362)
C	P22Tc-10*int*$^-$*c2ts*	*rec*$^+$*ata*$^+_{P22}$	3×10^{-7}	1 (79/7892)

Bacteria were infected with derivatives of P22Tc-10 as indicated above and as described in Materials and Methods. *trt*^{R} transductants obtained from these infections were replica-plated directly onto minimal medium in order to identify auxotrophs. Candidate auxotrophs were then cloned and auxotrophy verified. In experiment C, 120 separate mixtures of phage and bacteria were made; transductants coming from different mixtures are assumed to be independent. In experiments B and C, different recipient strains were used (DB7000 and NK80, respectively).

3 thiamine, 2 alanine, 2 lysine, 1 phenylalanine, 1 tyrosine, 1 threonine, and 25 required nutrients as yet unidentified. This distribution, with its predominance of proline, methionine, purine, and histidine auxotrophs is similar to that obtained after standard chemical mutagenesis of *Salmonella*.

If these tet^R auxotrophs indeed represent simple insertion of the tet^R element alone into the *Salmonella* chromosome, these transductants should not carry an P22 genes. 18 of these tet^R auxotrophs have been tested by marker rescue for the presence of alleles in any of 15 different P22 genes spanning the known genetic map; no rescue was observed.

Two types of experiments provide direct evidence that in these tet^R auxotrophs the tet^R element is intimately associated, both physically and functionally, with the simultaneously acquired auxotrophic mutations:

(1) When the tet^R determinant from these auxotrophs is transferred to another *Salmonella* strain by P22-mediated generalized transduction, all of the recipient bacteria which have become tet^R have also acquired the corresponding auxotrophic requirement (Table 2). Absolute co-transduction of the auxotrophy with the tet^R suggests that the two determinants are physically very closely linked.

(2) A large number of independent tet^R auxotrophs carrying a wide variety of auxotrophic mutations have been reverted to prototrophy. Revertants were obtained at frequencies of 10^{-9} to 3×10^{-7}, depending upon the strain. For 33/37 auxotrophs tested, reversion to prototrophy was always accompanied by loss of the tet^R determinant (Table 3). Thus, not only have these tet^R transductants simultaneously acquired a new auxotrophic mutation, but revertants of the new mutations have simultaneously lost tet^R.

These data strongly suggest that the acquisition of tetracycline resistance and of auxotrophy are the consequence of the same event, namely, the insertion of the tet^R element into the affected gene. The wide distribution of auxotrophic requirements

TABLE 2

Co-transduction of auxotrophy with tetracycline resistance

Donor strain (tet^R auxotroph)	Donor auxotrophy	Proportion of tet^R transductants acquiring donor auxotrophy
NK144	*leu*	73/73
NK147	*leu*	82/82
NK120	*his*	64/64
NK127	*his*	82/82
NK219	*his*	71/71
NK231	*his*	78/78
NK114	*gnd*†	73/73
	Total	523/523

Seven independent tet^R auxotrophs were used as donors in P22-mediated generalized transduction experiments. Lysates of P22int_3HT12/4 were grown on each donor and used to transduce recipient strain NK80 ($edd^-leu^+his^+gnd^+$) to tetracycline resistance on complete medium. tet^R transductants were then tested for growth on appropriately supplemented minimal plates to determine how many had also acquired the donor auxotrophy.

† *gnd* = gluconate dehydrogenase; in the presence of an edd^- mutation, a gnd^- mutation renders *Salmonella* unable to use gluconate as a carbon source.

TABLE 3

Reversion of tet^R *auxotrophs to prototrophy*

	Expt A (*recA*$^-$)	Expt B (*rec*$^+$)	Expt C (*rec*$^+$)	Total
(a) Number of auxotrophs tested for reversion	10	6	34	50
Number yielding any revertants	6	5	34	45
(b) Tetracycline-resistance phenotypes of revertants				
Number of auxotrophs yielding:				
only *tet*^{S} revertants	6	5	22	33
only *tet*^{R} revertants	0	0	2	2
tet^R and *tet*^{S} revertants	0	0	2	2
Total	6	5	26	37

(a) Several colonies of each 50 *tet*^{R} auxotrophic strains were individually inoculated in LB broth, and grown to saturation; 0·1-ml samples were then spread directly on minimal plates. Revertants were obtained at frequencies of 3×10^{-9} to 5×10^{-7}, depending on the particular auxotrophic strain. No difference was seen in the frequencies or types of revertants obtained at temperatures from 28°C to 41°C.

(b) Many revertants were then directly tested for presence of the *tet*^{R} determinant (on minimal plates + 25 μg tetracycline/ml). A total of 397 revertants from 89 independent clones of 37 different auxotrophs were tested.

obtained means, therefore, that the *tet*^{R} element can insert into a large number of different sites on the *Salmonella* chromosome. The observation that nearly all of the *tet*^{R} auxotrophs can revert to prototrophy further suggests that, in most cases, insertion of the *tet*^{R} element is not accompanied by loss of genetic information on the bacterial chromosome. It seems most likely that the host nucleotide sequences on either side of the insertion are preserved exactly; although conceivably insertion could sometimes produce small alterations which do not preclude subsequent reversion to prototrophy.

(c) *Polarity of* tet^R *insertions in the* Salmonella *histidine operon*

In order to characterize some of the presumed *tet*^{R} insertions with respect to their precise locations and to ascertain their effects on gene expression, we chose to examine closely 16 independent *his*$^-$ auxotrophs isolated in our experiments. The *his* operon of *Salmonella* is convenient because it has been extensively characterized both genetically and biochemically (Hartman *et al.*, 1971; Brenner & Ames, 1971); the only genes known to be required for biosynthesis of histidine are the nine structural genes which comprise this operon.

The *his*$^-$ auxotrophs were first tested for their ability to complement a series of F′ factors (which carry heterologous *his* genes from *E. coli*), each of which was mutant in one of the *his* structural genes. This intergeneric complementation system, as described by Atkins & Loper (1970) allows the assessment of complementation in the virtual absence of recombination. Table 4 shows that the 16 histidine auxotrophs fall into only four categories on the basis of their complementation patterns. In three of these categories, the auxotrophs exhibit pleiotropic defects: they fail to complement with episomes carrying mutations in two or three contiguous genes in the operon.

TABLE 4

Complementation patterns of his$^-$ tet^R *auxotrophs*

		his$^-$ mutation on episome†									
Promoters:‡		G$^-$ →···	D$^-$	C$^-$	B$^-$ →···	H$^-$	A$^-$	F$^-$	I$^-$ →···	E$^-$	Inferred location of lesion
his$^-$*tet*^{R} auxotrophs											
Class	Number of strains										
1	8	−	−	−	+	N.T.	+	+	+	+	*G* (or promoter)
2	3	+	−	−	+	N.T.	+	+	+	+	*D*
3	1	+	+	+	+	N.T.	+	+	−	−	*I*
4	4	+	+	+	+	N.T.	−/+	−/+	+	+	*H* or *A*?

16 independent *his*$^-$*tet*^{R} auxotrophs fell into only 4 classes on the basis of their ability to complement various *his*$^-$ mutations carried on *E. coli* F′*his* episomes (see Materials and Methods). The locations of *his*$^-$ lesions are inferred from the observed complementation patterns, the position of the promoters in the *his* operon, and the assumption that the lesions are polar. N.T. = not tested.

†*His*$^-$ mutations are listed in the same order as the corresponding genes in the *his* operon. In this orientation, *his* operator/promoter region is just to the left of gene *G*.

‡ Positions of arrows indicate sites of primary promoter and two low-level internal promoters.

Auxotrophs in the fourth category (class 4) complement all *his*⁻ mutants tested, but complementation for genes *A* and *F* was quite poor. Complementation using episomes mutant in gene *H* was not done.

Since all of these *his*⁻ auxotrophs revert to prototrophy (data not shown), the pleiotropic effects of the *tet*^{R} insertions are not attributable to a deletion. Using the same intergeneric complementation system, Atkins & Loper (1970) identified two internal secondary promoters in the *his* operon at the positions shown in Table 4. One of these promoters is in gene *C* and promotes transcription of distal genes, while the other is farther "downstream." The observation that *tet*^{R} *his*⁻ auxotrophs of classes 1 and 2 are $G^-D^-C^-$ and $G^+D^-C^-$, respectively, but are still capable of expressing genes distal to gene *C*, suggested that the *tet*^{R} insertion mutation is polar on distal genes and that this polarity extends only as far as the next downstream promoter. The complementation pattern of the class 3 auxotroph is consistent with such an hypothesis, as is that of the class 4 auxotrophs if the partial complementation seen for mutants in *A* and *F* is attributed to weak polarity.

If the polarity hypothesis is correct, each of the *tet*^{R} insertions should map within the most operator–proximal of the genes it affects. Thus, class 1 mutants should map in the operator/promoter or in gene *G*; class 2 mutants in gene *D*; and the class 2 mutant in gene *I*. Deletion mapping (shown in Fig 2) places unambiguously all of these *tet*^{R} auxotrophs in the expected genes.

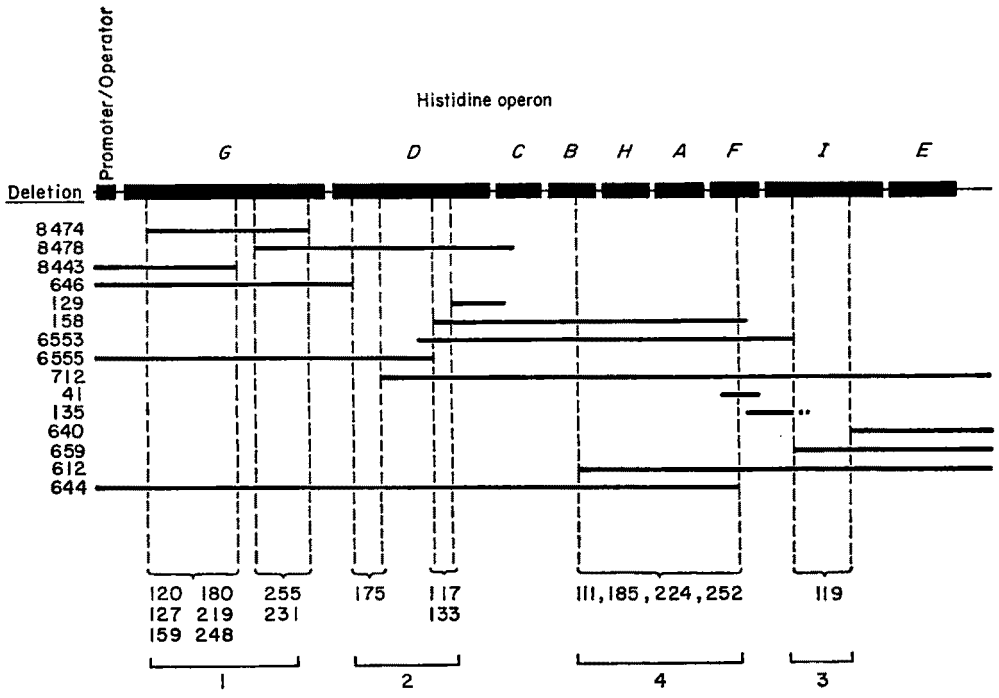


FIG. 2. Deletion mapping of *his*::*tet*^{R} auxotrophs.

Positions of 16 *his*::*tet*^{R} auxotrophs with respect to end points of known deletions in the histidine operon were determined by P22-mediated transductional crosses. Appearance of *his*$^+$ recombinants was scored after infection of appropriate deletion strains with transducing lysates grown on each of the *his*::*tet*^{R} auxotrophs. Horizontal bars indicate extent of *his* material deleted. Map is not drawn to scale. Bracketed numbers denote *his*::*tet*^{R} strains mapping in the indicated interval. Complementation classes 1 to 4 (from Table 4) also indicated.

The locations of *tet*^{R} auxotrophs in classes 1 and 2 have been confirmed by an independent method. The final step in the pathway for histidine biosynthesis, the conversion of histidinol to histidine, is catalyzed by the product of the *D* gene, histidinol dehydrogenase. Bacteria able to express *D* function can grow on minimal medium supplemented with histidinol even if they are defective in any other *his* function. For *tet*^{R} auxotrophs located in gene *D* itself, all revertants selected on histidinol should also be *his*$^+$. For *tet*^{R} auxotrophs located nearer to the promoter (i.e. "upstream"), it should be possible to obtain revertants which grow on histidinol in which the polar effect of the *tet*^{R} insertion on the expression of *D* function has been alleviated, but total operon function has not been restored and the "revertants" remain *his*$^-$.

Revertants of class 2 auxotrophs (mapping in *D*) which are able to grow on histidinol arise at frequencies of 10^{-7} to 10^{-9}. All such revertants tested (a total of 459 revertants picked from 7 to 10 clones of each of the three class 2 strains) had also become *his*$^+$, confirming that these *tet*^{R} insertions are in fact located in gene *D*. Revertants of class 1 auxotrophs (mapping in *G*) which grow on histidinol occur at frequencies of up to 10^{-3} and fewer than 1% of these revertants have become *his*$^+$, confirming that these insertions do *not* lie in gene *D*.

All of the above observations are consistent with the notion that the *tet*^{R} insertions in the *his* operon exert polar effects on the expression of distal genes.

All of the revertants of class 1 auxotrophs (in *G*) which were selected for ability to grow on histidinol had simultaneously lost the *tet*^{R} determinant. The vast majority of these revertants had not regained full *his*$^+$ function, and thus must represent some imperfect and/or partial excision of the *tet*^{R} element. Thus, *excision* of the *tet*^{R} element is usually not precise. This is in contrast to *insertion* of the element which is seen to be precise enough to allow restoration of gene function (as discussed above).

4. Discussion

The above experiments substantiate the idea that a genetic element which carries genes conferring tetracycline-resistance (flanked by an inverted repetition about 1400 base-pairs long) is capable of translocation as a unit from one DNA molecule to another. The *tet*^{R} element is found in nature on a variety of R-factors (Sharp *et al.*, 1973), was acquired by P22 (producing P22Tc-10 and P22Tc-106) and has now been observed to insert into a large number of different sites on the *Salmonella* chromosome, including six different sites in the *his* operon. Translocation of the *tet*^{R} element has the following genetic properties:

(1) When the *tet*^{R} element is inserted into a structural gene, gene function is abolished.

(2) When the *tet*^{R} element is inserted into a group of genes forming a single transcription unit (operon), it exerts a polar effect on the expression of promoter-distal genes.

(3) Insertion is relatively non-specific; the distribution on the *Salmonella* chromosome of *tet*^{R} insertion auxotrophs is similar to that obtained by chemical mutagenesis. On the other hand, there appears to be, within the *his* operon, some clustering of the sites of insertion.

(4) Insertion of the *tet*^{R} element virtually always occurs without loss of information from the molecule into which it inserts: most of the *tet*^{R} auxotrophs are capable of

reverting to prototrophy. Excision of the *tet*^{R} element, on the other hand, is usually not so precise: in the cases where we could test, there were hundreds of excisions which did *not* restore gene function for every excision that generated a prototrophic revertant.

(5) Insertion and excision of the *tet*^{R} element appears to be independent of the function of the *recA* gene of *Salmonella*.

Other laboratories have recently identified translocatable drug-resistance elements whose properties resemble those of the *tet*^{R} element. Heffron *et al.* (1975*a*,*b*) identified an ampicillin-resistance determinant which is flanked by a very short (150 base-pairs) inverted repetition. This element has been observed to translocate to at least a dozen sites within a 2800 base-pair segment of a small plasmid. Kopecko & Cohen (1975) report the *recA*-independent, apparently site-specific integration and translocation of a similar *amp*^{R} element. Most strikingly, Berg *et al.* (1975) describe two derivatives of coliphage λ which carry (as a simple insertion) a kanamycin-resistance element (derived from an R-factor) which is bounded by a 1400 base-pair inverted repetition. Thus, there appear to be many translocatable elements carrying drug-resistance determinants on R-factors which are associated with inverted repetitions.

However, Berg *et al.* also found derivatives of λ carrying a second type of kanamycin-resistance insertion (derived from a different R-factor) which do not have detectable inverted repetition; Gottesman & Rosner (1975) have also found a chloramphenicol-resistance determinant, which is translocatable into λ from coliphage P1, in which no reverse duplication was seen. These insertions may represent a different type of translocation element, or may be cases in which the inverted repetition is too short or too unstable to be detectable by standard visualization methods which depend on intra-molecular annealing.

In any case, it seems quite clear that translocatable genetic elements carrying drug-resistance determinants must make a substantial contribution to the reassortment of drug-resistant determinants which is seen among R-factors (Watanabe & Lyang, 1962; Clowes, 1972; Heffron *et al.*, 1975*a*).

In heteroduplex studies including the R-factor from which P22Tc-10 was made, Sharp *et al.* (1973) were able to correlate the *tet*^{R} determinant with a non-tandem inverted repetition having the same dimensions as the Tc-10 insertion. More recently, Ptashne & Cohen (1975) have shown that the inverted repetition associated with *tet*^{R} (on such an R-factor) is homologous with the IS3 insertion sequence of Malamy *et al.* (1972). The insertion sequences themselves were detected as individual insertion units which cause polar mutations (Shapiro, 1969; Jordan *et al.*, 1968; Malamy, 1970). In the cases of IS1 and IS2, polarity is attributable to the presence of transcription termination signals on the IS sequences themselves (Adhya *et al.*, 1974; Max Gottesman, personal communication). Thus, our *tet*^{R} element is bounded by two units which should be capable of functioning individually. This idea is supported by the preliminary observation that some of the *his*$^-$*tet*^{S} polarity relief revertants (obtained from class 1 *his*::*tet*^{R} insertions) can further revert to *his*$^+$ at high frequency. These could be instances in which excision of the *tet*^{R} determinant left behind a single IS sequence which can subsequently excise to restore *hisG*$^+$ function. (This explanation requires the additional assumption that the remaining IS sequence is non-polar or else can promote expression of *hisD* function; precedent exists for the idea that IS sequences might contain transcription termination signals when inserted in one

orientation, but serve as new promoters when inserted in the other (Saedler *et al.*, 1974).)

The IS sequences, singly and as non-inverted duplications, have recently been associated with a large number of unusual (often "illegitimate" and *recA*-independent) recombination phenomena—deletions, fusions, integration of F-factors, and amplification of R-factors (Sharp *et al.*, 1972; Ptashne & Cohen, 1975; Lee *et al.*, 1974; Davidson *et al.*, 1975; Hu *et al.*, 1975). The precise relationship of these phenomena to the properties of the inverted duplications of an IS3 sequence in the tet^R element remains to be worked out. We would like to suggest, based on our studies of the tet^R element, that a pair of inverted IS sequences might be essential for efficient and relatively stable translocation of the DNA between them to many new sites.

There are other bacterial systems for translocation of particular genetic elements which may or may not be mechanistically related to insertion of the tet^R element. The most striking example is bacteriophage Mu-1 which causes polar mutations by random insertion into the *E. coli* chromosome (for review, see Howe & Bade, 1975). Mutants of Mu which are capable of perfect excision much more frequently excise imperfectly (Bukhari, 1975), as is the case with the tet^R element. Coliphage λ is also capable of integration into a large number of sites (many of them structural genes) on the *E. coli* chromosome when the normal λ attachment site is missing (Shimada *et al.*, 1973). This integration is still dependent on the phage enzyme involved in site-specific integration at the normal site. Unlike Mu integration, however, integration of λ in the absence of the normal attachment site still shows considerable preference for particular sites, and excision of λ from these sites is usually precise (Shimada *et al.*, 1975). Both Mu and λ integration are, like that of the tet^R element, independent of *recA* function (Shimada *et al.*, 1975; Boram & Abelson, 1971).

Translocation of genetic material by such non-tandem inverted repetitions offers a powerful tool with which biological systems can perform substantial genetic rearrangements. Recombination between the inverted sequences themselves would generate an inversion of the material carried between them. Translocations by such elements could also result in duplications. Inverted repetitions (tandem and non-tandem) have been identified in the DNA of a large number of prokaryotic (Sharp *et al.*, 1972,1973; Berg *et al.*, 1975; Heffron *et al.*, 1975*a*,*b*; Daniell *et al.*, 1973) and eukaryotic (Wadsworth *et al.*, 1975; Wolfson & Dressler, 1972; Locker *et al.*, 1974; Manning *et al.*, 1975) organisms. In many of these cases, the inverted repetitions are associated with DNA rearrangements that can be explained in terms of translocatable segments. It seems possible therefore that a fundamental relationship may exist between translocatable segments of DNA and inverted repetitions.

We gratefully acknowledge the assistance of Elaine Lenk and the Massachusetts Institute of Technology, Department of Biology electron microscope facility; and the enthusiastic advice and ready access to strains from John Roth, Phil Anderson and John Scott. We also thank Norman Davidson, Tom Broker, Doug Berg, Fred Heffron, Stan Falkow, Lee Rosner and Stan Cohen for stimulating discussions and for access to unpublished results.

This work was supported in part by grant no. VC18D from the American Cancer Society, in part by grant no. GM21253 from the National Institutes of Health, and in part by grant no. GM18973 from the National Institutes of Health. One of us (D. B.) is supported by a Career Development award from the National Institutes of Health, no. GM70325.

Additionally, we thank Wanda Fischer for her help in preparing the manuscript.

REFERENCES

Adhya, S., Gottesman, M. & de Crombrugghe, B. (1974). *Proc. Nat Acad. Sci., U.S.A.* **71**, 2534–2538.

Atkins, J. F. & Loper, J. C. (1970). *Proc. Nat. Acad. Sci., U.S.A.* **65**, 925–932.

Berg, D. E., Davis, J., Allet, B. & Rochaix, J.-D. (1975). *Proc. Nat. Acad. Sci., U.S.A.* in the press.

Boram, W. & Abelson, J. (1971). *J. Mol. Biol.* **62**, 171–178.

Botstein, D. & Matz, M. J. (1970). *J. Mol. Biol.* **54**, 417–440.

Brenner, M. & Ames, B. N. (1971). In *Metabolic Regulation* (Vogel, H. J., ed.), pp. 349–387, Academic Press, New York.

Broker, T. R. & Lehman, I. R. (1971). *J. Mol. Biol.* **60**, 131–149.

Bukhari, A. (1975). *J. Mol. Biol.* **96**, 87–100.

Bukhari, A. & Metlay, M. (1973). *Virology*, **54**, 109–116.

Campbell, A. (1962). *Episomes Advan. Genet.* **11**, 101–145.

Chan, R. K. (1974). Ph.D. dissertation, Massachusetts Institute of Technology.

Chan, R. K. & Botstein, D. (1972). *Virology*, **49**, 257–267.

Chan, R. K. & Botstein, D. (1975). *Genetics*, in the press.

Chan, R. K., Botstein, D., Watanabe, T. & Ogata, Y. (1972). *Virology*, **60**, 833–898.

Clowes, R. C. (1972). *Bacteriol. Rev.* **36**, 361–405.

Daniell, E., Abelson, J., Kim, J. S. & Davidson, N. (1973). *Virology*, **51**, 237–239.

Davidson, N., Deonier, R. C., Hu, S. & Ohtsubo, E. (1975). *Microbiology*, **1**, in the press.

Garon, C. F., Berry, K. W. & Rose, J. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 2391.

Gottesman, M. M. & Rosner, J. L. (1975). *Proc. Nat. Acad. Sci., U.S.A.*, in the press.

Hartman, P. E., Hartman, Z., Stahl, R. C. & Ames, B. N. (1971). *Advan. Genet.* **16**, 1–34.

Heffron, D., Sublett, R., Hedges, R. W., Jacob, A. & Falkow, S. (1975*a*). *J. Bacteriol.* **122**, 250–256.

Heffron, F., Reubens, C. & Falkow, S. (1975*b*). *Proc. Nat. Acad. Sci., U.S.A.*, in the press.

Howe, M. M. & Bade, E. G. (1975). *Science*, in the press.

Hsu, M.-T. & Davidson, N. (1975). *Virology*, **58**, 229–239.

Hu, S., Ohtsubo, E., Davidson, N. & Saedler, H. (1975). *J. Bacteriol.* **102**, 764–775.

Jordan, E., Saedler, H. & Starlinger, P. (1968). *Mol. Gen. Genet.* **102**, 353–364.

Kopecko, D. J. & Cohen, S. N. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 1373–1377.

Lee, C. S., Davis, R. W. & Davidson, N. (1970). *J. Mol. Biol.* **48**, 1–22.

Lee, H. J., Ohtsubo, E., Deonier, K. & Davidson, N. (1974). *J. Mol. Biol.* **89**, 585–594.

Levine, M. & Smith, H. O. (1964). *Science*, **146**, 1581–1582.

Locker, J., Rabinowitz, M. & Getz, G. S. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 1366–1370.

Malamy, M. (1970). In *The Lac Operon* (Beckwith, J. R. & Zipser, D., eds), pp. 359–373, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Malamy, M. H., Fiandt, M. & Szybalski, W. (1972). *Mol. Gen. Genet.* **119**, 207–222.

Manning, J. E., Schmid, C. W. & Davidson, N. (1975). *Cell*, **4**, 141–155.

Miyake, T. & Demerec, M. (1960). *Genetics*, **45**, 755–762.

Ptashne, K. & Cohen, S. N. (1975). *J. Bacteriol.* **122**, 776–787.

Raj, A. S., Raj, A. Y. & Schmieger, H. (1974). *Mol. Gen. Genet.* **135**, 175–184.

Saedler, H., H., Reif, J., Hu, S. & Davidson, N. (1974). *Mol. Gen. Genet.* **132**, 265–289.

Schmieger, H. (1972). *Mol. Gen. Genet.* **119**, 75–88.

Scott, J. F. & Roth, J. R. (1975). *Proc. Nat. Acad. Sci., U.S.A.*, in the press.

Shapiro, J. A. (1969). *J. Mol. Biol.* **40**, 93–105.

Sharp, P. A., Hsu, M.-T., Ohtsubo, E. & Davidson, N. (1972). *J. Mol. Biol.* **71**, 471–497.

Sharp, P. A., Cohen, S. N. & Davidson, N. (1973). *J. Mol. Biol.* **75**, 235–255.

Shimada, K., Weisberg, R. A. & Gottesman, M. E. (1973). *J. Mol. Biol.* **80**, 297–314.

Shimada, K., Weisberg, R. A. & Gottesman, M. E. (1975). *J. Mol. Biol.* **93**, 415–430.

Smith, H. O. & Levine, M. (1964). *Virology*, **27**, 229–231.

Smith, H. O. & Levine, M. (1967). *Virology*, **31**, 207–216.

Stahl, M. M. & Stahl, F. W. (1971). In *The Bacteriophage Lambda* (Hershey, A. D., ed.), pp. 431–442, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Susskind, M. M., Botstein, D. & Wright, A. (1974). *Virology*, **62**, 350–366.

Tye, B. K., Chan, R. K. & Botstein, D. (1974). *J. Mol. Biol.* **85**, 485–500.
Watanabe, T. & Lyang, K. W. (1962). *J. Bacteriol.* **84**, 422–430.
Watanabe, T., Ogata, Y., Chan, R. K. & Botstein, D. (1972). *Virology*, **50**, 874–882.
Wolfson, J. & Dressler, D. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3054–3057.
Zinder, N. & Lederberg, J. (1952). *J. Bacteriol.* **64**, 679–699.

STUDIES ON THE VIRULENCE OF BACTERIOPHAGE-INFECTED STRAINS OF CORYNEBACTERIUM DIPHTHERIAE[1]

VICTOR J. FREEMAN

Department of Public Health and Preventive Medicine, University of Washington, School of Medicine, Seattle, Washington

Received for publication February 26, 1951

The relationship of naturally occurring avirulent strains to virulent strains of *Corynebacterium diphtheriae* is an unanswered question in the epidemiology of diphtheria and in the evolution of the diphtheria bacillus. The detailed investigations reported here have revealed that avirulent strains of *C. diphtheriae* infected with bacteriophage have yielded virulent *C. diphtheriae strains.*

MATERIALS AND METHODS

Cultures. The avirulent cultures[2] of *Corynebacterium diphtheriae* used in the following experiments were part of a shipment of field cultures received from the Division of Laboratories of the California State Department of Public Health in 1949. The subcultures used in these experiments have been kept in the lyophilized state since receipt. Cultures no. 770 and no. 1180 were two different isolations from a single diphtheria contact. Cultures no. 1174, 411, and 444 represented three separate isolations from a single case of diphtheria. All five cultures were designated *mitis* and avirulent by the Division of Laboratories of the California State Department of Public Health. Epidemiologically, there was no known relationship between the two individuals from whom the five cultures were obtained.

To be certain that no change in virulence had occurred during transfer to and from the lyophilized state, all of these strains were tested both intradermally and subcutaneously in guinea pigs, and all were found to be avirulent. As a further check, the five avirulent strains were submitted to the routine virulence-testing procedures of two additional laboratories, and all results were reported as negative.[3] *In vitro* plate toxigenicity[4] tests conducted repeatedly on these cultures also were negative. Before use in final experiments, each avirulent strain was subjected to four successive single colony isolations.

Bacteriophages. The bacteriophages used in these experiments were designated

[1] Supported in part by a grant from the Division of Research Grants and Fellowships of the National Institutes of Health, United States Public Health Service.

[2] The criteria for "avirulent" strains are typical morphologic and cultural characteristics of *C. diphtheriae* without the property of toxigenicity.

[3] These tests were carried out through the courtesy of Miss Marie Mulhern, Director, Public Health Laboratory, Seattle-King County Health Department, and Miss Donna Kerr, Assistant Director, Division of Laboratories, British Columbia Department of Health and Welfare.

[4] The terms "virulence" and "toxigenicity" are used interchangeably throughout this paper.

arbitrarily as A and B, with no intention that such designations be regarded as nomenclature for phages of *C. diphtheriae*. Phage A was received from Dr. Allen Ferris of Melbourne, Australia. It is a weak phage that produces only partial clearing of susceptible cultures on agar media (figure 1). Phage B (figure 2) was received from Miss Sheila Toshach at the School of Hygiene, Toronto, Canada. The latter phage is more active than phage A, producing glass-clear lytic areas on agar media (figure 1). Lysis of the avirulent strains could be demonstrated in liquid media only with phage B. Both phages were received originally as filtrates of lysed virulent cultures of *C. diphtheriae*.

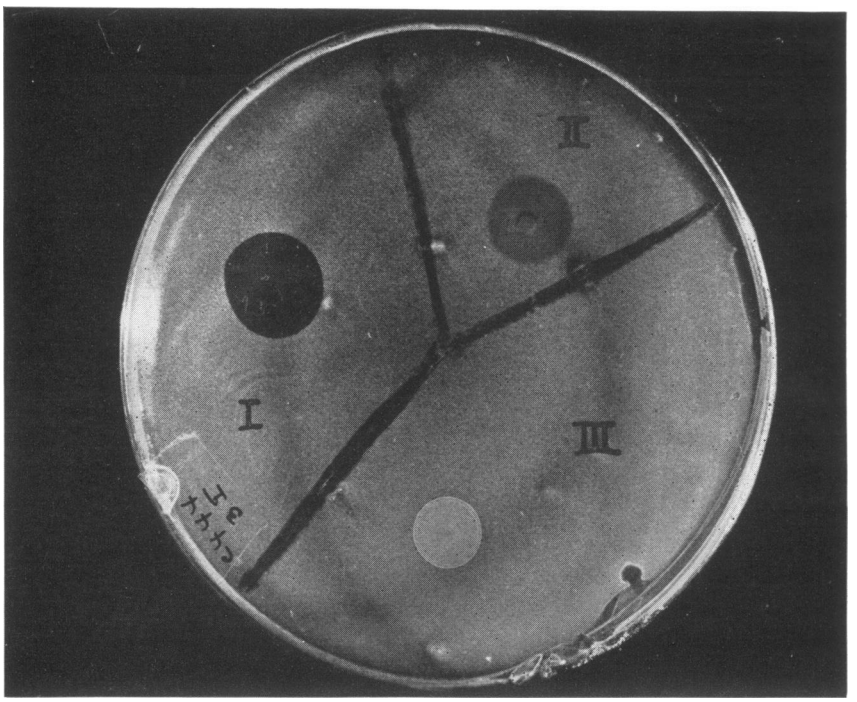


Figure 1. Bacteriophage lysis of an avirulent strain of *C. diphtheriae* (no. 444) growing on infusion agar. I, lysis by phage B. II, lysis by phage A. III, lysogenesis demonstrated by a strain newly rendered virulent.

The phages were prepared for use by making three successive single plaque isolations from agar plate cultures. The first plate was inoculated with a mixture of strain no. 444 and one of the bacteria-free phage filtrates. The resulting "adapted" phages were then produced in quantity on solid media, after which they were passed through Seitz EK filter pads. These filtered lysates were used undiluted in the subsequent experiments.

Media. Difco heart infusion broth was used for both liquid and solid media. The final pH of both broth and agar was 7.2 to 7.4. The agar medium for the *in vitro* virulence tests was prepared according to the recent modification of King *et al.* (1950). Agar plates made up from Mueller's serum tellurite medium (Difco)

were used for the isolation of cultures. Incubation of all media was carried out at 37 C.

Preparation of "lysates." The phage "lysates" used for testing toxigenicity contained whole bacterial cells as well as bacteriophage, extracellular products, and the products of cell lysis. These lysates were prepared in two different ways. In the first method, a solid medium was employed. Several agar plates were inoculated and spread with approximately 0.5 ml of an 18-hour broth suspension of the appropriate avirulent culture. The excess fluid was removed with a pipette. The inoculated plates were then dried at 37 C for approximately half an hour and were afterwards inoculated with 0.5 ml of the appropriate filtered bacteriophage

Figure 2. Electron micrograph of bacteriophage B. Preparation by replication method, followed by chromium shadowing. Magnification × 21,200. (This photomicrograph was prepared with the technical assistance of Mr. Oliver Rowe, electron microscopist, Engineering Experimental Station, University of Washington.)

lysate. Following removal of the excess fluid, the plates were incubated at 37 C overnight. The next morning the surface growth on the agar plates was washed off with 3 to 5 ml of 0.85 per cent saline. Suspensions prepared in this manner were tested by the injection of 0.1 ml intradermally and 1.0 ml subcutaneously into guinea pigs. Control suspensions were prepared in the same way except for the use of saline in place of the bacteriophage lysate.

In the second method, tubes containing 3 ml of broth were inoculated with 3 drops (approximately 0.15 ml) of an 18-hour culture and incubated until growth was just visible (usually about 1 hour). The tubes were then inoculated with 3 drops of the appropriate filtered bacteriophage lysate and returned to the incubator until the next morning. Control broth cultures were left in the incubator after the initial inoculation. Broth from the incubated culture tubes was tested

by the injection of 0.1 ml intradermally and 1.0 ml subcutaneously into guinea pigs. Lysis in broth occurred fairly consistently when the susceptible avirulent strains were mixed with phage B in the manner described. The lysis was incomplete, but a reduction in turbidity was readily evident on comparison with control cultures. Phage A failed to show any evidence of lysis of these strains growing in broth.

Virulence tests. Intradermal testing was done on 450- to 500-g guinea pigs, according to the method of Fraser (1937). Usually 8 to 12 tests were made on each animal. In all instances a naturally virulent strain of *C. diphtheriae* was included as a positive control. Subcutaneous tests were conducted in guinea pigs weighing 250 to 300 g by inoculating 1.0 ml of test fluid beneath the skin of the animal's flank. Care was taken to avoid the introduction of any fluid into the peritoneum. Since preliminary experiments did not reveal any difference in results when different brands of commercial antitoxin were used, the particular brand has not been specified in the account of results given below. Test animals in both intradermal and subcutaneous methods received 1,000 units of diphtheria antitoxin, the controls in the subcutaneous group receiving theirs 4 hours before, and all of the intradermal group 4 hours after, inoculation of the initial test lysates.

In vitro toxigenicity tests were carried out as previously described (Freeman, 1950), except for the modification in medium mentioned above. The *in vitro* plates were inoculated with the same test fluids as were used in the animal virulence tests, both series of tests being started on the same day. As with the *in vivo* tests, the brand of commerical antitoxin used has not been specified. Differences similar to those previously described for the production of secondary lines (Freeman, 1950) were observed in these experiments when the appropriate antitoxin was employed. A naturally toxigenic strain was included on all plates as a positive control.

Final experiments. In order to ensure that all possibility of contamination of the avirulent strains with virulent cultures of *C. diphtheriae* might be excluded, an entirely separate and previously unused laboratory was set up for the final experiments. In this final series, the same test sample harvested from solid medium was inoculated intradermally and subcutaneously in guinea pigs and on *in vitro* plates all within the same day. To further ensure comparability of results, the mixtures of avirulent culture with virulent culture filtrate were prepared in exactly the same manner as that already outlined for preparation of the culture-bacteriophage lysates.

RESULTS

Phage susceptibility of avirulent strains employed. Table 1 shows the susceptibility of the five avirulent strains of *Corynebacterium diphtheriae* to the phages A and B. The readings given are for visible lysis only. It is conceivable that some lysis might have occurred in the broth cultures to which phage A was added, but only quantitative analysis would reveal this. Control materials spotted on agar plates included the sterile broth medium and homologous culture grown in

broth. All such controls were negative for lysis. Culture 411 obviously is differentiated from the other four avirulent strains by reason of its phage resistance. This was the only characteristic observed that provided such differentiation. In all other respects, both cultural and morphologic, all five avirulent strains were indistinguishable.

TABLE 1

Bacteriophage susceptibility of avirulent strains of C. diphtheriae

STRAIN NO.	ON AGAR MEDIUM		IN BROTH	
	Phage A*	Phage B*	Phage A	Phage B
444	2+†	4+	0	1-2+
1174	2+	4+	0	1-2+
1180	2+	4+	0	1-2+
770	2+	4+	0	1-2+
411	0	0	0	—0

* Material spotted on plate previously inoculated with culture indicated in first column.

† Degree of clearing: 0 = none; 1+ = slight; 2+ = moderate; 3+ = marked; 4+ = complete.

TABLE 2

*Intradermal tests of bacteriophage lysates in guinea pigs**

STRAIN NO.	AVIRULENT CULTURE GROWN ON AGAR						AVIRULENT CULTURE GROWN IN BROTH					
	Control†		Phage A filtrate†		Phage B filtrate†		Control		Phage B added		Seitz filtrate of incubated culture phage B mixture	
	T	C	T	C	T	C	T	C	T	C	T	C
444	0	0	0	0	4+‡	0	0	0	4+	0	2+	0
1174	0	0	0	0	4+	0	0	0	4+	0	2+	0
1180	0	0	0	0	4+	0	0	0	4+	0	2+	0
770	0	0	0	0	4+	0	0	0	4+	0	2+	0
411	0	0	0	0	0	0	0	0	0	0	—	—

* The intradermal method used involves a second inoculation of the test substances 4 hours after the initial injections and immediately following an intraperitoneal injection of 1,000 units of diphtheria antitoxin. The test results are recorded under "T" and the control antitoxin results under "C."

† Material used for the second inoculation of the plates (details described under "Preparation of Lysates"); saline was used in the control cultures.

‡ The symbol 4+ indicates the presence of both erythema and necrosis involving an area greater than 1 cm^2; 2+ indicates a similar reaction, except for the absence of necrosis.

Demonstration of Toxigenicity of Lysates in Vivo

Intradermal tests. Table 2 summarizes the effects of the different culture lysates inoculated intradermally in guinea pigs. All positive reactions were similar and were indistinguishable from typically positive reactions caused by characteristically virulent strains of *C. diphtheriae*. From these results it can be seen that

a dermal necrotic factor was present in the phage B lysates of the avirulent strains. Even though phage A did show partial lysis on four of the avirulent strains, the phage A lysates failed to show any evidence of dermal toxicity. That the necrotic factor in the phage B lysates may have been true diphtheria toxin was suggested by the fact that diphtheria antitoxin invariably prevented the necrosis from occurring.

The broth lysates that caused dermal toxicity were retested after being passed through Seitz EK filter pads to determine whether the factor under study was extracellular. The reactions produced by the filtered lysates were reduced in intensity but were comparable to a control of filtered diphtheria toxin equally diluted.[5] The filtered lysates were actively lytic for the original susceptible strains, in addition to being toxic under the conditions described.

TABLE 3

Subcutaneous tests of bacteriophage lysates in guinea pigs*

STRAIN NO.	CULTURE PLUS SALINE	CULTURE PLUS PHAGE A	CULTURE PLUS PHAGE B	CULTURE PLUS PHAGE B AND ANTITOXIN
444	0/3†	0/1	4/4	0/2
1174	0/1	0/1	2/2	0/1
1180	0/1	0/1	2/2	0/1
770	0/1	0/1	2/2	0/1
411	0/1	0/1	0/1	0/1
Total.....	0/7	0/5	10/11	0/6

* All cultures and culture lysates were washed off agar media with 0.85 per cent saline and inoculated in 1.0-ml doses.

† The numerator represents the number of guinea pigs that died; the denominator, the total number tested.

The absence of toxicity manifested by strain 411 was recorded repeatedly. This negative result was regarded as a twofold control. First, it provided a limited control on the phage per se, for one could conclude that, at least in the quantity of the original inoculum, the phage was not toxic. Secondly, this result controlled the possibility that any "carry-over" toxin might have been responsible for the reactions. In other words, even though the phage B filtrates were known to be toxic (see last column, table 2), the degree of dilution involved in the preparation of the lysates was adequate to prevent necrosis occurring from any added toxin. It would seem extremely doubtful that the amount of toxin carried over from the original virulent culture filtrate could have maintained measurable potency through three successive single plaque isolations on avirulent cultures.

Subcutaneous tests. The subcutaneous tests recorded in table 3 confirm the results of the intradermal tests and in addition demonstrate that the toxin concerned had lethal as well as necrotic properties. All animals in each of the con-

[5] The diphtheria toxin was provided through the courtesy of Dr. W. E. Ward of Cutter Laboratories.

trol groups and in the group receiving phage A lysate survived and showed no evidence of ill effect from the inoculations other than slight induration, which disappeared within 48 hours. The only survivor of the unprotected phage B lysate group received a suspension of the one avirulent strain in this series (411) that was resistant to phage lysis. This animal also showed no significant reaction to the inoculation.

All 10 animals that succumbed to infection with the phage B lysate died within 48 hours. Autopsy findings in all instances revealed a moderate to marked gelatinous hemorrhagic exudate at the site of inoculation, markedly hemorrhagic adrenal glands, and distension of some or all of the following abdominal organs: stomach, gall bladder, small bowel, large bowel, and urinary bladder. Gelatinous exudate and hemorrhagic adrenal glands have been reported consistently as characteristic of diphtheria intoxication (Topley *et al.*, 1946). Isolation of cultures from these animals is described below.

An additional 300-g guinea pig was inoculated with 1 ml of undiluted phage B filtrate. Although local necrosis resulted, the animal survived without any permanent residual effects. As pointed out above, the phage B filtrates were, of necessity, toxic. Consequently, local necrosis was to be expected. But the fact that death did not result was regarded as further evidence that the phage itself was not the virulent factor.

The probability that the toxic factor (or factors) demonstrated in the above experiments was true diphtheria toxin was considered great. Both dermal necrotic and lethal reactions produced by the phage B lysates were indistinguishable from such reactions produced by naturally toxigenic strains. Also, in all instances where diphtheria antitoxin was employed, necrotic and lethal reactions were entirely prevented. In addition, the fact that filtered phage lysates were still toxic showed that, as with true diphtheria toxin, the lysate toxin was extracellular. Preliminary heat stability tests have shown that the lysate toxin, like diphtheria toxin (Smith and Martin, 1948), was thermolabile, being destroyed by heating to 56 C for 15 minutes.

Demonstration of Toxigenicity of Lysates in Vitro

Additional confirmation of the production of toxigenicity from avirulent strains of *C. diphtheriae* was provided by the *in vitro* plate tests. Such tests also provided a control against host factors that otherwise might have been difficult to exclude. As illustrated in figure 3, the toxin-antitoxin precipitate formed by the avirulent culture phage B lysate was more distinct than that formed by the control virulent culture. This has been a consistent observation with all the *in vitro* tests. The precipitate produced by the phage B lysate appeared earlier than that produced by the control virulent culture, sometimes being evident as early as 18 hours. It was produced with all brands of commercial antitoxin tested, a result that was found true only for virulent diphtheria cultures (Freeman, 1950). Secondary lines were produced by both control and test cultures when the appropriate antitoxin was employed, but here again these lines appeared earlier and were definitely more pronounced with the phage lysate than with either the avirulent or virulent culture controls.

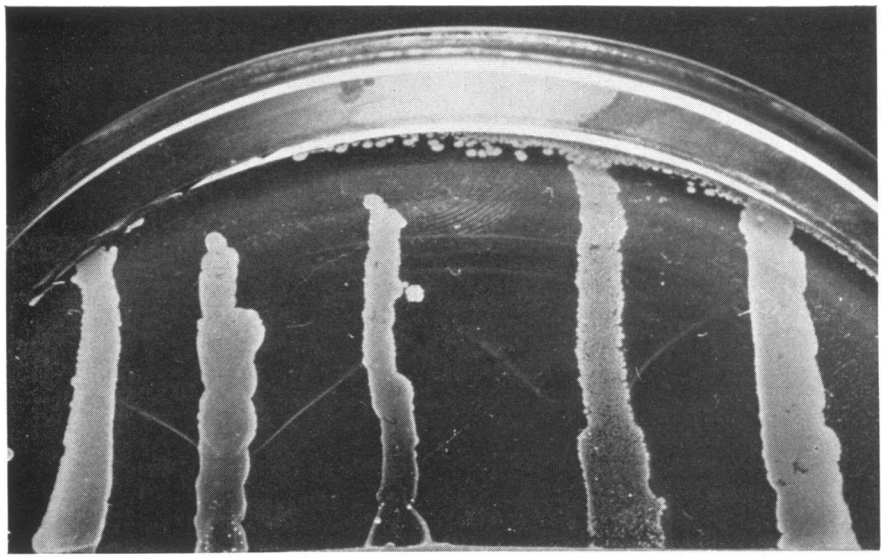


Figure 3. In vitro plate virulence test. Culture *A* shows a negative reaction. It is strain 444 mixed only with saline. Culture *B* is strain 444 mixed with phage B. The characteristic positive arrowhead precipitate is readily evident. *C* is strain 444 mixed with phage A, giving a negative reaction. *D* is a naturally virulent control strain showing a positive reaction. *E* is strain 444 mixed with a virulent culture filtrate, showing a negative reaction.

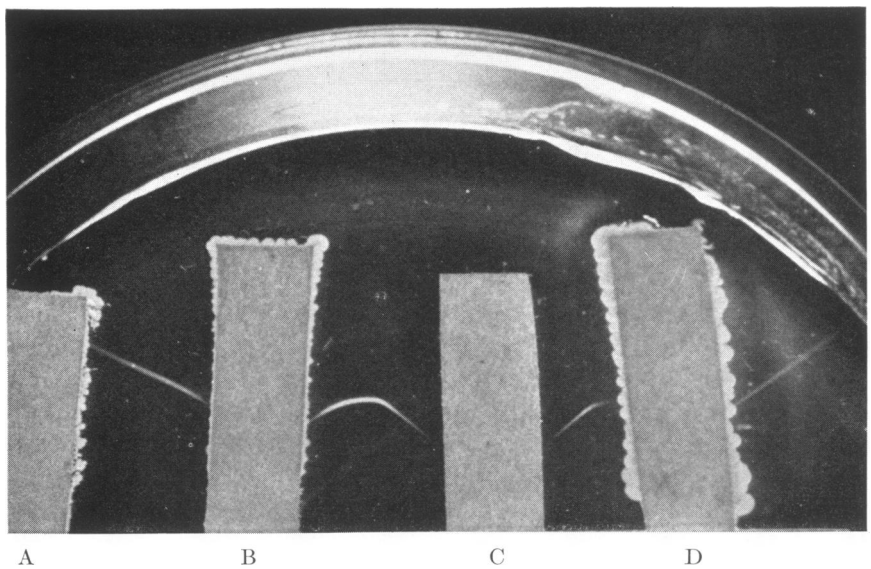


Figure 4. Double diffusion gradient analysis. Filter strip *A* was soaked in a broth culture of strain 444; *B* was soaked in a culture of strain 444 mixed with phage B; *C* was soaked in known diphtheria toxin; *D* was soaked in a culture of a control virulent strain. Note the fusion of the known diphtheria antitoxin precipitates with the toxins from the virulent culture and the avirulent culture-bacteriophage mixture.

Further evidence suggesting that the filtrable toxic factor produced by the avirulent culture phage B lysate was true diphtheria toxin was obtained by applying Elek's principle of double diffusion gradients (1949*a*) to the antigen-antibody system involved in the production of the precipitates just described. Figure 4 illustrates the results. The linear precipitates produced by the avirulent culture lysate, the diphtheria toxin, and the virulent diphtheria culture have fused with one another. Such fusion is indicative of the homologous nature of the toxic antigens diffusing out, since heterologous antigen-antibody precipitates cross one another instead of fusing (Elek, 1949*b*; Ouchterlony, 1949).

Properties of Cultures from Lysates and Autopsies

In vitro toxigenicity tests. These were carried out to assess the relative proportions of virulent and avirulent organisms in the cultures resulting from the mix-

TABLE 4

In vitro toxigenicity tests on single colonies from avirulent cultures and culture lysates

		NO. COLONIES TESTED	POSITIVE	NEGATIVE	PER CENT NEGATIVE
I	Avirulent cultures without phage B	107	0	107	100.0
II	Avirulent culture phage B lysates—isolations from animal autopsies	156	129	27	20.9
III	Avirulent culture phage B lysates—direct isolations without animal passage	61	53	8	13.1
IV	Subculture from one of the virulent colonies in group II	72	72	0	0.0

ture of avirulent strains with phage B, both before and after animal passage. Also tested were unlysed avirulent control cultures and the progeny of a known toxigenic single colony culture isolated from a guinea pig autopsy strain. *In vivo* checks on some of the *in vitro* tests always yielded confirming results. Cultures isolated from guinea pigs that died following inoculation with the phage B lysates were obtained from the gelatinous exudate and the peritoneal cavity. The majority of *C. diphtheriae* cultures were isolated from the gelatinous exudates, although an appreciable number of cultures isolated from the peritoneum also were positive for *C. diphtheriae* organisms.

Table 4 summarizes the results of the toxigenicity tests. It can be seen that in the culture lysates the majority of organisms isolated both before and after animal passage were toxigenic. No toxigenic strain has yet been isolated from the avirulent control cultures. Tests on 72 single colony subcultures, the progeny of a toxigenic single colony isolated from a guinea pig autopsy, were 100 per cent positive.

Phage susceptibility and lysogenicity of culture isolations. All virulent cultures isolated from lysates or autopsies were found to be resistant to lysis by phage B, but, like the avirulent cultures from which they originated, they showed partial lysis by phage A. When tested by the method of Fisk (1942) for lysogenicity, these same phage-B-resistant virulent strains all were found to carry a phage that, in its gross characteristics, could not be distinguished from phage B. Apparently, exposure of the phage-susceptible avirulent cultures to phage B resulted in the production of virulent lysogenic strains of *C. diphtheriae.*

A representative sampling of the 35 avirulent cultures isolated from the single colony subcultures of the phage B lysates invariably yielded strains that were still phage-B-susceptible and capable of producing toxigenic cultures on exposure to phage B. These avirulent strains were not lysogenic.

Stability of newly isolated virulent strains. Evidence of the stability of these newly isolated virulent strains was indicated by the fact that repeated stock culture passage did not reduce their virulence to any appreciable degree. Also, second and third animal passages with these "new" virulent strains in even smaller doses and under more rigid conditions (applied to a scratched skin surface after the method of Orskov, 1948) still caused rapid death with the characteristic post-mortem picture.

Preliminary immunity tests. Frobisher *et al.* (1947) have shown that animals inoculated with avirulent diphtheria cultures demonstrate some resistance to subsequent challenge with virulent strains. It was decided to test the 12 survivors listed in table 3 that had had avirulent culture mixed with saline or phage A. A week following the initial inoculations, the animals were given second inoculations of the same mixtures. Two weeks after the second inoculations, the animals were challenged with 0.5-ml doses of broth suspensions of newly isolated virulent strains inoculated subcutaneously. All animals died within 48 hours and showed on post-mortem examination the characteristic picture of diphtheria intoxication referred to above.

Six additional guinea pigs, one that had received two doses of 1 ml each of diphtheria fluid toxoid (Cutter), one that received heart infusion broth in the same doses, two that received 1,000 units each of diphtheria antitoxin, and two uninoculated, were challenged with one of the newly isolated virulent cultures applied according to the Orskov scratch method. The animal that had received the diphtheria toxoid and those that received the antitoxin survived without ill effect. The two uninoculated animals and the guinea pig inoculated with the broth died within 48 hours, showing the characteristic picture of diphtheria intoxication.

Effect of Virulent Culture Filtrate on Avirulent Strains

In order to determine whether all naturally virulent strains of *C. diphtheriae* possibly were associated with a filtrable factor capable of producing virulence in previously avirulent strains, the following experiment was conducted. A naturally virulent culture of *C. diphtheriae* growing in broth was passed through a Seitz EK filter pad, and the filtrate was handled in the same manner as described

above under "Preparation of Lysates," except that the virulent culture filtrate was substituted for the phage filtrate. Mixtures were prepared for the five avirulent strains and tested for toxigenicity both *in vitro* and *in vivo* at the same time as the lysate tests recorded above, and on the same plates and animals. All such tests were negative for toxigenicity. These results would tend to rule against any extracellular product of virulent organisms as the agent involved in rendering avirulent strains virulent and suggest that the phage per se or phage by-products were implicated.

It would have been desirable to test the filtrate of a virulent strain that had been lysed by phage B and rendered free of phage as a control on phage by-products. But the difficulty of obtaining such a phage-free filtrate that has not at the same time been divested of possibly significant phage extracellular factors is obviously great.

DISCUSSION

In 1917 Rosenau and Bailey reported on their investigations into the problem of avirulent diphtheria bacilli and their relation to diphtheria immunity. They concluded that the avirulent bacilli constituted a group so different from the typical Klebs-Loeffler organism as to throw doubt on the appropriateness of the name "avirulent diphtheria bacilli." In 1929 Okell noted that practically all observers agreed that avirulent diphtheria bacilli could never become virulent either *in vivo* or *in vitro*. On the basis of serologic investigations, Okell concluded that there seemed no reason to modify the prevailing view that the avirulent diphtheria bacillus was not even potentially a cause of disease. On the other hand, Frobisher and his group have emphasized the probable importance of nontoxigenic diphtheria strains in naturally acquired diphtheria immunity. They observed, not infrequently, the isolation of totally avirulent strains from undoubted cases of clinical diphtheria, and their experiments have been designed to investigate what they describe as the "subtle, unexplained, pathogenic power" of these nontoxigenic strains (Frobisher, Parsons, and Updyke, 1947). The material reported herein would appear to support the Frobisher group.

The question of the mechanism by which the transition of avirulent to virulent diphtheria organisms occurs is one that will require careful and extensive investigation. The possibility that a mixture of avirulent organisms with a very few virulent cells might occur either as a natural mixture from the nasopharynx or as the result of accidental contamination has been considered. An extremely small proportion of virulent organisms could be mixed with the avirulent cells, and still it might be conceivable that the whole culture would be regarded as avirulent. Since phage B lyses the avirulent but not the virulent form, a mixture of such a culture with the phage would allow the virulent strain to predominate. However, the fact that the avirulent cultures used in all of the reported experiments had been subjected to four successive single colony isolations, and that subsequent subcultures from these "purified" strains in every one of 107 single colony isolations failed to yield a virulent culture, would argue against the possibility of initial mixture with virulent cells. It would be desirable to re-

peat some of the foregoing experiments with avirulent cultures that had been prepared from single cell isolations.

If one assumes that the avirulent cultures were absolutely pure, and if one recognizes the results recorded above concerning the failure of a virulent culture filtrate to elicit toxigenicity in avirulent strains, then the most likely hypothesis to explain the phenomenon of virulence conversion is the spontaneous development of toxigenic mutants, with selection by phage lysis. The finding that all the newly isolated virulent strains tested for phage B susceptibility were resistant to lysis by phage B is consistent with such a hypothesis. The inability of phage A lysates to demonstrate toxigenicity might have been due to quantitative limitations, since lysis of the avirulent cultures was only partial, and therefore they may still have overgrown the virulent mutants.

The theory of mutation in the study of *Corynebacterium diphtheriae* is not a new concept. Crowell (1926) obtained an avirulent culture by a single cell isolation from a virulent parent strain that had been cultured from a single cell. He attempted in a great variety of ways to cause the avirulent strain to revert to its toxigenic form, but without success. As a consequence, Crowell concluded that the mutation was irreversible. There are many other instances cited in the literature of the isolation of avirulent strains from virulent parent cultures, but the reverse situation apparently has not been described. Variation from the virulent to the "atoxic" form of *C. diphtheriae* in the presence of bacteriophage has been reported by Blair (1924) and Stone and Hobby (1934).

The release by phage lysis of preformed toxin or of endotoxin that conceivably might be oxidized to toxin (Frobisher *et al.*, 1947) is a hypothesis that cannot be excluded at present, but on the basis of preliminary work it is considered less likely. If the inability of phage A lysates to demonstrate toxigenicity is qualitative rather than quantitative, when compared with the action of phage B, then the possibility that the complex phage-bacterium relationship per se is involved must be considered. Investigations concerning these various hypotheses are in progress.

Regardless of the fact that the underlying mechanism is not understood, the knowledge that avirulent cultures of *C. diphtheriae* can become virulent in the presence of specific bacteriophage is of importance to any consideration of the many perplexing problems that have confronted bacteriologists and epidemiologists interested in the study of diphtheria. If the virulence of the diphtheria bacillus should prove dependent not only on its toxigenic ability and its invasive power but also on the degree of its association with a specific bacteriophage, then some of the difficulties involved in understanding the complex problems of bacterial metabolism and immunity as they occur in the diphtheria case or carrier might be partially solved.

ACKNOWLEDGMENTS

The author is indebted to Drs. C. A. Evans, H. C. Douglas, A. S. Lazarus, R. S. Weiser, and N. Groman for their valuable criticism, and to Miss Una Morse for her technical assistance. Use of the electron microscope was arranged through

the courtesy of Professor F. Burt Farquharson, Director, Engineering Experimental Station, University of Washington. Photographs of agar plate cultures were prepared by the Department of Medical Photography, University of Washington.

SUMMARY

Virulent strains of *Corynebacterium diphtheriae* were isolated from four of five avirulent cultures of *C. diphtheriae* when the four cultures were incubated with a specific bacteriophage B filtrate by which they were readily lysed. No virulent strains could be isolated when these same cultures were incubated with a phage A filtrate by which they were only partially lysed. The fifth avirulent culture was not susceptible to either phage and yielded no virulent strains when incubated with either phage filtrate.

Virulence (toxigenicity) tests were carried out *in vivo* by both intradermal and subcutaneous methods and *in vitro* by the plate technique, complete correlation being obtained by all methods. Controls of cultures incubated with saline were negative in all instances when tested for virulence.

The extracellular nature of the toxic factor produced by the newly isolated virulent strains was demonstrated by filtration through a Seitz EK filter pad. Occurrence of the toxic reactions in guinea pigs inoculated with these virulent strains was shown to be preventable by the administration of diphtheria antitoxin.

Tests on single colonies isolated from avirulent cultures and culture lysates revealed that 100 per cent of colonies isolated from the avirulent cultures without phage were avirulent, whereas over 80 per cent of colonies from the avirulent culture phage B lysates were virulent.

The virulent strains isolated from the avirulent culture lysates were found to be stable. Also they were phage-B-resistant and lysogenic. The phage they were found to carry could not be distinguished from phage B.

A filtrate of a naturally virulent strain of *C. diphtheriae* was tested on the avirulent cultures in a manner identical with that of the phage filtrates. No virulent strains were isolated by this method.

It is suggested that an initial mixture of virulent with avirulent cells is unlikely, and therefore that the most likely hypothesis is the spontaneous development of toxigenic mutants with selection by phage lysis.

REFERENCES

BLAIR, J. E. 1924 A lytic principle (bacteriophage) for *Corynebacterium diphtheriae*. J. Infectious Diseases, **35**, 401–406.

CROWELL, M. J. 1926 Morphological and physiological variations in the descendants of a single diphtheria bacillus. J. Bact., **11**, 65–74.

ELEK, S. D. 1949*a* The serological analysis of mixed flocculating systems by means of diffusion gradients. Brit. J. Exptl. Path., **30**, 484–500.

ELEK, S. D. 1949*b* The plate virulence test for diphtheria. J. Clin. Path., **2**, 250–258.

FISK, R. T. 1942 Studies on staphylococci. J. Infectious Diseases, **71**, 153–165.

FRASER, D. T. 1937 The intracutaneous virulence test for *Corynebacterium diphtheriae*. Am. J. Pub. Health, **27**, Supp. to no. 3, 121–124.

FREEMAN, V. J. 1950 Influence of type and concentration of antitoxin on the *in vitro* toxigenicity test for *C. diphtheriae*. Public Health Rep., **65**, 875–882.

FROBISHER, M., JR., PARSONS, E. I., AND UPDYKE, E. 1947 The correlation of laboratory and clinical evidence of virulence of *C. diphtheriae*. Am. J. Pub. Health, **37**, 543–548.

KING, E. O., FROBISHER, M., JR., AND PARSONS, E. I. 1950 Further studies on the *in vitro* test for virulence of *Corynebacterium diphtheriae*. Am. J. Pub. Health, **40**, 704–709.

OKELL, C. C. 1929 The relationship of virulent to avirulent diphtheria bacilli. J. Hyg., **29**, 309–312.

ORSKOV, J. 1948 Chloroform-fat reactions with diphtheria bacilli. Acta Path. Microbiol. Scand., **25**, 829–834.

OUCHTERLONY, O. 1949 Antigen-antibody reactions in gels. Acta Path. Microbiol. Scand., **26**, 507–515.

ROSENAU, M. J., AND BAILEY, G. H. 1917 Effect of repeated injections of avirulent diphtheria bacilli, *B. hofmanni* and *B. xerosis*, in guinea-pigs. J. Infectious Diseases, **37**, 97–111.

SMITH, D. T., AND MARTIN, D. S. 1948 Zinsser's textbook of bacteriology. 9th ed. Appleton-Century-Crofts, New York.

STONE, F. M., AND HOBBY, G. L. 1934 A coccoid form of *C. diphtheriae* susceptible to bacteriophage. J. Bact., **27**, 403–417.

TOPLEY, W. W. C., WILSON, G. S., AND MILES, A. A. 1946 Principles of bacteriology and immunity. 3d ed. Williams & Wilkins Co., Baltimore, Md.

Week 5

Stress Response

Bacterial Growth

Molecular Microbiology (1991) 5(1), 49–59 0950382X9100008Q

Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*

R. Lange and R. Hengge-Aronis*
Fakultät für Biologie, Postfach 5560, Universität Konstanz, 7750 Konstanz, Germany.

Summary

During carbon-starvation-induced entry into stationary phase, *Escherichia coli* cells exhibit a variety of physiological and morphological changes that ensure survival during periods of prolonged starvation. Induction of 30–50 proteins of mostly unknown function has been shown under these conditions. In an attempt to identify C-starvation-regulated genes we isolated and characterized chromosomal C-starvation-induced *csi*::*lacZ* fusions using the λp*lac*Mu system. One operon fusion (*csi*2::*lacZ*) has been studied in detail. *csi*2::*lacZ* was induced during transition from exponential to stationary phase and was negatively regulated by cAMP. It was mapped at 59 min on the *E. coli* chromosome and conferred a pleiotropic phenotype. As demonstrated by two-dimensional gel electrophoresis, cells carrying *csi*2::*lacZ* did not synthesize at least 16 proteins present in an isogenic *csi*2$^+$ strain. Cells containing *csi*2::*lacZ* or *csi*2::Tn*10* did not produce glycogen, did not develop thermotolerance and H_2O_2 resistance, and did not induce a stationary-phase-specific acidic phosphatase (AppA) as well as another *csi* fusion (*csi*5::*lacZ*). Moreover, they died off much more rapidly than wild-type cells during prolonged starvation. We conclude that *csi*2::*lacZ* defines a regulatory gene of central importance for stationary phase *E. coli* cells. These results and the cloning of the wild-type gene corresponding to *csi*2 demonstrated that the *csi*2 locus is allelic with the previously identified regulatory genes *katF* and *appR*. The *katF* sequence indicated that its gene product is a novel sigma factor supposed to regulate expression of catalase HPII and exonuclease III (Mulvey and Loewen, 1989). We suggest that this novel sigma subunit of RNA polymerase defined by *csi*2/*katF*/*appR* is a central early regulator of a large starvation/stationary phase regulon in *E. coli* and propose '*rpoS*' ('σ^S') as appropriate designations.

Received 18 June, 1990; revised 14 August, 1990. *For correspondence. Tel. (7531) 882039; Fax (7531) 883688.

Introduction

Escherichia coli can survive in such diverse environments as the mammalian gut or freshwater and must therefore possess powerful mechanisms to sense and react appropriately to dramatic changes in the availability of nutrients, osmolarity, temperature and other external factors. Among these, nutrient limitation appears to have the most global influence on cellular physiology and morphology.

Unlike, for instance, *Bacillus subtilis*, *E. coli* does not form endospores under starvation conditions and was therefore classified as a non-differentiating bacterium (Matin *et al.*, 1989). Nevertheless, morphological effects can be observed in starving, i.e. non-growing, stationary *E. coli* cells, which become smaller and sphere-shaped. Physiological adaptations, however, are most dramatic. During the transition into stationary phase glycogen is synthesized as a storage compound (Okita *et al.*, 1981) and the cells develop a remarkable resistance against heat shock and H_2O_2 (Jenkins *et al.*, 1988). Several stationary-phase-induced proteins have been reported, some of which are related to these phenotypes. Among these are catalase HPII (*katE*; Loewen *et al.*, 1985), exonuclease III (*xthA*; Sak *et al.*, 1989), acidic phosphatase (*appA*; Touati *et al.*, 1987), the microcins B17 and C7 (Genilloud *et al.*, 1989; Diaz-Guerra *et al.*, 1989), the morphogene BolA and FtsZ, a gene product involved in cell division (Aldea *et al.*, 1989). The corresponding genes, however, can only be a small subset of all stationary-phase-induced genes. On two-dimensional O'Farrell electrophoresis gels, 30–50 protein spots have been identified which were induced in *E. coli* and *S. typhimurium* cells entering stationary phase because of carbon limitation (Groat *et al.*, 1986; Spector *et al.*, 1986). Using multiple peptidase mutants, it was shown that protein degradation is required for survival of prolonged starvation (Reeve *et al.*, 1984), probably because synthesis of starvation proteins is dependent on endogenous amino acid resources created by proteolysis. Approximately two-thirds of these proteins are not found in an adenylate cyclase (*cya*)-defective mutant, indicating a control by cAMP and the catabolite activator protein (CAP, encoded by *crp*). Since *cya* and *crp* survive prolonged starvation periods equally well as wild-type strains, cAMP-dependent starvation proteins are not essential, at least under the starvation conditions tested (Schultz *et al.*, 1988). However, during glucose-starvation-induced entry

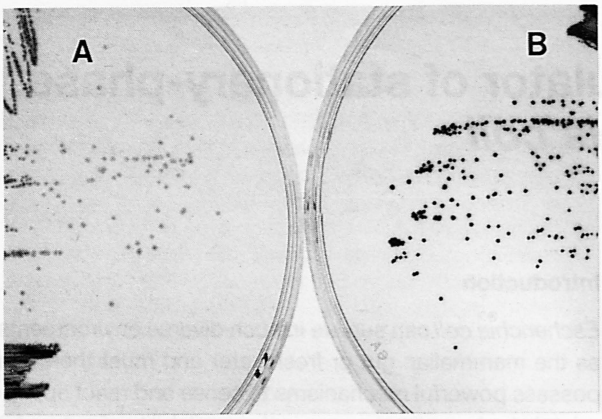


Fig. 1. Expression of *csi2*::*lacZ* on plates. MC4100 carrying *csi2*::*lacZ* was streaked on M9/XG plates containing 0.4% glucose (A) and 0.04% glucose ('starvation' plates; B).

into stationary phase, *E. coli* cells drastically increase their synthesis of cAMP, most of which is excreted into the medium (Buettner *et al.*, 1973). It is an unsolved paradox that cells appear to waste huge amounts of ATP to produce and excrete cAMP at a time when the energy source is becoming limited. The class of cAMP-independent starvation proteins includes the heat-shock proteins GroEL and DnaK and several other polypeptides that could also be induced by nitrogen and phosphate starvation (Schultz *et al.*, 1988).

The regulatory mechanisms responsible for the induction of starvation proteins can only be speculated about, since no general regulatory proteins have been identified so far. Guanosintetraphosphates have been implicated because a *relA*-independent increase in the intracellular ppGpp level correlates with the onset of starvation (Metzger *et al.*, 1989). Another potential factor in regulation is DNA topology. Plasmid DNA isolated from starving cells indicates a relaxation of DNA in stationary phase (Balke and Gralla, 1987). In the case of osmoregulated genes, it was demonstrated that changes in DNA supercoiling can have considerable effects on expression of global stimulons (Higgins *et al.*, 1988).

Using the λp*lac*Mu system (Bremer *et al.*, 1988) we isolated chromosomal *csi*::*lacZ* fusions. An adenylate-cyclase-defective mutant was used to identify cAMP-independent genes essential for survival during starvation. Several such *csi*::*lacZ* fusions have been characterized so far. This paper focuses on results obtained with one transcriptional fusion (*csi2*::*lacZ*) which allowed the identification of a central regulator of a large class of starvation proteins in *E. coli*.

Results

Isolation of csi::lacZ *fusions*

For the isolation of *csi*::*lacZ* fusions the randomly transposing λp*lac*Mu55 and λp*lac*Mu15 phages were used to obtain operon and protein fusions, respectively (Bremer *et al.*, 1988). *lacZ* fusions to *csi* genes were expected to synthesize higher levels of β-galactosidase on XG indicator plates with a low glucose concentration (0.02 or 0.04%; 'starvation' plates) than on plates with a high glucose content (0.4%). Since *lacZ* fusions to 'ordinary' catabolite-repressible operons (*mal*, *mgl*) were also induced under these conditions (our unpublished results), the isolation of *csi*::*lacZ* fusions was carried out in an adenylate-cyclase-deficient (Δ*cya*) strain in which neither catabolite-repressible operons nor non-essential cAMP/CAP-dependent starvation genes are expressed.

A Δ*cya* strain does not synthesize LamB, which is required as a receptor for λp*lac*Mu phages. We therefore introduced *zjb729*::Tn*10*, which is located in the intergenic region between *malK* and *lamB* and allows low-level constitutive expression of LamB under the control of the Tn*10* p_{out} promoter (Brass *et al.*, 1984).

Using this strain (RO1) as a receptor for λp*lac*Mu phages and the plate screen mentioned above, several chromosomal *csi*::*lacZ* fusions were isolated and characterized. All *csi*::*lacZ* fusions were P1-transduced into the wild-type strain MC4100 and into isogenic Δ*cya* strains (RO1, RH76). Preliminarily these fusions have been designated with numbers. The fusion *csi2*::*lacZ* characterized in this paper is a transcriptional fusion obtained with λp*lac*Mu55.

Kinetics of csi2::lacZ *expression*

A reduction in growth rate appeared to be the signal for induction of *csi2*::*lacZ*. This was demonstrated by the high β-galactosidase activity on starvation plates (0.04% glucose) where colonies exhibited a reduced growth rate limited by the diffusion of glucose (Fig. 1). In liquid minimal medium, however, fast growth stopped abruptly after exhaustion of glucose and only weak expression of *csi2*::*lacZ* was observed (Fig. 2A). Introduction of an adenylate cyclase mutation (Δ*cya*) reduced the growth rate of the *csi2*::*lacZ*-carrying strain (but not of a *csi*$^+$ strain) and resulted in increased expression of the fusion (Fig. 2B). In liquid rich medium, transition into stationary phase was accompanied by a gradual reduction of growth rate, during which a five-fold induction of *csi2*::*lacZ* was observed (Fig. 2C). During fast exponential growth no β-galactosidase was synthesized and pre-existing activity was diluted by cell division. In a similar experiment the early exponential phase was prolonged by repeated dilution and a basal expression level of 0.1 μmol min^{-1} mg^{-1} was determined (data not shown). Two-fold induction of *csi2*::*lacZ* could be obtained by a carbon 'downshift' from glucose to succinate, which also reduced the growth rate. When cells growing on minimal medium

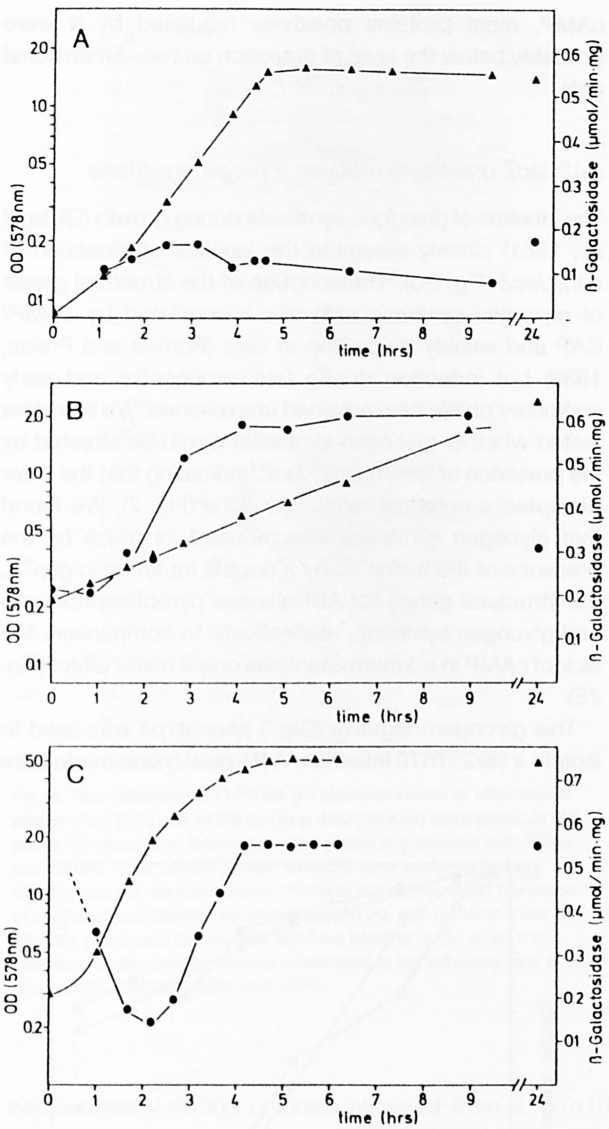


Fig. 2. Kinetics of induction of *csi2*::*lacZ*. Expression of *csi2*::*lacZ* was measured in MC4100 (*cya*$^+$) (A, C) and RO1 (Δ*cya*) (B). Cells were grown in M9 containing 0.1% glucose (A, B) or LB (C). OD_{578} was followed (triangles), and specific β-galactosidase activity was determined (circles).

with glucose were transferred to fresh medium without any carbon source, growth stopped immediately and no induction of *csi2*::*lacZ* was observed. Transfer into a phosphate-free medium (Groat *et al.*, 1986) resulted in induction of *csi2*::*lacZ* (to the fully induced level of 0.4–0.5 µmol min^{-1} mg^{-1}). In an adenylate-cyclase-deficient strain, however, phosphate starvation did not further increase the expression of the already induced *csi2*::*lacZ* fusion. Abrupt deprivation of nitrogen sources did not induce *csi2*::*lacZ* either in 'downshift' experiments or in cultures growing with a limiting concentration of ammonium chloride during (rather abrupt) entry into stationary phase. On the other hand, induced levels of β-galactosidase activity were observed in a culture growing slowly (with a doubling time of 140 min) with 2 mM glutamine as a nitrogen source or during gradual transition into stationary phase of a culture supplemented with a limiting concentration of yeast extract as a complex nitrogen source (data not shown).

csi2::lacZ *is negatively regulated by cAMP*

Surprisingly, we found higher expression and late exponential-phase induction of *csi2*::*lacZ* in a Δ*cya* strain (Fig. 2B). This suggested negative control by cAMP and we therefore tested the effect of cAMP directly. Strain MC4100 Δ*cya csi2*::*lacZ* was grown on minimal glucose medium and cAMP was added during mid-exponential phase. Expression of *csi2*::*lacZ* stopped immediately and pre-existing β-galactosidase was diluted during further growth (Fig. 3). Moreover, growth accelerated, indicating that at least in a Δ*cya* mutant, the gene disrupted by *csi2*::*lacZ* plays an important role during exponential growth in glucose minimal medium.

Chromosomal location of csi2::lacZ

Fusion *csi2*::*lacZ* was mapped on the *E. coli* chromosome using a collection of Tn*10* insertions distributed over the entire chromosome (Singer *et al.*, 1989). For approximate mapping, strains carrying Tn*10* insertions within 10-min intervals were pooled, whereas for fine mapping, single Tn*10* insertions were used. *csi2*::*lacZ* maps almost

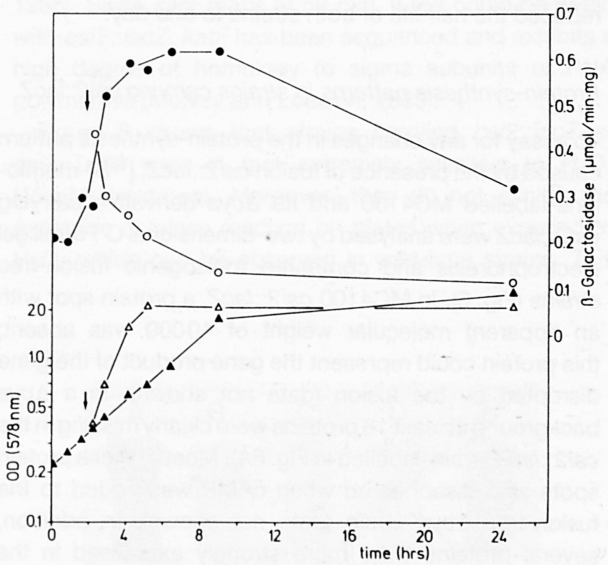


Fig. 3. *csi2*::*lacZ* is negatively regulated by cAMP. RO1 (Δ*cya*) carrying *csi2*::*lacZ* was grown on M9 containing 0.1% glucose, and OD_{578} (triangles) and specific β-galactosidase activity (circles) were determined. At the time indicated by the arrow the culture was divided and cAMP (5 mM) was added (open symbols).

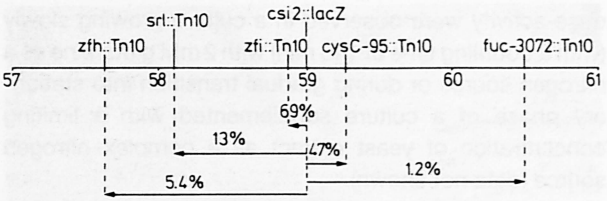


Fig. 4. Chromosomal location of *csi2*::*lacZ*. Co-transduction frequencies to Tn*10* insertions in marker genes of the 59-min region of the chromosome are shown. Arrowheads point to the selected markers in P1 transductions. Between 133 and 1645 transductants were screened for relevant markers in the various P1 transductions.

exactly at 59 min. Figure 4 shows co-transduction frequencies to surrounding gene markers and Tn*10* insertions. The gene order *srl–zfi*::Tn*10–csi2*::*lacZ–cysC* was confirmed by a three-factor cross in which a P1 lysate grown on a strain carrying *srl*::Tn*5* and *zfi*::Tn*10* (RO20) was transduced into strain MC4100 carrying *csi2*::*lacZ*, selecting for tetracycline resistance (Tc^R). Among Tc^R, Srl^- transductants, fusion-containing and fusion-free strains were obtained with equal frequency (3%), indicating a location of *csi2*::*lacZ* to the right of *zfi*::Tn*10*.

csi*2*::lacZ *reduces survival during prolonged starvation*

The loss of an essential *csi* gene due to disruption by *lacZ* can be expected to reduce the ability to survive prolonged starvation periods. Figure 5 demonstrates that this was the case for *csi2*::*lacZ*. Whereas cultures of a wild-type strain (MC4100) and an isogenic Δ*cya* derivative each exhibited a half-life of approximately six days, *csi2*::*lacZ* reduced the half-life of both strains to one day.

Protein-synthesis patterns in strains carrying csi*2*::lacZ

To assay for any changes in the protein-synthesis pattern caused by the presence of fusion *csi2*::*lacZ*, [^{35}S]-methionine-labelled MC4100 and its Δ*cya* derivative carrying *csi2*::*lacZ* were analysed by two-dimensional O'Farrell gel electrophoresis and compared to isogenic fusion-free strains (Fig. 6). In MC4100 *csi2*::*lacZ*, a protein spot with an apparent molecular weight of 40000 was absent; this protein could represent the gene product of the gene disrupted by the fusion (data not shown). In a Δ*cya* background at least 16 proteins were clearly missing in the *csi2*::*lacZ* strain (labelled in Fig. 6A). Most of these protein spots also disappeared when cAMP was added to the fusion-free Δ*cya* strain (data not shown). In addition, several proteins were more strongly expressed in the *csi2*::*lacZ* strain (labelled in Fig. 6B). We conclude that *csi2*::*lacZ* probably disrupts a regulatory gene that positively controls at least 16 other genes. Since in a *cya*$^+$ strain this regulatory gene is itself negatively controlled by cAMP, most proteins positively regulated by it were probably below the level of detection on two-dimensional gels.

csi*2*::lacZ *drastically reduces glycogen synthesis*

The kinetics of glycogen synthesis during growth (Okita *et al.*, 1981) closely resemble the kinetics of induction of *csi2*::*lacZ* (Fig. 2C). Transcription of the structural genes of glycogen-synthetic enzymes is regulated by cAMP/CAP and weakly by ppGpp *in vitro* (Romeo and Preiss, 1989) but induction during late exponential and early stationary phase has remained unexplained. We therefore tested whether glycogen synthesis would be affected by the presence of fusion *csi2*::*lacZ*, indicating that the latter disrupted a potential regulatory gene (Fig. 7). We found that glycogen synthesis was reduced as much by the presence of the fusion as by a double mutation in *glgCA*, the structural genes for ADPglucose pyrophosphorylase and glycogen synthase, respectively. In comparison, the lack of cAMP in a Δ*cya* mutant has only a minor effect (Fig. 7E).

This glycogen-negative (Glg^-) phenotype was used to isolate a *csi2*::Tn*10* insertion. A P1 pool lysate made from

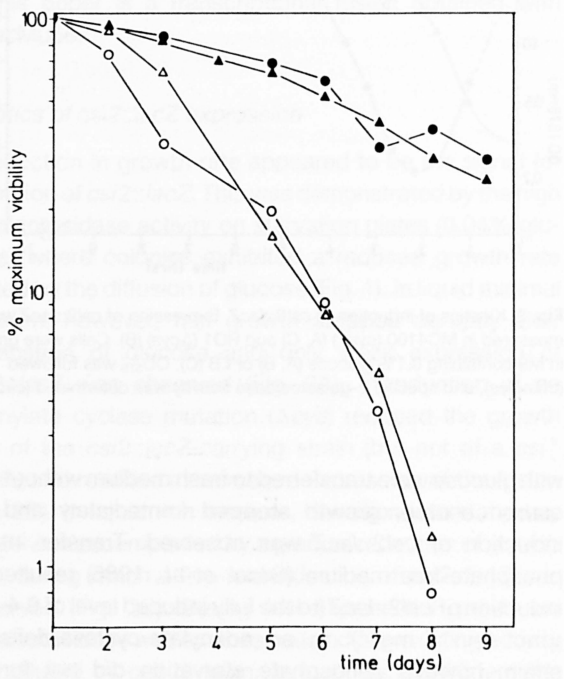


Fig. 5. Starvation survival of strains carrying *csi2*::*lacZ*. MC4100 (circles) and RO1 (Δ*cya*; triangles) carrying *csi2*::*lacZ* (open symbols) or no fusion (closed symbols) were grown on M9 and 0.1% glucose. After stationary phase was reached, incubation was continued for nine days under the same conditions. Viable cells were counted as colonies plated on LB-plates after appropriate dilution. One hundred percent viability corresponds to the number of viable cells counted one day after the culture had reached stationary phase.

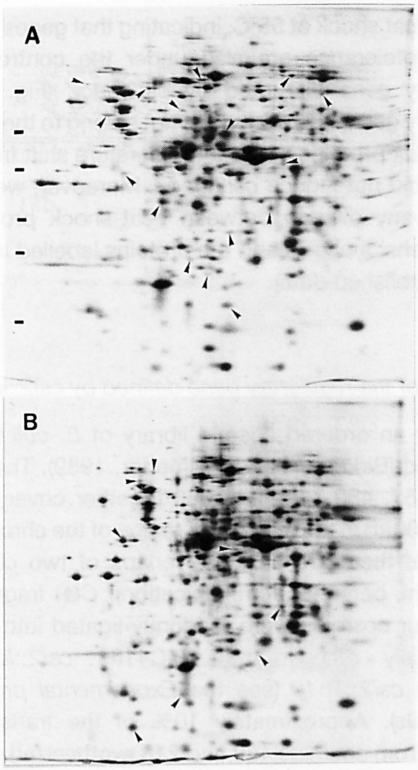


Fig. 6. Two-dimensional O'Farrell gel electorphoresis of total cellular protein. RH76 (A) and RH76 carrying *csi2*::*lacZ* (B) were grown in M9 and 0.1% glucose to late exponential phase and labelled with [^{35}S]-L-methionine. Total cellular protein extracts were analysed by two-dimensional gel electrophoresis. Proteins not expressed in the presence of *csi2*::*lacZ* are labelled by arrowheads in (A), and proteins more strongly expressed due to *csi2*::*lacZ* are labelled in (B). In (A) the positions of standard proteins are indicated at the left-hand side of the figure (67K, 45K, 36K, 29K, 24K, 20K).

approximately 40000 colonies obtained from a λ-Tn*10* 'hop' into the chromosome of MC4100 (M. Ehrmann, personal communication) was transduced into MC4100 carrying *csi2*::*lacZ*, and 40 white Tc^R colonies selected on XG/tetracycline indicator plates were purified. One of these 40 strains was Glg$^-$, indicating a *csi2*::Tn*10* insertion. This Tn*10* insertion could cross out *csi2*::*lacZ* with 100% efficiency in P1 transductions and exhibited the same co-transduction frequencies with nearby marker genes as *csi2*::*lacZ*. All phenotypes described in this paper for *csi2*::*lacZ* were also found for this Tn*10* insertion, which was therefore designated as *csi2*::Tn*10*. The other 39 Tn*10* insertions proved to be co-transducible with *csi2*::*lacZ*.

csi2::lacZ *inhibits the expression of a stationary-phase-specific acidic phosphatase (AppA)*

Fusion *csi2*::*lacZ* maps at the same chromosomal position as *appR*, a gene locus regulating *appA*-directed acidic phosphatase synthesis (Touati *et al.*, 1986). *appA* does not belong to the *pho* regulon, it is negatively regulated by cAMP, and it is induced when cells enter stationary phase (Touati *et al.*, 1987). We therefore tested whether *csi2*::*lacZ* and *appR* were allelic. Table 1 shows that stationary MC4100 *csi2*::*lacZ* cells exhibited a 14-fold reduction in acidic phosphatase activity relative to the wild-type strain. For spontaneous *appR* mutants, a four-fold reduction was reported (Touati *et al.*, 1986), indicating that these were not null mutants. This result indicated that *csi2*::*lacZ* and *appR* represent the same gene locus. This was further confirmed by the ability of a Δ*crp csi2*::*lacZ* strain but not of an isogenic fusion-free Δ*crp* strain to grow on succinate (data not shown), a phenotype described for *appR* Δ*crp* mutants (Touati *et al.*, 1986).

Table 1 also suggests that negative regulation of *appA* by cAMP is both indirect via regulation of *csi2*(*appR*) by cAMP and direct at the level of *appA* expression, since introduction of Δ*cya* or Δ*crp* into the *csi2*::*lacZ* strain resulted in a five-fold increase in AppA activity.

Strains carrying csi2::lacZ *or* csi2::*Tn*10 *do not develop* H_2O_2 *resistance*

Glucose-starved stationary wild-type cells exhibit a remarkable resistance against treatment with H_2O_2 (15–50mM; Jenkins *et al.*, 1988), which is dependent on catalase HPII and exonuclease III. These enzymes are encoded, respectively, by *katE* (Loewen *et al.*, 1985) and *xthA* (Sak *et al.*, 1989), both of which are positively controlled by *katF* (Sak *et al.*, 1989; Sammartano *et al.*, 1986). Since *katF* maps at 59 min, it too could be allelic with *csi2*::*lacZ*. *katF* has been sequenced and exhibits a high degree of homology to sigma subunits of RNA polymerase (Mulvey and Loewen, 1989).

Figure 8 shows that strains carrying *csi2*::*lacZ* or *csi2*::Tn*10* were in fact extremely sensitive to H_2O_2 (15mM) treatment. Moreover, they do not exhibit the explosive catalase reaction on plates when exposed to H_2O_2 which can be observed in wild-type strains. This

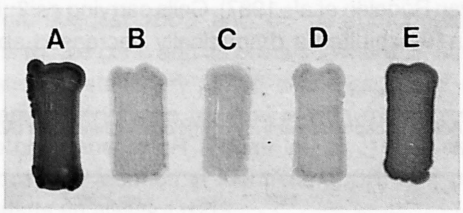


Fig. 7. Glycogen accumulation during growth on plates. The strains were grown overnight on glycogen test plates and stained with iodine vapors. A, MC4100; B, MC4100 *csi2*::*lacZ*; C, MC4100 *csi2*::Tn*10*; D, REI17 (*glgCA*); E, RH76 (Δ*cya*).

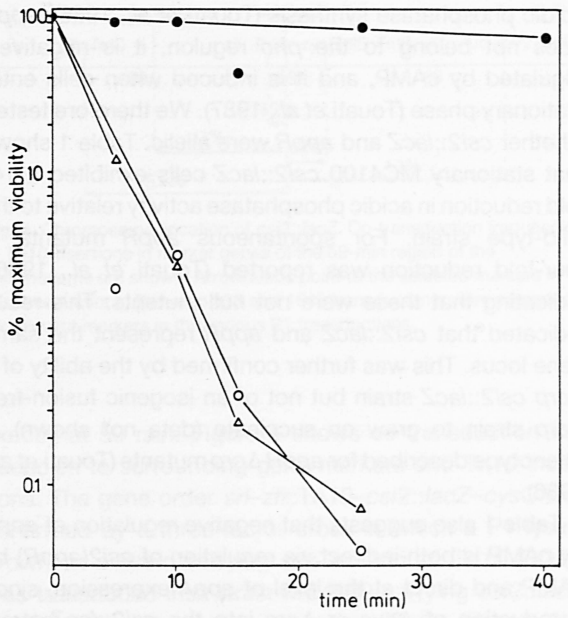


Fig. 8. Survival of H_2O_2 treatment. Stationary-phase cultures of MC4100 (closed circles), MC4100 *csi2*::*lacZ* (open circles) and MC4100 *csi2*::Tn*10* (open triangles) were exposed to 15 mM H_2O_2, and viable cell numbers were determined by diluting and plating aliquots onto LB-plates. One hundred percent viability corresponds to the viable cell number determined immediately before the addition of H_2O_2.

result and, finally, the comparison of the cloned *csi2* gene to the cloned *katF* unequivocally demonstrated that the two genes are identical (see below). Interestingly, addition of 60 μM H_2O_2 to growing cells did not have any influence on the level of expression of *csi2*::*lacZ* (data not shown) although under these conditions adaptive H_2O_2 resistance is observed (Jenkins *et al.*, 1988).

Strains carrying csi2::lacZ *or* csi2::*Tn*10 *are impaired in stationary-phase thermotolerance*

Upon entry into stationary phase, wild-type cells develop a pronounced tolerance against transient heat shocks at 55–57°C (Jenkins *et al.*, 1988). Although the heat-shock proteins GroEL and DnaK are induced in glucose-starved cells, induction of the classical *rpoH*-controlled heat-shock genes alone is not sufficient for this thermotolerance (van Bogelen *et al.*, 1987). Cells carrying *csi2*::*lacZ* or *csi2*::Tn*10* exhibited a dramatically decreased ability to survive heat shock at 55°C, indicating that genes involved in thermotolerance are also under the control of the regulatory gene disrupted by *csi2*::*lacZ* (Fig. 9). This regulatory gene, however, does not belong to the family of heat-shock proteins, since a temperature shift from 30°C to 42°C did not induce *csi2*::*lacZ*. Moreover, we did not observe any overlap between heat-shock proteins on two-dimensional gels and the proteins labelled in Fig. 6A (our unpublished data).

Table 1. Acidic phosphatase activity of stationary-phase cells (AppA; nmol min^{-1} mg^{-1}).

Strain	*csi*$^+$	*csi2*::*lacZ*
MC4100	38.2	2.7
MC4100 Δ*cya*	44.8	14.4
MC4100 Δ*crp*	67.7	17.7

Cloning of the regulatory gene defined by csi2::lacZ

Recently an ordered cosmid library of *E. coli* K12 was described (Birkenbihl and Vielmetter, 1989). The cosmid clones 557, 480, 218, and 333 together cover approximately 100 kb in the 58–60 min region of the chromosome such that there are always overlaps of two clones (R. Birkenbihl, personal communication). *Cla*I fragments of these four cosmids were randomly ligated into pBR322 and finally recovered in MC4100 *csi2*::*lacZ* and MC4100 *csi2*::Tn*10* (see the *Experimental procedures* for details). Approximately 10% of the transformants derived from cosmids 480 and 218 synthesized glycogen and exhibited catalase activity. Plasmids were prepared from six transformants derived from each cosmid. All 12 plasmids carried an identical 4.35 kb *Cla*I insert which had been present in both cosmids. A restriction map of this fragment is shown in Fig. 10A. An internal 2.4 kb *Kpn*I

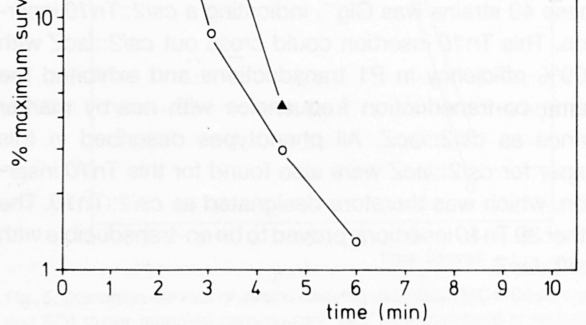


Fig. 9. Survival of heat shock at 55°C. Stationary-phase cells of MC4100 (closed circles), MC4100 *csi2*::*lacZ* (open circles) and MC4100 *csi2*::Tn*10* (triangles) were transferred to prewarmed tubes and viable cell numbers were determined by plating aliquots onto LB-plates. One hundred percent viability corresponds to the viable cell number determined immediately before heat shock.

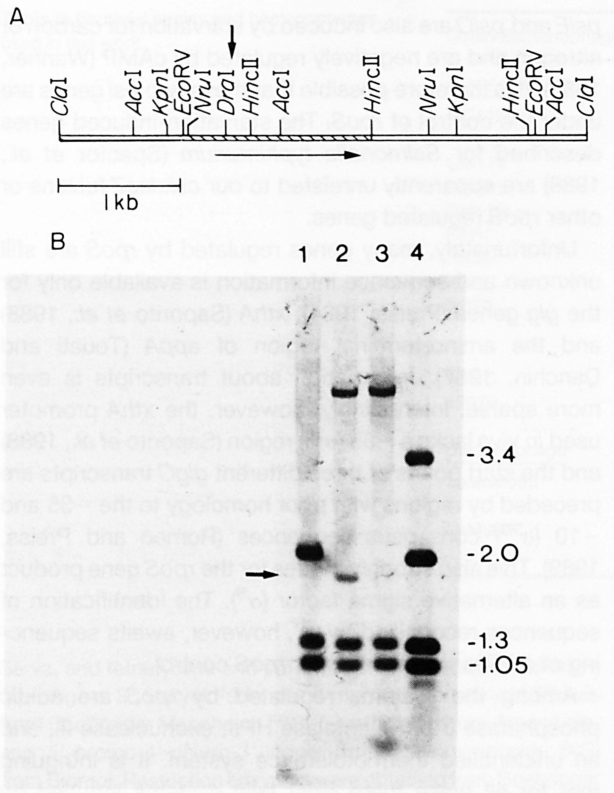


Fig. 10. Restriction map of the *csi2*::*lacZ* complementing *Cla*I fragment (A) and localization of *csi2*::*lacZ* within this fragment in the chromosome (B).

A. The 4.35 kb *Cla*I insert cloned into pBR322 (yielding pRH320) is shown. The cross-hatched fragment between the left *Nru*I site and the central *Hin*cII site corresponds to the region sequenced by Mulvey and Loewen (1989). The horizontal and the vertical arrows indicate the position of the open reading frame corresponding to *katF*/*csi2* and the position of the *csi2*::*lacZ* insertion in the corresponding region in the chromosome, respectively.

B. Chromosomal DNA of MC4100 (1), MC4100 *csi2*::*lacZ* (2), MC4100 *csi2*::Tn*10* (3) and pRH320 plasmid DNA was digested with *Cla*I and *Nru*I. Agarose gel electrophoresis and Southern transfer were followed by hybridization to the digoxigenine-labelled total *Cla*I insert of pRH320. Also a vector band (3.4 kb) is visible because of slight contamination of the probe with pBR322 vector sequences and can be used as a size standard. The arrow indicates a fragment (1.8 kb) from which the position of the fusion joint can be derived (see text).

fragment is in good agreement with the restriction map of a fragment carrying *katF* (Mulvey and Loewen, 1989). The internal 2.9 kb *Eco*RV and 2.0 kb *Nru*I fragments, both of which carry *katF*, were subcloned into pBR322, and complemented the glycogen- and catalase-negative phenotypes conferred by *csi2*::*lacZ* and *csi2*::Tn*10*. Southern hybridization experiments (Fig. 10B) with *Cla*I/*Nru*I-digested chromosomal DNA from wild-type and *csi2*::*lacZ* and *csi2*::Tn*10* insertion strains and the 4.35 kb *Cla*I fragment as a probe indicated that both insertions are within the 2.0 kb *Nru*I fragment. The first restriction site cut within λp*lac*Mu55 under these conditions is a *Cla*I site 1.4 kb downstream of the fusion point, and the last one is a *Cla*I site 3.45 kb upstream of the end of λp*lac*Mu (Bremer *et al.*, 1988). The 1.8 kb fragment labelled with an arrow in Fig. 10B, therefore, indicated that the fusion joint of *csi2*::*lacZ* was 0.4 kb to the right of the left *Nru*I restriction site, i.e. within the *katF* open reading frame. This demonstrated unequivocally that *csi2* and *katF* are allelic.

Discussion

In the present paper we describe the isolation and characterization of a starvation-inducible *lacZ* fusion with an extremely pleiotropic phenotype. Strains carrying *csi2*::*lacZ* or a Tn*10* insertion in the same gene were defective in a variety of stationary-phase-specific functions, among these glycogen synthesis, H_2O_2 resistance, thermotolerance, and activity of acidic phosphatase (AppA). They were seriously impaired in their ability to survive prolonged starvation periods and were deficient in synthesis of at least 15–20 proteins, as demonstrated by two-dimensional gel electrophoresis. These phenotypes point to a role as a central regulator of starvation functions for the gene disrupted by *csi2*::*lacZ*.

Fusion *csi2*::*lacZ* maps at the same location (59 min on the *E. coli* chromosome) as *katF*. This gene positively regulates catalase HPII (*katE*) and exonuclease III (*xthA*), which are required for H_2O_2 resistance (Sak *et al.*, 1989). Cloning of the gene from the 59-min region which complemented the *csi2*::*lacZ* phenotypes established its identity with *katF*. Also, the fusion joint of *csi2*::*lacZ* could be localized within the *katF* open reading frame. The *katF* sequence exhibits high homology to *rpoD*, which encodes the 'house-keeping' sigma subunit of RNA-polymerase (σ^{70}) in *E. coli*. It was suggested, therefore, that *katF* acted as a minor sigma factor in the control of an unknown protection or repair regulon to which *katE* and *xthA* belong (Mulvey and Loewen, 1989). In the light of our results it becomes clear that *csi2*/*katF* is a major regulatory gene for a large starvation regulon that comprises practically all stationary-phase-specific functions of *E. coli* described so far. In view of its sequence we consider it more than likely that the *csi2*/*katF* gene product acts as a sigma factor modifying the promoter-sequence recognition of RNA polymerase during transition into stationary phase. We therefore suggest '*rpoS*' and 'σ^S' as appropriate designations for the gene and its product, where 'S' stands for starvation or stationary phase. Since *csi2*::*lacZ* exhibits the same phenotypes as strains with mutations in the regulatory gene *appR* (Touati *et al.*, 1986) and their chromosomal location is identical, *appR*, too, probably represents an allele of *rpoS*.

The induction profile of the operon fusion *csi2*::*lacZ* indicated that *rpoS* is induced at the transcriptional level during transition from the exponential phase into the

stationary phase. Induction could be observed irrespective of whether the limiting compound was a source of carbon, of nitrogen or of phosphate. The inducing signal appears to be a reduction in growth rate, since various treatments with this effect, such as growth on plates with low glucose, carbon-source downshifts from glucose to succinate, or slow growth with glutamine as a nitrogen source, also resulted in induction. In various media, such as LB or minimal medium with a limiting concentration of complex nitrogen sources, transition into stationary phase is characterized by a gradual reduction in growth rate and results in induction of *csi*2::*lacZ*. Abrupt deprivation of carbon or nitrogen sources, which results in an immediate growth stop, did not induce *csi*2::*lacZ*. Interestingly, in similar phosphate-starvation experiments, increased expression of *csi*2::*lacZ* was observed. The kinetics of *csi*2::*lacZ* induction under various conditions show that exponentially growing cells maintain a low basal level of *rpoS* expression. Although we have not been able to detect any deficiency due to *csi*2::*lacZ* in exponentially growing wild-type cells, a three-fold increase in doubling time and a higher expression of *csi*2::*lacZ* in a Δ*cya* mutant could indicate that *rpoS* also plays a role during growth in an adenylate-cyclase-deficient mutant.

rpoS is negatively controlled by cAMP. The negative cAMP effect is partly reflected in the regulation of structural genes of the *rpoS* regulon. *appA* is repressed by cAMP, probably indirectly via the cAMP effect on *rpoS* expression as well as by a direct cAMP/CAP interaction with *appA* promoter sequences (Touati and Danchin, 1987). Also, negative regulation by cAMP might explain induction of *csi*2::*lacZ* by abrupt starvation for phosphate. Adenylate cyclase is stimulated by phosphate (Peterkofsky *et al.*, 1989). A sudden lack of phosphate may therefore result in a decrease in the intracellular level of cAMP, thus allowing better expression of *csi*2::*lacZ*.

For the glycogen-synthetic genes (*glgB*, *glgCA*), direct positive control by cAMP/CAP has been observed in *in vitro* transcription/translation assays (Romeo and Preiss, 1989). However, we observed only a minor reduction of glycogen synthesis in Δ*cya*, Δ*crp* or Δ*cya*Δ*crp* mutants, compared to a dramatic reduction caused by disruptions in *rpoS*. *rpoS* may be absolutely required for the expression of the *glg* genes, or for the expression of a factor required for activation of the *glg*-encoded enzymes. The same absolute requirement was found for H_2O_2 resistance, thermotolerance and for expression of another starvation-inducible operon fusion (*csi*5::*lacZ*) that maps around 99 min on the chromosome (our unpublished data).

A series of phosphate-starvation-inducible (*psi*) *lacZ* fusions was described by Wanner and McSharry (1982). Although none of our *csi*::*lacZ* fusions mapped at the chromosomal locations of any of the described *psi* genes, *psiE* and *psiO* are also induced by starvation for carbon or nitrogen and are negatively regulated by cAMP (Wanner, 1983). It is therefore possible that these two *psi* genes are under the control of *rpoS*. The starvation-induced genes described for *Salmonella typhimurium* (Spector *et al.*, 1988) are apparently unrelated to our *csi*::*lacZ* fusions or other *rpoS* regulated genes.

Unfortunately, many genes regulated by *rpoS* are still unknown and sequence information is available only for the *glg* genes (Preiss, 1984), *xthA* (Saporito *et al.*, 1988) and the amino-terminal region of *appA* (Touati and Danchin, 1987). Information about transcripts is even more sparse. Interestingly, however, the *xthA* promoter used *in vivo* lacks a −35 (σ^{70}) region (Saporito *et al.*, 1988) and the start points of three different *glgC* transcripts are preceded by regions with poor homology to the −35 and −10 (σ^{70}) consensus sequences (Romeo and Preiss, 1989). This also strongly argues for the *rpoS* gene product as an alternative sigma factor (σ^S). The identification of sequences recognized by σ^S, however, awaits sequencing of additional genes under *rpoS* control.

Among the systems regulated by *rpoS* are acidic phosphatase (AppA), catalase HPII, exonuclease III, and an unidentified thermotolerance system. It is intriguing that for all these there exist twin systems induced by specific stresses. These include alkaline phosphatase which is controlled by *phoR*/*phoB* (Wanner, 1987), the *oxyR*-regulated catalase HPI (Christman *et al.*, 1985), various DNA-repair systems that belong to the *lexA*/*recA*-dependent SOS regulon (Walker, 1987), and the heat-shock system, which is under the control of another alternative sigma factor, *rpoH* (Neidhardt and van Bogelen, 1987). Whereas these systems are induced in reaction to phosphate limitation, oxidative stress, DNA damage and temperature shifts, respectively, the *rpoS*-dependent systems with similar functions are not induced by these specific stimuli. It appears to have been more 'economic' to evolve alternate systems for all these functions, rather than subjecting single systems to highly complex multiple controls. In this respect, the *rpoS*-regulated functions belong to a 'superglobal' regulatory programme comparable in its complexity and physiological significance to the sporulation programme of *B. subtilis*. It is possible that in this programme *rpoS* is the first regulatory gene followed by other unidentified regulators.

Experimental procedures

Media and chemicals

Media were purchased from Difco and prepared according to Miller (1972). Concentrations of carbon sources and supplements are indicated in the various experiments. Kanamycin and ortho-nitrophenyl-galactopyranoside (ONPG) were purchased from

Table 2. Bacterial strains and bacteriophages used in this work.

Strain/Bacteriophage	Relevant genotype	Source
E. coli K12		
MC4100	F⁻, (*argF–lac*)*U169*, *araD139*, *rpsL150*, *ptsF25*, *flbB5301*, *rbsR*, *deoC*, *relA1*	Silhavy *et al.* (1984)
GM114	MC4100, *zfi-551*::Tn*10*	G. May *et al.* (1986)
GM115	MC4100, *zfh-552*::Tn*10*	G. May *et al.* (1986)
GM119	MC4100, *srl*::Tn*10*	G. May *et al.* (1986)
REI17	MC4100, *glgAC*, *malK*::*lacZ*, *mall*, *ugp*::Tn*10*	J. Reidl
RH76	MC4100, Δ*cya851*	This study
RO1	RH76, *zjb-729*::Tn*10*	This study
RO20	F⁻, (*argF–lac*)*U169*, *araD139*, *rpsL150*, (*glpA–glpT*)*593*, *gyrA*, Δ*phoA8*, *recA*, *srl*::Tn*5*, *zfi-551*::Tn*10*	This study
12173	*cysC-95*::Tn*10*	Singer *et al.* (1989)
12079	*fuc-3072*::Tn*10*	Singer *et al.* (1989)
Bacteriophage		
λp*lac*Mu55	Mucts62, *ner*⁺, *A'*am1093, *'uvrD'*, Mu*S''trp'*, *lacZ*⁺, *lacY*⁺, *lacA'*, *imm*λ	Bremer *et al.* (1988)
λpMu507	Mucts62, Mu*A*⁺*B*⁺, λ*c*ts857, *S*am7	Magazin *et al.* (1977)

Serva, and tetracycline and para-nitrophenyl-phosphate (PNPP) from Sigma. 3',5'-cyclo-adenosine-monophosphate (cAMP) was from Boehringer Mannheim, [^{35}S]-L-methionine from Amersham, and 5'-bromo-4'-chloro-3'-indolyl-β-D-galactopyranoside (XG) from Biomol. Restriction enzymes were obtained from Boehringer Mannheim, BRL (*Hin*cII) or New England Biolabs (*Acc*I). T4-DNA ligase was from Pharmacia. Chemicals for two-dimensional gel electrophoresis were purchased from BioRad. All other chemicals were obtained from Merck.

Bacterial strains, phages and growth conditions

Bacteria were grown aerobically at 37°C, and growth was monitored by measuring optical density at 578nm (OD_{578}). Bacterial strains and phages used in this study are listed in Table 2.

Genetic techniques and DNA manipulations

For P1 transduction, standard techniques were used (Miller, 1972). Chromosomal *lacZ* fusions were isolated by co-infection with λpMu507 as a helper phage and the fusion phages λp*lac*-Mu55 or λp*lac*Mu15 and selection for kanamycin-resistant strains (Bremer *et al.*, 1988). For DNA manipulations, procedures described by Silhavy *et al.* (1984) and Sambrook *et al.* (1989) were followed.

For cloning of the wild-type gene from the 59-min region that complemented *csi*2::*lacZ* phenotypes, four cosmid clones (nos 557, 480, 218, and 333; R. Birkenbihl, personal communication) carrying DNA from this region were digested with *Cla*I, randomly ligated into pBR322, and transformed into a highly transformable strain. Approximately 1000 transformants derived from each cosmid were pooled, and the plasmids were prepared and transformed into MC4100(*csi*2::*lacZ*) and MC4100(*csi*2::Tn*10*). Replicas of the transformation plates were assayed for glycogen (Glg) accumulation and catalase (Cat) activity (see below). Glg⁺ Cat⁺ transformants were purified and used for plasmid preparation.

Chromosomal DNA was isolated using the procedure of Silhavy *et al.* (1984). For Southern transfer, labelling of probes, and hybridization, protocols provided by Sambrook *et al.* (1989) and Boehringer Mannheim (non-radioactive DNA labelling and detection kit) were followed.

Survival assays

Long-time starvation survival was assayed by growing cells in 50ml of M9 medium containing 0.1% glucose. During growth and then for up to 10d samples were withdrawn for determination of OD_{578} and viable cell numbers on LB-plates after appropriate dilution in 0.9% NaCl.

For the heat-shock survival assay, cells were grown overnight in LB. Stationary cells were washed and diluted in 0.9% NaCl to a density of about 5000 cells ml^{-1}. One-millilitre samples were put into prewarmed glass tubes (55°C) and at the times indicated 0.1ml portions were plated directly onto LB-plates to determine viable cell numbers.

For H_2O_2 treatment, cells were grown overnight in LB, washed and resuspended in 0.9% NaCl to an OD_{578} of 1.0. H_2O_2 was added to a final concentration of 15mM. At the times indicated, 0.1-ml samples were withdrawn, diluted immediately in 0.9% NaCl and plated onto LB-plates.

Plate assays

β-galactosidase activity was assayed qualitatively on various plates containing 50 μg ml^{-1} XG.

Accumulation of glycogen was tested by growing cells as single colonies or in patches on plates containing 1% glucose (Latil-Damotte and Lares, 1977). Instead of overlaying the plates with an iodine salt solution, they were stained with iodine vapours by exposure to crystalline iodine for 2–5 min. Glycogen-accumulating strains stain dark brown; staining is reversible, can be repeated several times, and does not kill the cells.

Catalase activity was tested qualitatively by applying small drops of H_2O_2 (30%) on to colonies or patches grown overnight on LB plates.

Enzyme assays

Quantitative β-galactosidase assays were performed as described by Miller (1972). Acidic phosphatase (AppA) was assayed in stationary-phase overnight cultures exactly as described by Dassa *et al.* (1982).

Two-dimensional O'Farrell gel electrophoresis

This was performed according to the method of O'Farrell *et al.* (1977). Cells were grown in M9 medium containing 0.1% glucose to late exponential phase. The OD_{578} of 1-ml samples was adjusted to 0.15 with the same medium. Ten to fifteen micro-Curies of [^{35}S]-L-methionine (>1000 Ci mmol^{-1}) was used to label a 1-ml sample for 2 min, followed by a 1-min chase with 0.2 mM unlabelled methionine. Samples were precipitated with trichloroacetic acid (10%), washed with acetone and resuspended in first-dimensional sample buffer. Two-dimensional gel electrophoresis was carried out in a BioRad mini-Protean device according to the directions given by the manufacturer. Separation in the first dimension was by non-equilibrium pH gradient electrophoresis, where the pH gradient ranged from approximately pH 4.8 to pH 7 (left and right side, respectively, in Fig. 6). Running conditions were as follows: 10 min of pre-electrophoresis at 500 V without a buffer change, 10 min of electrophoresis at 500 V, followed by 4 h 50 min at 750 V. In the second dimension, sodium dodecyl sulphate/polyacrylamide (12%) gels were run for 100 min at 135 V. Gels were stained with Coomassie Brilliant Blue and dried. Fuji X-ray films were used for autoradiography.

Acknowledgements

We thank Rainer Birkenbihl, Erhard Bremer, Gerhard May and Joachim Reidl for bacterial strains and phages, and Carol Gross and her co-workers for providing us with a collection of mapped Tn*10* insertions. We appreciate the expert help of Vickie Koogle in preparing the manuscript. We are grateful to Winfried Boos, in whose laboratory this work was carried out, for his interest in the project and many helpful discussions. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 156).

References

Aldea, M., Garrido, T., Hernandez-Chico, C., Vicente, M., and Kushner, S.R. (1989) Induction of a growth-phase dependent promoter triggers transcription of *bolA*, an *Escherichia coli* morphogene. *EMBO J* **8:** 3923–3931.

Balke, V.L., and Gralla, J.D. (1987) Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J Bacteriol* **169:** 4499–4506.

Birkenbihl, R.P., and Vielmetter, W. (1989) Cosmid-derived map of *E. coli* strain BHB2600 in comparison to the map of strain W3110. *Nucl Acids Res* **17:** 5057–5069.

Brass, J.M., Manson, M.D., and Larson, T.J. (1984) Transposon Tn*10* dependent expression of the *lamB* gene in *Escherichia coli*. *J Bacteriol* **159:** 93–99.

Bremer, E., Silhavy, T.J., and Weinstock, G.M. (1988) Transposition of λ*plac*Mu is mediated by the A protein altered at its carboxy-terminal end. *Gene* **71:** 177–186.

Buettner, M.J., Spitz, E., and Rickenberg, H.V. (1973) Cyclic 3′,5′-monophosphate in *Escherichia coli*. *J Bacteriol* **114:** 1068–1073.

Christman, M.F., Morgan, R.W., Jacobson, F.S, and Ames, B.N. (1985) Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in *Salmonella typhimurium*. *Cell* **41:** 753–762.

Dassa, E., Cahu, M., Desjoyaux-Chevel, B., and Boquet, P. (1982) The acid phosphatase with optimum pH of 2.5 of *Escherichia coli*. *J Biol Chem* **257:** 6669–6676.

Diaz-Guerra, L., Moreno, F., and SanMillan, J.L. (1989) *appR* gene product activates transcription of microcin C7 plasmid genes. *J Bacteriol* **171:** 2906–2908.

Genilloud, O., Moreno, F., and Kolter, R. (1989) DNA sequence, products and transcriptional pattern of the genes involved in production of the DNA replication inhibitor microcin B17. *J Bacteriol* **171:** 1126–1135.

Groat, R.G., Schultz, J.E., Zychlinsky, E., Bockman, A.T., and Matin, A. (1986) Starvation proteins in *Escherichia coli*: kinetics of synthesis and role in starvation survival. *J Bacteriol* **168:** 486–493.

Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., May, G., and Bremer, E. (1988) A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* **52:** 569–584.

Jenkins, D.E., Schultz, J.E., and Matin, A. (1988) Starvation-induced cross-protection against heat or H_2O_2 challenge in *Escherichia coli*. *J Bacteriol* **170:** 3910–3914.

Latil-Damotte, M., and Lares, C. (1977) Relative order of *glg* mutations affecting glycogen biosynthesis in *Escherichia coli* K12. *Mol Gen Genet* **150:** 325–329.

Loewen, P.C., and Triggs, B.L. (1984) Genetic mapping of *katF*, a locus that with *katE* affects the synthesis of a second catalase species in *Escherichia coli*. *J Bacteriol* **160:** 668–675.

Loewen, P.C., Switala, J., and Triggs-Raine, B.L. (1985) Catalase HPI and HPII in *Escherichia coli* are induced independently. *Arch Biochem Biophys* **243:** 144–149.

Magazin, M., Howe, M., and Allet, B. (1977) Partial correlation of the genetic and physical maps of bacteriophage Mu. *Virology* **77:** 677–688.

Matin, A., Auger, E.A., Blum, P.H., and Schultz, J.E. (1989) Genetic basis of starvation survival in nondifferentiating bacteria. *Annu Rev Microbiol* **43:** 293–316.

May, G., Faatz, E., Villarejo, M., and Bremer, E. (1986) Binding protein dependent transport of glycin betaine and its osmotic regulation in *Escherichia coli* K12. *Mol Gen Genet* **205:** 225–233.

Metzger, S., Schreiber, G., Aizenman, E., Cashel, M., and Glaser, G. (1989) Characterization of the *relA1* mutation and a comparison of *relA1* with new *relA* null alleles in *Escherichia coli*. *J Biol Chem* **264:** 21146–21152.

Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Mulvey, M.R., and Loewen, P.C. (1989) Nucleotide sequence of *katF* of *Escherichia coli* suggest KatF protein is a novel σ transcription factor. *Nucl Acids Res* **17:** 9979–9991.

Neidhardt, F.C., and van Bogelen, R.A. (1987) In Escherichia coli *and* Salmonella typhimurium. Neidhardt, F.C. (ed.). Washington, D.C.: American Society for Microbiology, pp. 1334–1345.

O'Farrell, P.Z., Goodman, H.M., and O'Farrell, P.H. (1977) High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell* **12:** 1133–1142.

Okita, T.W., Rodriguez, R.L., and Preiss, J. (1981) Biosynthesis of bacterial glycogen: Cloning of the glycogen biosynthetic enzyme structural genes of *Escherichia coli*. *J Biol Chem* **256:** 6944–6952.

Peterkofsky, A., Svenson, I., and Amin, N. (1989) Regulation of *Escherichia coli* adenylate cyclase activity by the phospho-enolpyruvate:sugar phosphotransferase system. *FEMS Microbiol Rev* **63:** 103–108.

Preiss, J. (1984) Bacterial glycogen synthesis and its regulation. *Annu Rev Microbiol* **38:** 419–458.

Reeve, C.A., Bockman, A.T., and Matin, A. (1984) Role of protein degradation in the survival of carbon starved *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol* **157:** 758–763.

Romeo, T., and Preiss, J. (1989) Genetic regulation of glycogen biosynthesis in *Escherichia coli*: *in vitro* effects of cyclic AMP and guanosine 5′-diphosphate-3′-diphosphate and analysis of *in vitro* transcripts. *J Bacteriol* **171:** 2773–2782.

Sak, B.D., Eisenstark, A., and Touati, D. (1989) Exonuclease III and the catalase hydroperoxidase II in *Escherichia coli* are both regulated by the *katF* gene product. *Proc Natl Acad Sci USA* **86:** 3271–3275.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sammartano, L.J., Tuveson, R.W., and Davenport, R. (1986) Control of sensitivity to inactivation by H_2O_2 and broad-spectrum near-UV radiation by the *Escherichia coli katF* locus. *J Bacteriol* **168:** 13–21.

Saporito, S.M., Smith-White, B.J., and Cunningham, R.P. (1988) Nucleotide sequence of the *xth* gene of *Escherichia coli* K12. *J Bacteriol* **170:** 4542–4547.

Schultz, J.E., Latter, G.I., and Matin, A. (1988) Differential regulation by cyclic AMP of starvation protein synthesis in *Escherichia coli*. *J Bacteriol* **170:** 3903–3909.

Silhavy, T.J., Berman, M.L., and Enquist, L.W. (1984) *Experiments With Gene Fusions*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Singer, M., Baker, T.A., Schnitzler, G., Deischel, S.M., Goel, M., Dove, W., Jaacks, K.J., Grossman, A.D., Erickson, J.W., and Gross, C.A. (1989) A collection of strains containing genetically linked alternating antibiotic resistance elements. *Microbiol Rev* **53:** 1–24.

Spector, M.P., Aliabadi, Z., Gonzales, T., and Foster, J.W. (1986) Global control in *Salmonella typhimurium*: two-dimensional electrophoretic analysis of starvation-, anaerobiosis- and heat shock-inducible proteins. *J Bacteriol* **168:** 420–424.

Spector, M.P., Park, Y.K., Tirgari, S., Gonzales, T., and Foster, J.W. (1988) Identification and characterization of starvation-regulated genetic loci in *Salmonella typhimurium* by using Mu*d*-directed *lacZ* operon fusions. *J Bacteriol* **170:** 345–351.

Touati, E., Dassa, E., and Boquet, P.L. (1986) Pleiotropic mutations in *appR* reduce pH 2.5 acid phosphatase expression and restore succinate utilization in *crp*-deficient strains of *Escherichia coli*. *Mol Gen Genet* **202:** 257–264.

Touati, E., and Danchin, A. (1987) The structure of the promoter and amino terminal region of the pH 2.5 acid phosphatase structural gene (*appA*) of *E. coli*: a negative control of transcription mediated by cyclic AMP. *Biochimie* **69:** 215–221.

Touati, E., Dassa, E., Dassa, J., and Boquet, P.L. (1987) In *Phosphate Metabolism and Cellular Regulation in Microorganisms*. Torriani-Gorini, A., Rothman, F.G., Silver, S., Wright, A., and Yagil, E. (eds). Washington, D.C.: American Society for Microbiology, pp. 31–40.

van Bogelen, R.A., Acton, M.A., and Neidhardt, F.C. (1987) Induction of the heat shock regulon does not produce thermotolerance in *Escherichia coli*. *Genes Devel* **1:** 525–535.

Walker, G. (1987) In Escherichia coli *and* Salmonella typhimurium. Neidhardt, F.C. (ed.). Washington, D.C.: American Society for Microbiology, pp. 1346–1357.

Wanner, B.L. (1983) Overlapping and separate controls on the phosphate regulon in *Escherichia coli* K12. *J Mol Biol* **166:** 283–308.

Wanner, B.L. (1987) In Escherichia coli *and* Salmonella typhimurium. Neidhardt, F.C. (ed.). Washington, D.C.: American Society for Microbiology, pp. 1326–1333.

Wanner, B.L., and McSharry, R. (1982) Phosphate-controlled gene expression in *Escherichia coli* K12 using Mu*d1*-directed *lacZ* fusions. *J Mol Biol* **158:** 347–363.

The activity of σ^E, an *Escherichia coli* heat-inducible σ-factor, is modulated by expression of outer membrane proteins

Joan Mecsas, Pierre E. Rouviere,[1] James W. Erickson,[2] Timothy J. Donohue, and Carol A. Gross[1,3]

Department of Bacteriology, University of Wisconsin–Madison, Madison, Wisconsin 53706 USA

σ^E and σ^{32} are two heat- and ethanol-inducible σ-factors in *Escherichia coli*. The σ^{32} regulon is also induced by unfolded and misfolded proteins in the cytoplasm, and the function of many of the proteins in the σ^{32} regulon is to bind to cytoplasmic proteins and assist them in folding or unfolding. To further understand the function of the σ^E regulon, we searched for mutants that affected σ^E activity. Our results indicate that a signal generated by expression of outer membrane proteins modulates σ^E activity. Specifically, σ^E activity is induced by increased expression of OMPs and is reduced by decreased expression of OMPs. In addition, mutations that cause misfolded OMPs induce σ^E activity. This signal is generated after the fate of OMPs and periplasmic proteins diverge in the secretory pathway and is not the result of an accumulation of OMP precursors in the cytoplasm. Our results indicate that this effect of OMPs is specific to the σ^E regulon, because none of the above mutations affect σ^{32} activity. We propose that the σ^E regulon is involved in processes that occur in extracytoplasmic compartments and that these two heat-inducible regulons may have distinct but complementary roles of monitoring the state of proteins in the cytoplasm (σ^{32}) and outer membrane (σ^E).

[*Key Words*: σ-Factors; protein export; outer membrane proteins; heat shock; σ^E]

Received September 13, 1993; revised version accepted October 14, 1993.

In bacterial cells the σ-subunit directs RNA polymerase to initiate transcription at promoter sites on the DNA (Burgess et al. 1969). The primary σ-factor in the cell is responsible for transcription of most genes during exponential growth. In addition, alternative σ-factors direct transcription of sets of genes whose products are needed for specific functions, such as sporulation, nitrogen fixation, or flagella synthesis (Gross et al. 1992). Alternative σ-factors are often activated by changes in environmental or cellular conditions that generate morphological and/or molecular cues, signaling the need for the gene products in the regulon under control of a particular σ-factor. Elucidation of these signal-transduction pathways provides insights about global control of gene activity in prokaryotic cells.

The activity of two *Escherichia coli* alternative σ-factors, σ^{32} and σ^E (σ^{24}), increases after temperature upshift or exposure to ethanol (Grossman et al. 1984; Erickson et al. 1987; Straus et al. 1987; Erickson and Gross 1989; Wang and Kaguni 1989). RNA polymerase (E) containing σ^{32} ($E\sigma^{32}$) transcribes the heat shock genes with products that consist primarily of chaperones and proteases. These heat shock proteins function to process partially folded proteins in the cytoplasm and target them for secretion, proteolysis, or refolding (Straus et al. 1988; Hoffmann et al. 1992; Martin et al. 1992; Wild et al. 1992a); to aid in assembly and disassembly of complex protein structures (Alfano and McMacken 1989; Dodson et al. 1989; Zylicz et al. 1989); and to protect the cell from severe stresses (i.e., 10% ethanol and 50°C) (Neidhardt and VanBogelen 1987). Interestingly, under conditions of severe stress, only $E\sigma^E$ transcribes *rpoH* (Erickson et al. 1987), the gene encoding σ^{32}. Because σ^{32} is unstable (Straus et al. 1987), continued transcription of *rpoH* by $E\sigma^E$ is probably essential for cell survival under these conditions. Thus, σ^E was initially thought to be a supplementary heat shock σ-factor that functioned to maintain high concentrations of σ^{32} during severe conditions.

The only other *E. coli* gene known to be transcribed by $E\sigma^E$ is *degP* (*htrA*) (Erickson and Gross 1989), which encodes a periplasmic endopeptidase essential at temperatures above 42°C (Lipinska et al. 1988, 1989, 1990; Strauch and Beckwith 1988; Strauch et al. 1989). Although the physiological substrates of DegP have not been identified, it can degrade colicin A (Cavard et al. 1989). Because $E\sigma^E$ is the only form of RNA polymerase that transcribes *degP*, the requirement for DegP at 42°C is circumstantial evidence that σ^E plays an essential role in protecting cells from stress caused by high tempera-

Present addresses: [1]Department of Oral Biology, University of California, San Francisco, San Francisco, California 94143 USA; [2]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720 USA.
[3]Corresponding author.

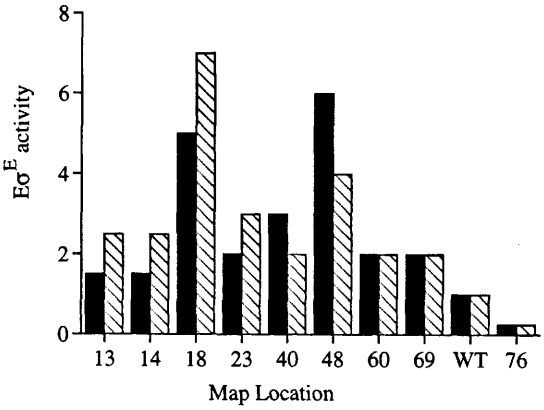


Figure 1. Changes in $E\sigma^E$ activity at the *rpoHP3* and *degP* promoters resulting from pISE plasmids or mini-Mu insertion (*dse*). $E\sigma^E$ activity at *rpoHP3* was assayed by monitoring β-galactosidase activity from a single copy Φλ[*rpoHP3–lacZ*] fusion (solid bars). $E\sigma^E$ activity at the *degP* promoter was assayed by S1 mapping of RNA (hatched bars). Assays were done in the CAG16052 strain background for the pISE strains (map positions 13, 14, 18, 23, 40, 48, 60, and 69) and in MC1061Φλ[*rpoHP3–lacZ*] for the *dse* allele (map position 76). The wild-type values were obtained from CAG16052 + pEG5005 for the pISE strains and MC1061Φλ-[*rpoHP3–lacZ*] for the *dse* strain. All values were normalized to wild type and are averages of measurements made at least twice.

tures. However, the periplasmic location of DegP suggests that the $E\sigma^E$ regulon may protect extracytoplasmic compartments from stress.

The experiments in this paper focus on identifying the signals that modulate $E\sigma^E$ activity. We find that the activity of $E\sigma^E$ increases when outer membrane proteins (OMPs) are overproduced and decreases when they are underproduced. Additional experiments suggest that the signal generated by OMPs is likely to originate in the periplasm or at the outer membrane (OM). This signal, unlike heat or ethanol, does not affect $E\sigma^{32}$ activity. We suggest that σ^E responds to signals in the extracytoplasmic compartments of the bacterial cell.

Results

Identification of positive regulators of $E\sigma^E$ activity

To elucidate the role of σ^E in *E. coli* cell physiology, we used two genetic strategies to identify potential positive regulators of $E\sigma^E$ activity. The first strategy was predicated on the rationale that increasing the copy number of genes encoding positive regulators of $E\sigma^E$ will increase $E\sigma^E$ activity. The second strategy was based on the rationale that loss-of-function mutations in positive regulators will decrease $E\sigma^E$ activity. To monitor $E\sigma^E$ activity, gene fusions containing only 37 bases of the *rpoHP3* promoter were used to express various reporter genes. Thus, changes in transcription from these fusions should result solely from changes in $E\sigma^E$ activity at the promoter.

To identify genes that increased $E\sigma^E$ activity when present in multicopy, a random plasmid library of *E. coli* DNA (Groisman and Casadaban 1986) was introduced into a reporter strain in which expression of chloramphenicol acetyltransferase is driven from *rpoHP3*. Cells growing on higher concentrations of chloramphenicol than the parental strain could contain plasmids that increased sigma E activity (pISE). The inserts on those plasmids were mapped to nine regions of the *E. coli* genome (Table 1) by hybridization to the Kohara phage gene bank (Kohara et al. 1987). Because most of the DNA regions were isolated only once, more genes may exist that increase $E\sigma^E$ activity when cloned on a multicopy plasmid. The increase in $E\sigma^E$ activity for one representative of each region was verified and quantified by examining expression from two different $E\sigma^E$ promoters. Plasmids caused a 1.5- to 7-fold increase in $E\sigma^E$ activity, and, in general, the increase in expression at *rpoHP3* and *degP* were similar (Fig. 1).

To identify genes that decreased $E\sigma^E$ activity when inactivated, transposon mutagenesis with mini-Mu (Castilho et al. 1984) was performed in a reporter strain in which expression of *galK* is driven from *rpoHP3*. Cells with normal expression of $E\sigma^E$ are red on MacConkey–galactose plates, whereas those with decreased $E\sigma^E$ activity should be white. In addition, cells with decreased $E\sigma^E$ activity should be temperature sensitive as a result of decreased transcription of *degP*, which is required for growth at high temperature. This screen identified one locus, *dse* (decreased sigma E), mapping to 76 min by P1 transduction and to Kohara phage 619–621 by hybridization (Table 1). The *dse* insertion caused a fourfold decrease in expression at both the *rpoHP3* and *degP* promoters (Fig. 1).

Table 1. *Mapping of loci altering $E\sigma^E$ activity*

Locus[a]	Number of isolates	Kohara phage[b]	Known genes[c,d]
ise13	1	160–162	***ompT***, *envY*, *appY*
ise14[e]	1	168	*lipA*
ise18	12	205–206	***ompX***[f]
ise23	1	225–226	
ise40	1	335–336	*htpX*, *prc*
ise48	2	373	***ompC***
ise60	1	458–459	*argA*
ise69	1	518–519	*mtrA*, *deaD*, *pnp*, *yhb*
dse76	16	619–621	***ompR***, ***envZ***, *bioH*, *pckA*

[a]The number indicates the map position of the locus on *E. coli* chromosome; (*ise*) increased σ^E activity; (*dse*) decreased σ^E activity.
[b]Plasmids were mapped to Kohara phage (see Materials and methods)
[c]Genes mapping to these phage were identified using GeneScape (Bouffard et al. 1992).
[d]Genes in boldface type encode proteins that are located in the outer membrane, or positive regulators of OMPs.
[e]Cells containing pISE14 were mucoid.
[f]*ompX* is described in J. Mecsas, R. Welch, J.W. Erickson, and C.A. Gross (in prep).

Overproduction of OMPs increases Eσ^E activity

A number of loci were identified that increased Eσ^E activity when present in multicopy. We initially focused on characterizing the insert on plasmid pISE18 because this region of the genome was isolated independently 12 times and gave one of the largest (five- to sevenfold) increases in Eσ^E activity. DNA sequence analysis of the region responsible for this phenotype indicated that the gene causing the increase in Eσ^E activity encodes an OMP, OmpX (J. Mecsas, R. Welch, J.W. Erickson, and C.A. Gross, in prep.). Because several other pISEs contained inserts that mapped to regions known to encode OMPs (Table 1), it seemed possible that overproducing a variety of OMPs could induce Eσ^E activity. To test this, multicopy plasmids encoding OM porins OmpF or OmpC, the OM protease OmpT, or OmpX were introduced into cells containing Φλ[*rpoHP3–lacZ*]. Overproduction of OmpC, OmpF, or OmpX caused an increase in Eσ^E activity at both *rpoHP3* and the *degP* promoters (Fig. 2A,B). Furthermore, Figure 3 shows that these multicopy plasmids caused the expected increase of these proteins in the OM. Overproduction of OmpT, which was tested only at the *rpoHP3* promoter, also caused a small but reproducible increase in Eσ^E activity (Fig. 2A), whereas pBR322 had no effect on Eσ^E activity. This suggests that overexpression of a variety of OMPs induces Eσ^E activity and that pISE13 and pISE48 were identified because they encode OmpT and OmpC, respectively.

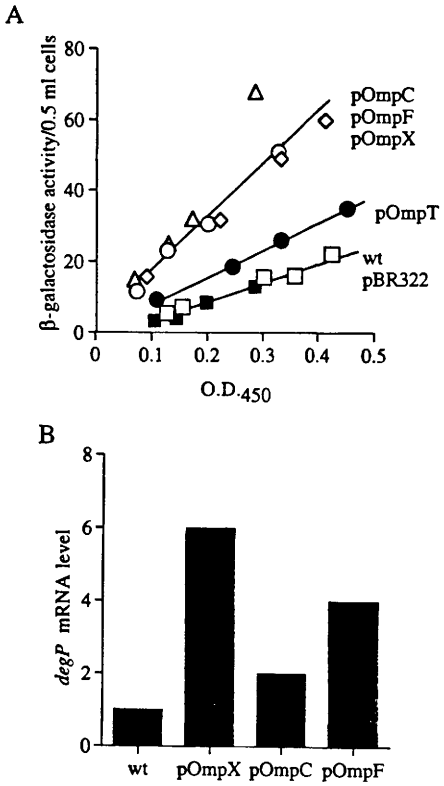


Figure 2. (*A*) Eσ^E activity in strains overexpressing OMPs from multicopy plasmids. β-Galactosidase activity in the parental strain (□) was compared with strains containing either pMY111 (OmpC, △), pMY222 (OmpF, ○), pJE100 (OmpX, ◇), pML21 (OmpT, ●), or pBR322 (■). The data shown are representative of at least three experiments. The increase in *rpoHP3* activity was 3- to 8-fold for strains overexpressing OmpC, OmpF, and OmpX and 1.4- to 2-fold for the strain overexpressing OmpT. The r^2 values (variance) for a linear fit of the points from each strain are all >0.975 for the data in this experiment. (*B*) *degP* mRNA levels in MC1061Φλ[*rpoHP3–lacZ*] and isogenic strains overexpressing OmpC, OmpF, or OmpX as assayed by S1 mapping. RNA was purified from cells grown at 30°C and harvested at an OD_{450} of 0.3–0.4. Values were normalized to wild type and are averages of two to five measurements.

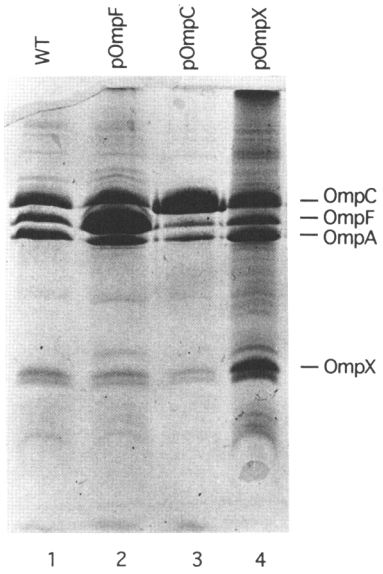


Figure 3. OMP profiles of MC1061Φλ[*rpoHP3–lacZ*] cells and isogenic strains containing multicopy plasmids expressing OMPs. Equivalent amounts of cells were loaded in each lane. (Lane *1*) Proteins from wild-type cells; (lane *2*) pOmpF (pMY222); (lane *3*) OmpC (pMY111); (lane *4*) OmpX (pJE100).

Underproduction of OMPs decreases Eσ^E activity

In the screen designed to generate loss-of-function mutations that caused a decrease in Eσ^E activity, one locus was identified 16 independent times (Table 1). To determine the gene inactivated by the mini-Mu transposon, one transposon with flanking sequences was cloned. Restriction mapping indicated that the insertion was located in *ompR*, a transcriptional activator of two major OM porins, OmpC and OmpF (Forst et al. 1988; Mizuno and Mizushima 1990). As expected, isolation of OMs from cells containing this insertion showed little OmpC and OmpF (data not shown). We considered two possible explanations for the decrease in Eσ^E activity of these mutants. Eσ^E activity could be responding to the decreased amount of OMPs present in the strain. Alternatively, Eσ^E activity could depend on a pathway that is dependent on OmpR transcriptional activity. These two possibilities can be distinguished by overexpressing OmpX in a Δ*ompR* strain. According to the first hypothesis, Eσ^E activity will be high, whereas in the second, it

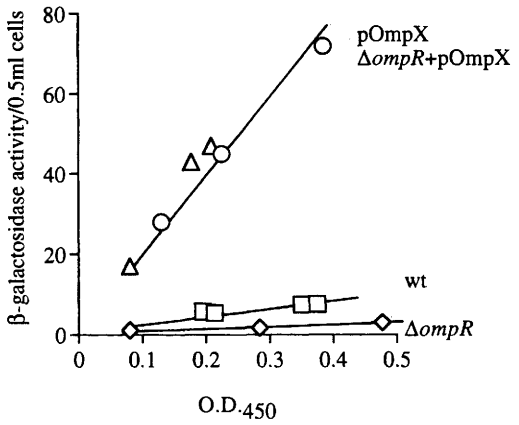


Figure 4. Overexpression of OmpX induces $E\sigma^E$ activity independent of *ompR*. β-Galactosidase activity in MC1061Φλ-[*rpoHP3–lacZ*] (wt, □) was compared with isogenic strains containing Δ*ompR* (◇), overproducing OmpX (△), or overproducing OmpX in a Δ*ompR* background (○). β-Galactosidase activity in Δ*ompR* strain was three- to five-fold lower than the wild-type strain, and activity in the Δ*ompR* + pJE100 strain was three- to six-fold higher than wild type in three experiments (r^2>0.96).

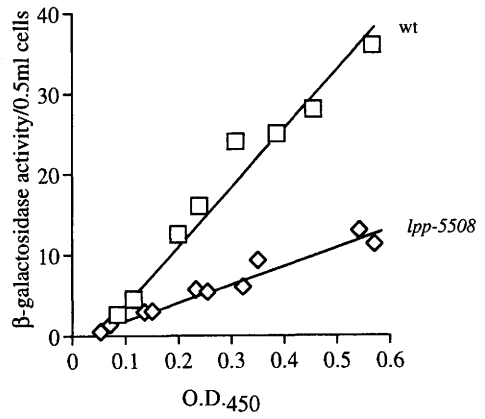


Figure 5. β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] (wt, □) compared with an isogenic strain containing *lpp5508* (◇). The difference in β-galactosidase activity ranged from three- to fivefold in two experiments (r^2>0.95).

will remain low. As shown in Figure 4, $E\sigma^E$ activity is induced by overproduction of OmpX in a Δ*ompR* background. Thus, OmpR is not a transcriptional activator for $E\sigma^E$. Instead, $E\sigma^E$ activity appears to correlate with the amount of OMPs produced.

We asked whether decreasing the levels of other OMPs also affected $E\sigma^E$ activity. Lpp is the major OM lipoprotein in *E. coli* (Hirota et al. 1977). In an *lpp5508* background Lpp is undetectable, and the *E. coli* OM is less stable and more sensitive to low amounts of detergents and chelating reagents. $E\sigma^E$ activity was decreased in an *lpp5508* background (Fig. 5), supporting the idea that $E\sigma^E$ activity is reduced in cells with decreased OMP content.

Cellular localization of signal-inducing $E\sigma^E$ activity by overexpression of OMPs

To further understand how production of OMPs affects $E\sigma^E$ activity, the location of the signal generated by overproducing OMPs was determined. First, the location of the bulk of the overproduced OmpX was determined by fractionation of the cell into cytoplasmic, inner membrane, periplasm, and OM components. As expected, all detectable OmpX was in the OM (Fig. 6). This result, however, does not exclude the possibility that an undetected pool of OmpX was mislocalized or jammed in the secretion apparatus and, thus, caused the increase in $E\sigma^E$ activity. The following experiments were done to test these possibilities.

If OMP precursors in the cytoplasm generate a signal that induces $E\sigma^E$ activity, then cells that accumulate OMP precursors in the cytoplasm should have a high level of $E\sigma^E$ activity. In fact, the converse was observed. In a *secB* deletion strain, partially folded OMP precursors accumulate in the cytoplasm (Kumamoto 1991), but $E\sigma^E$ activity was reduced twofold (Fig. 7A). Thus, accumulation of OMP precursors in the cytoplasm does not induce $E\sigma^E$ activity.

Overproduction of OMPs could titrate components of the inner membrane secretion apparatus, thereby generating a signal to increase $E\sigma^E$ activity. Because periplasmic proteins and OMPs share some cytoplasmic and inner membrane components of the secretion apparatus, overproduction of periplasmic proteins should also titrate these components of secretion apparatus. $E\sigma^E$ activity, however, was unaffected when periplasmic proteins DegP or β-lactamase were overexpressed from high-copy-number plasmids (Fig. 7B). Western blot anal-

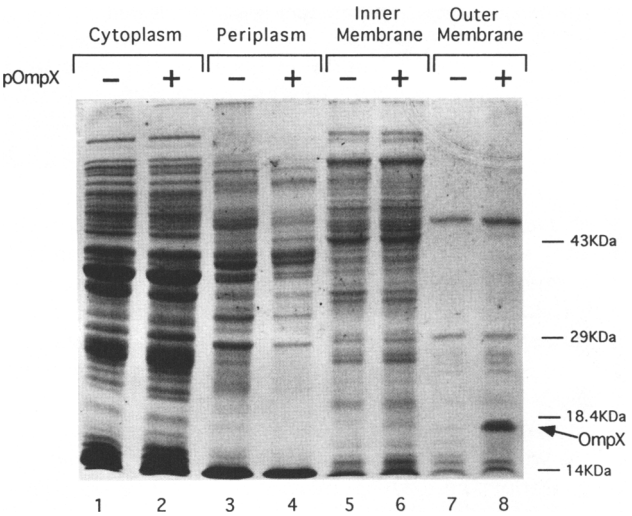


Figure 6. Protein profiles in different cell compartments in MC1061Φλ[*rpoHP3–lacZ*] wild-type cells and the otherwise isogenic strain overproducing OmpX (pJE100). Cells were fractionated as described in Materials and methods. Cytoplasmic (lanes *1,2*), periplasmic (lanes *3,4*), inner membrane (lanes *5,6*), and OMPs (lanes *7,8*), from wild type (lanes *1,3,5,7*) and cells containing pJE100 (lanes *2,4,6,8*) were electrophoresed on a SDS–polyacrylamide gel. The OmpC and OmpF proteins were not resolved because this gel did not contain urea.

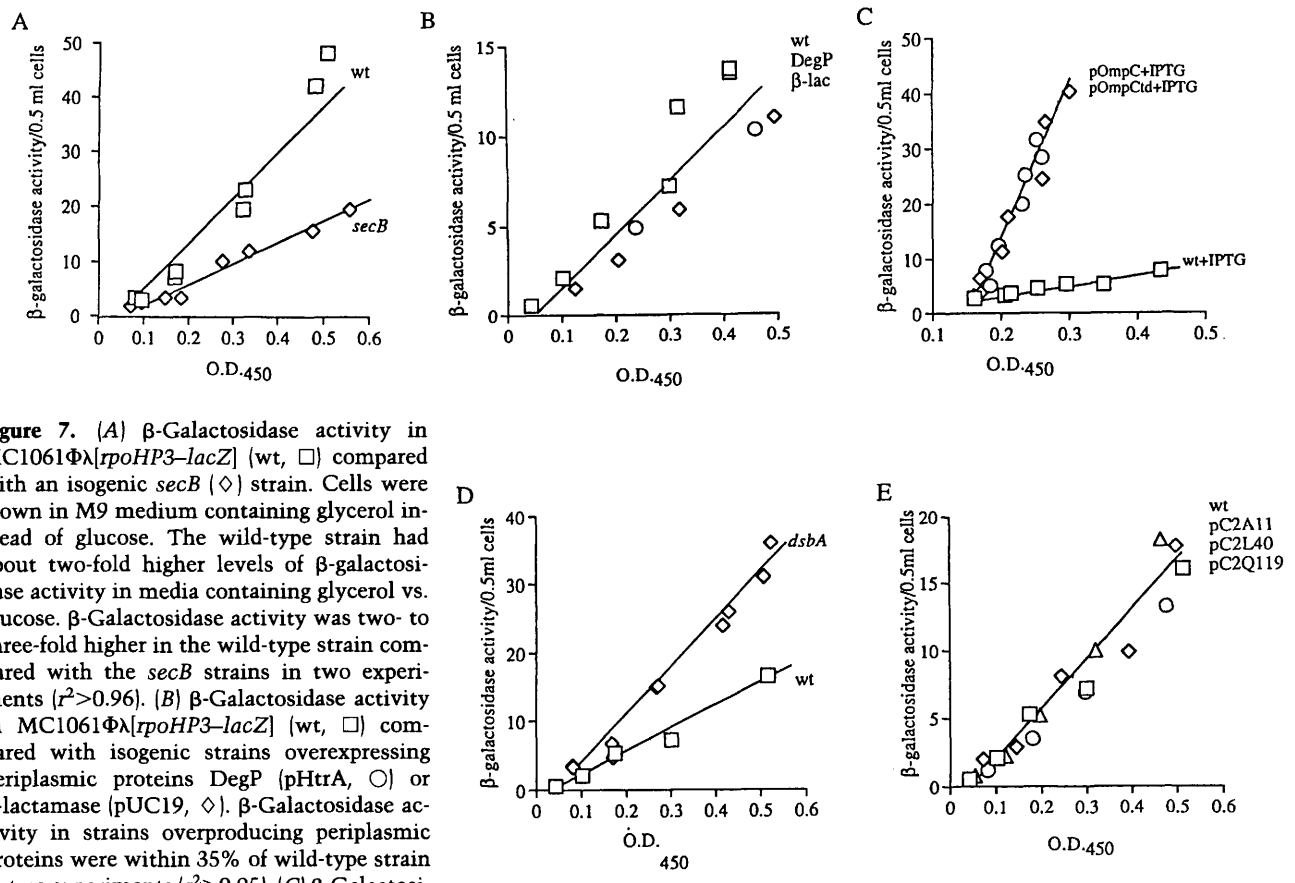


Figure 7. (*A*) β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] (wt, □) compared with an isogenic *secB* (◇) strain. Cells were grown in M9 medium containing glycerol instead of glucose. The wild-type strain had about two-fold higher levels of β-galactosidase activity in media containing glycerol vs. glucose. β-Galactosidase activity was two- to three-fold higher in the wild-type strain compared with the *secB* strains in two experiments (r^2>0.96). (*B*) β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] (wt, □) compared with isogenic strains overexpressing periplasmic proteins DegP (pHtrA, ○) or β-lactamase (pUC19, ◇). β-Galactosidase activity in strains overproducing periplasmic proteins were within 35% of wild-type strain in two experiments (r^2>0.95). (*C*) β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] (wt, □) compared with isogenic strains overexpressing IPTG-inducible OmpC (pGMC1, ○) or OmpCtd (pKMCtd, △). Strains were grown to an OD_{450} of 0.15, and IPTG was added to a final concentration of 1 mM. After addition of IPTG, the rate of β-galactosidase synthesis in strains overexpressing OmpC and OmpCtd increased 10- to 15-fold compared with the wild-type strain in three experiments. Growth of the pOmpCtd overproducing strain stopped 35 min after IPTG addition, whereas growth of the OmpC overproducing strain slowed immediately after IPTG addition (r^2>0.91). (*D*) β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] (wt, □) compared with an isogenic strain containing *dsbA* (◇). The *dsbA* strain had 1.5- to 2.5-fold higher β-galactosidase activity than the wild-type strain in three experiments (r^2>0.97). (*E*) β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] (wt, □) compared with isogenic strains containing CycA–AP fusion proteins, pC2A11 (△), pC2L40 (○), and pC2Q119 (◇). β-Galactosidase activity in strains expressing CycA–AP fusion proteins were within 25% of the wild-type strain (r^2>0.93).

ysis of DegP confirmed that expression from a multicopy plasmid increased the level of both the precursor and processed forms of DegP (data not shown). In addition, overproduction of periplasmic proteins MalE and alkaline phosphatase had no effect on $E\sigma^E$ activity (data not shown). These experiments suggest that overproduction of OMPs does not generate a signal for $E\sigma^E$ activity by titrating components of the secretion apparatus that are shared with periplasmic proteins. Furthermore, $E\sigma^E$ activity is not induced by increases in periplasmic proteins.

After transport through the inner membrane, the secretion pathways of periplasmic and OMPs diverge. Therefore, overproduced periplasmic proteins may not be in the same environment as overproduced OMPs. To specifically ask whether $E\sigma^E$ activity is induced by OMPs that are mislocalized in the periplasm or jammed in the secretion apparatus, $E\sigma^E$ activity was examined in a strain containing an OmpC variant that lacks 2 amino acids in the mature portion of the protein, OmpCtd (Catron and Schnaitman 1987). This variant is secreted through the inner membrane but is poorly incorporated in the OM and inhibits incorporation of wild-type OmpC, OmpF, and OmpA into the OM. Induction of OmpCtd increases the activity of $E\sigma^E$ 10- to 15-fold (Fig. 7C). Thus, $E\sigma^E$ activity is induced under conditions that slow down the processing of OMPs after they have passed the inner membrane but before insertion in the OM.

Overproducing OMPs could cause them to aggregate or fold more slowly, resulting in accumulation of misfolded proteins. The *E. coli dsbA* gene product was shown recently to facilitate disulfide bond formation of proteins in both the OM (OmpA) and the periplasmic space (AP) (Bardwell et al. 1991). $E\sigma^E$ activity at the *rpoHP3* promoter in a Δ*dsbA* background was induced about twofold (Fig. 7D). A similar effect on a *degP–lacZ* fusion has been observed by P. Danese and T. Silhavy

(pers. comm.). This suggests that $E\sigma^E$ activity may be induced by misfolded OMPs or periplasmic proteins.

Finally, to test whether misfolded periplasmic proteins induce $E\sigma^E$ activity, $E\sigma^E$ activity was tested in strains expressing protein fusions between portions of *Rhodobacter sphaeroides* CycA, a periplasmic *c*-type cytochrome (not found in *E. coli*), and alkaline phosphatase. These protein fusions are exported to the *E. coli* periplasm, and the CycA portion is then rapidly degraded. This indicates that these hybrid proteins do not fold properly (Brandner et al. 1991). Expression of these misfolded periplasmic proteins did not alter $E\sigma^E$ activity (Fig. 7E).

Altering levels of the DegP protease has no effect on $E\sigma^E$ activity

Results with the Δ*dsbA* strain suggested that $E\sigma^E$ activity may be induced by misfolded OMPs. Although the substrates for the periplasmic DegP protease are unknown, these could include misfolded proteins or other molecules involved in signaling $E\sigma^E$ activity. Two experiments were designed to determine whether DegP levels, and presumably activity, influenced $E\sigma^E$ activity. A Δ*degP* strain should accumulate DegP substrates; however, $E\sigma^E$ activity was unaffected (Fig. 8A). This indicates that either accumulation of DegP substrates does not induce $E\sigma^E$ activity or that other uncharacterized periplasmic proteases compensate for the lack of DegP.

Overproduced OMPs that induce $E\sigma^E$ activity could also be potential substrates for DegP. Thus, an increase in DegP in an OmpC overproducing strain might decrease $E\sigma^E$ activity. However, $E\sigma^E$ activity remained high in a strain with both DegP and OmpC encoded by multicopy plasmids (Fig. 8B). Thus, the signal generated by overproducing OmpC is not directly affected by the level of DegP protease.

OMP levels do not affect expression from $E\sigma^{32}$ promoters

Environmental stresses known to induce $E\sigma^E$ activity (temperature upshift and exposure to ethanol) also induce $E\sigma^{32}$ activity. It was not clear, however, whether all conditions that increase $E\sigma^E$ activity also increase $E\sigma^{32}$ activity. To test this, $E\sigma^{32}$ activity was assayed in a strain containing the *rpoDpHS* promoter fused to *lacZ* on a prophage (Wild et al. 1992) under conditions that alter OMP levels. Neither the plasmids increasing OMP levels nor chromosomal mutations in *ompR* or *dsbA* had any effect on $E\sigma^{32}$ activity (Fig. 9A). As reported previously, a *secB* mutation induced $E\sigma^{32}$ activity fourfold (Wild et al. 1992a), presumably because a buildup of partially folded proteins in the cytoplasm activates the σ^{32} regulon (Fig. 9B). Thus, changes in the levels of OMPs affects only the $E\sigma^E$ regulon.

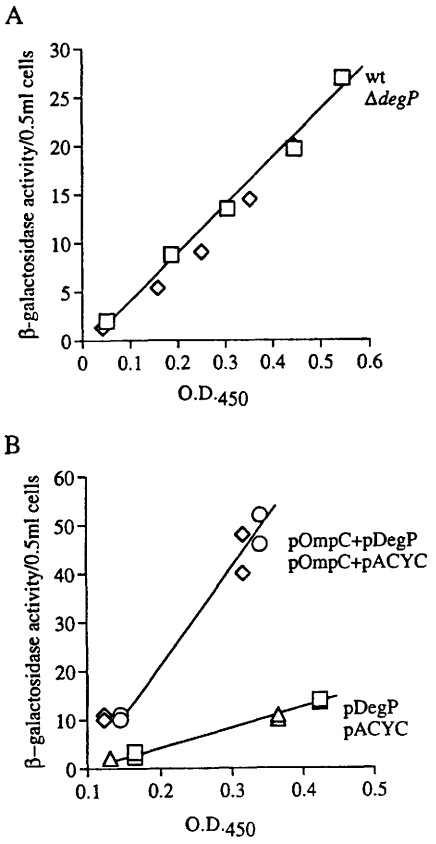


Figure 8. (*A*) β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] (wt, □) compared with an isogenic strain containing *degP* (◇). β-Galactosidase activity of these two strains was within 10% in three experiments (r^2>0.98). (*B*) β-Galactosidase activity in MC1061Φλ[*rpoHP3–lacZ*] isogenic strains overexpressing OmpC (pMY111, ◇), OmpC, and DegP (pMY111 and pKS17, ○) compared with cells overexpressing DegP (pSK17, △), or wild-type cells (pACYC184, □). β-Galactosidase activity between cells overproducing OmpC or OmpC and DegP was within 10%; these cells had four- to sixfold more β-galactosidase activity than cells overproducing DegP alone or wild-type cells (r^2>0.97).

Discussion

We describe the results of experiments designed to identify positive regulators of $E\sigma^E$ activity in *E. coli*. Initially, two genetic strategies were used to allow identification of a broad range of regulators because each approach has limitations. The screen for loss-of-function mutants fails to identify regulators that are essential, are not required for growth at high temperature, or have redundant functions. The multicopy selection fails to identify genes that cannot be cloned on multicopy vectors or gene products with activity that does not increase with increasing gene dosage. The initial results from these complementary approaches indicate that $E\sigma^E$ activity changes when the level of OMPs is altered. On the basis of these observations, we showed that increasing the amount of OmpX, OmpT, OmpF, or OmpC increases $E\sigma^E$ activity. Furthermore, mutations that increase the amount of misfolded OMPs or reduce the efficiency of OMP incorporation into the OM induce $E\sigma^E$ activity. Decreasing specific OMPs, Lpp or OmpF and OmpC, or decreasing

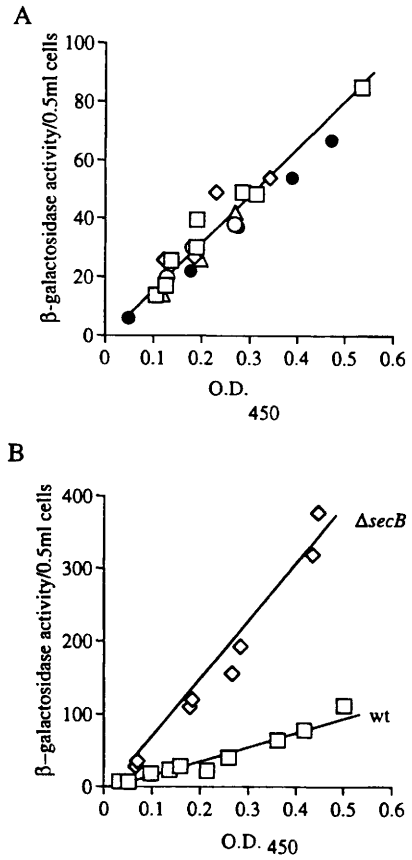


Figure 9. (*A*) β-Galactosidase activity in MC1061Φλ[*rpoDpHS–lacZ*] (wt, □) compared with isogenic strains overexpressing OmpC (pMY111, ◇) or OmpX(pJE100, △), or with insertions in *ompR* (○) or *dsbA* (●). Differences in β-galactosidase activity among these strains were within 20% (r^2>0.95). (*B*) β-Galactosidase activity in MC1061Φλ[*rpoDpHS–lacZ*] (wt, □) compared with an isogenic strain containing *secB* (◇). The *secB* strain had four-fold higher β-galactosidase activity than the wild-type strain in two experiments (r^2>0.94).

OMP levels indirectly by SecB limitation reduces $E\sigma^E$ activity. Thus, one signal that modulates $E\sigma^E$ activity is linked to production of OMPs.

Where in the cell do overproducing OMPs generate a signal?

Although all of the detectable overproduced wild-type OMPs are in the OM, the signal need not originate there. OMPs transverse several cellular compartments before insertion into the OM (for review, see Pugsley 1993). They are synthesized as precursor proteins containing an amino-terminal signal sequence, maintained in a translocation-competent form by interaction with cellular chaperones such as SecB, targeted to the membrane by interaction with SecA, transported through the cytoplasmic membrane by the SecY/E translocase, and cleaved by a signal peptidase as they exit the secretion machinery. These steps are common to both periplasmic proteins and OMPs. OMPs are then localized to the OM either through a periplasmic intermediate (Sen and Nikaido 1990) or by direct transfer to the OM at attachment sites (Bayer et al. 1982). This latter process is not well understood. The signal generated by gross changes in production of OMPs could be located in any of these compartments.

Our results suggest that the signal controlling $E\sigma^E$ activity is not generated during the early steps in the translocation of OMPs. A Δ*secB* strain accumulates precursors of periplasmic proteins and OMPs in the cytoplasm (Kumamoto 1991) but does not induce $E\sigma^E$ activity. This suggests that the signal generated by overproducing OMPs does not result from accumulation of cytoplasmic precursor proteins or from *secB* limitation. Overexpression of proteins normally located in the periplasm also does not induce $E\sigma^E$ activity. Because periplasmic proteins use the same translocase as OMPs (Pugsley 1993), it is unlikely that overproduced OMPs generate a signal by titrating the translocase. Finally, expression of foreign proteins localized to the periplasm or accumulation of DegP substrates does not alter $E\sigma^E$ activity. This suggests that a signal is not generated by aberrant periplasmic proteins. However, overexpressing OmpCtd (Catron and Schnaitman 1987), an aberrant OMP that is translocated but poorly incorporated into the OM, does induce $E\sigma^E$. On the basis of these observations, we propose that the signal for $E\sigma^E$ originates after the translocation pathways of OMP and periplasmic proteins diverge. The signal could be generated as the OMP exits the cytoplasmic membrane, in the periplasm, or in the OM. The extracellular location of the signal necessitates that a signal-transduction pathway exists to propagate the signal back through the inner membrane to influence the activity of $E\sigma^E$ in the cytoplasm.

What is the signal?

Although we have not quantitatively measured OMPs, it is known that the total amount of OMPs in the membrane can change by as much as 30% (Diedrich and Fralick 1982). Our data suggest that the total amount of OMPs produced directly correlates with $E\sigma^E$ activity. We consider two different ways in which varying the amount of OMP production could generate a signal that modulates $E\sigma^E$ activity. First, changing the amount of OMP production could change the composition of the OM. Alternatively, if we assume that a change in OMP levels reflects a change in their rate of synthesis, a change in the flow rate of OMPs through the system that targets OMPs to the OM could generate a signal.

In the first scenario, the composition of the OM could vary by changing either the absolute amount or the relative ratios of individual OMPs. Altering the absolute amount of OMPs could change the ratio of protein to either lipid or peptidoglycan and generate a signal, possibly by changing the structural integrity or fluidity of the cell envelope. However, to the extent that we have assayed this, $E\sigma^E$ is not sensing structural integrity of the cell envelope. Both the *lpp5508* mutation (Hirota et al. 1977) and overproduction of OmpX, OmpC, and OmpF

(data not shown) result in increased sensitivity to detergents and chelating reagents, yet these alterations have opposite effects on $E\sigma^E$ activity. Alternatively, the signal could be generated by a change in relative ratios of individual OMPs. Overproducing some OMPs leads to underrepresentation of other OMPs in the OM (Click et al. 1988). If one OMP negatively regulates $E\sigma^E$ activity, then overproducing other OMPs could decrease its level whereas underproducing other OMPs could increase its level. Although such a protein may exist, we were unable to identify one protein that varied consistently with $E\sigma^E$ activity (data not shown). For instance, OmpA increases in the Δ*ompR* strain and decreases in a Δ*secB* strain, yet both strains had decreased $E\sigma^E$ activity.

A second way of generating the signal modulating $E\sigma^E$ activity can be derived by considering the pathway that targets OMPs to the OM (Bayer et al. 1982; Sen and Nikaido 1990; Pugsley 1993). OMPs must be sorted from periplasmic proteins, transported to the OM, and folded properly. Although the specific components of this pathway are poorly understood, the pathway is likely to involve chaperones to maintain OMPs in appropriate protein conformations, molecules that target OMPs to the OM, and possibly molecules that insert OMPs in the membrane. A change in the flow rate of OMPs through this pathway could generate the signal modulating $E\sigma^E$ activity. For example, a chaperone involved in the transport of proteins to the OM could also interact with the signal-transduction pathway modulating $E\sigma^E$ activity. Overproduction of OMPs could titrate or alter the activity of this chaperone and signal for an increase in $E\sigma^E$ activity, whereas underproduction of OMPs would have the converse effect.

Why are both the $E\sigma^E$ and $E\sigma^{32}$ regulons induced by heat?

We believe that the answer to this question arises from a consideration of signals that uniquely induce each regulon. Conditions that increase the amounts of unfolded, partially folded, or misfolded proteins in the cytoplasm induce $E\sigma^{32}$ but not $E\sigma^E$. Conversely, conditions that increase OMPs in compartments outside the cytoplasm induce $E\sigma^E$ but not $E\sigma^{32}$. Two distinct regulons may exist to allow cells to respond to processes or signals that are unique to each compartment. Thus, just as the products of the $E\sigma^{32}$ regulon are involved in responding to environmental stimuli affecting the processing of cytoplasmic proteins, the products of the $E\sigma^E$ regulon could be involved in responding to environmental stimuli affecting the OM or the processing of OMPs. Although $E\sigma^E$ transcribes *rpoH*, transcription from the $E\sigma^E$ promoter accounts for <5% of total transcription of *rpoH* at 30°C (Erickson et al. 1987) Furthermore, $E\sigma^{32}$ activity is posttranscriptionally regulated (Straus et al. 1987, 1989). Thus, increases in $E\sigma^E$ activity do not necessarily increase $E\sigma^{32}$ activity at moderate temperatures.

In Gram-negative bacteria, the composition of the OM changes in response to many environmental conditions including temperature, osmolarity, dessication, starvation, and growth in other hosts. The products of the $E\sigma^E$ regulon could be involved in responding to environmental changes that affect processing of OMPs. In addition to the periplasmic location of DegP, circumstantial evidence exists for an extracytoplasmic role of the products of the $E\sigma^E$ regulons. $E\sigma^E$ recognizes promoters for two secreted hydrolytic enzymes (Erickson and Gross 1989), *dagA* (Buttner et al. 1988) and *phlA1* (Givskov et al. 1988) from *Streptomyces coelicolor* and *Serratia liquefaciens*, respectively. If σ^E homologs in *S. coelicolor* and *S. liquefaciens* are involved in transcription of these genes, the $E\sigma^E$ regulon may consist of extracytoplasmic proteins that are involved in altering the protein composition of the OM, periplasm, or extracellular space. The *degP* homolog in *Salmonella* (*htrA*) plays a role in virulence and appears to have an $E\sigma^E$ promoter (Johnson et al. 1991). This suggests the exciting possibility that $E\sigma^E$ activity may be induced during growth in and colonization of animals.

An appealing, but speculative, idea is that these two σ-factors may have complementary but distinct functions; both may be involved in protein processing—the $E\sigma^{32}$ regulon in the cytoplasm and the $E\sigma^E$ regulon in the extracytoplasmic compartments. It is clear that $E\sigma^{32}$ is induced by misfolded proteins in the cytoplasm; our data indicate that $E\sigma^E$ may be induced by misfolded OMPs in other compartments. We and others (P. Danese and T. Silhavy, pers. comm.) have shown that the Δ*dsbA* mutation, which creates misfolded or partially folded proteins in the periplasm and OM, induces $E\sigma^E$. A second $E\sigma^E$ inducer, OmpCtd, may cause accumulation of misfolded or improperly localized OMPs in an extracellular compartment. Finally, we have argued that overproducing OMPs may induce $E\sigma^E$ activity by saturating the pathway that inserts OMPs into the OM, thus increasing the pool of incompletely folded proteins in the extracytoplasmic compartments of the cell.

That two *E. coli* σ-factors may be induced by similar molecular cues in different cellular compartments is reminiscent of the situation in eukaryotic cells where unfolded or damaged proteins in different compartments or organelles induce chaperones by parallel but independent mechanisms. Unfolded proteins in the cytoplasm induce the heat shock response, possibly by titrating out Hsp70 (Craig and Gross 1991). Unfolded proteins in the endoplasmic reticulum induce a set of proteins residing in this compartment by interacting with BIP, an Hsp70 family member found exclusively in the endoplasmic reticulum (Kozutsumi et al. 1988; Mori et al. 1992). Future experiments addressing the exact nature of the signal that modulates $E\sigma^E$ activity and how the signal is transduced back to the cytoplasm should provide biologists with insights into how different cell compartments respond to similar environmental or physiological cues.

Materials and methods

Media and strains

All media were prepared as described (Miller 1972). Liquid cultures were grown in M9 glucose supplemented with vitamins

and all 20 common amino acids, except for experiments involving the *secB::kan* allele. In those experiments, both wild-type and *secB::kan* strains were grown in M9 glycerol supplemented with vitamins and all 20 common amino acids. Solid media used were MacConkey lactose, MacConkey galactose, LB, and M9 glycerol supplemented with amino acids and vitamins. Drug concentrations were 50 μg/ml of ampicillin, 30 μg/ml of kanamycin (Kan), 10 μg/ml of tetracycline (Tet), and 20–75 μg/ml of chloramphenicol (Cam).

CAG16028 Δ(*araCOIBA*, *leu*)*7696*, *araD139*, *galK*, *galU*, *hsdR*, Δ(*lac*)*X74*, *recA56*, *rpsL*, a derivative of MC1061 was used for making the RNA for the S1 assays. Most β-galactosidase assays were done in CAG16028 lysogenized with either the Φλ[*rpoHP3–lacZ*] resulting in CAG16037 or Φλ[*rpoDpHS–lacZ*] resulting in CAG16074 (Wild et al. 1992). CAG16038 was a derivative of C600 *galK* containing pJEK61 (Erickson and Gross 1989). CAG16040 was a derivative of C600 *galK* containing plasmid with PgalK driving *galK* expression. For identification and initial characterization of the pISE plasmids, M8820 MuCts (Groisman and Casadaban 1986) was lysogenized with λΦ[*rpoHP3–lacZ*] generating CAG16045 and transformed with P3CAT generating CAG16052 to select for increased $E\sigma^E$ activity. The *secB::kan* allele (Kumamoto and Beckwith 1985) was introduced into CAG16037 and CAG16074 using P1 transduction. To facilitate manipulation of the *lpp5508* allele (Hirota et al. 1977), a Tn*10*tet (*zdh-925*::Tn*10*) was linked to this mutation (Singer et al. 1989). The linked Tet^R was then transferred to CAG16037 or CAG16074 by P1 transduction. The presence of the *lpp5508* allele was determined by increased sensitivity to SDS and EDTA (Hirota et al. 1977). The *dsbA::kan* (Bardwell et al. 1991) and *degP:: kan* (Strauch and Beckwith 1988) alleles were transferred to CAG16037 and CAG16074 by P1 transduction.

Plasmids and phage

The following plasmids were used to express the proteins indicated in parentheses and have been described previously: pKS17 (DegP) (Strauch and Beckwith 1988), pMY111 (OmpC) (Mizuno et al. 1983), pMY222 (OmpF) (Ramakrishnan et al. 1985), pML21 (OmpT) (Grodberg et al. 1988), pC2A11, pC2L40, and pC2Q119 (CycA–AP fusions) (Brandner et al. 1991), pGMC1 (P_{tac}-OmpC) and pKMCtd (P_{tac}-OmpCtd) (Catron and Schnaitman 1987), and pHtrA (DegP) (Lipinska et al. 1988). To make pP3CAT, which was used in the multicopy selection, a 37-bp fragment of the P3 promoter of *rpoH* (Erickson and Gross 1989) was cloned into pKK232-8 (Pharmacia), resulting in pJM14. A 2.4-kb *Hgi*AI fragment of pJM14 was cloned into the 2.8-kb *Bsa*A1 fragment of pACYC184 resulting in pP3CAT. pJEK61 (Erickson and Gross 1989) was used in the screen for loss-of-function mutants. The pISE plasmids and the parental pEG5005 plasmids were generated as described (Groisman and Casadaban 1986). pJE100 is described (J. Mecsas, R. Welch, J. Erickson, and C.A. Gross, in prep.). The Φλ[*rpoHP3–lacZ*] was constructed by cloning the *SalI–Eco*RI fragment of pJEK61 into pRS415 and crossing the resulting plasmid with λRS45 as described (Simons et al. 1987). Φλ[*rpoDpHS–lacZ*] was described (Wild et al. 1992b).

Genetic selections

To select for genes that increase $E\sigma^E$ activity when overexpressed, a multicopy library was made using a mini-Mu system (Groisman and Casadaban 1986). CAG16052 was infected with a high-titer lysate containing this library, and appropriate dilutions were plated on LB plates supplemented with Kan and 20–75 μg/ml of Cam. Approximately 7000 Kan^R colonies were plated on a total of four plates at each Cam concentration. Unexpectedly, the presence of the pEG5005 vector sequences caused an increase in the resistance of CAG16052 to Cam; thus, the efficiency of plating was higher than expected. Cells that grew faster than the CAG16052 + pEG5005 strain were picked: Eighteen colonies were picked on 20 μg/ml; 14 colonies were picked on 40 μg/ml; and 8 colonies were picked on 50 μg/ml and 75 μg/ml. Colonies were purified twice; plasmid DNA was isolated and used to transform CAG16045 to show that the increase in $E\sigma^E$ activity was linked to the plasmid. Of the original 48 colonies picked, 22 resulted in an increase in $E\sigma^E$ activity at both the *rpoHP3* and *degP* promoters and were mapped to the Kohara phage.

To generate cells with insertions in positive regulators of genes for $E\sigma^E$ activity, mini-Mu transposon mutagenesis was performed as described (Castilho et al. 1984) using the *Mud*II1734 transposon. CAG16038 was infected with a mini-Mu lysate, plated on MacConkey plates containing galactose and Kan, and grown at either 20°C or 30°C. White Kan^R colonies were patched onto MacConkey–galactose plates and grown at 30°C and 44°C. Mutants that grew poorly at 44°C were taken from the 30°C plate, colony purified, and tested for Mu^S. To test that the insertion specifically reduced $E\sigma^E$ activity, P1 was grown on the white, temperature-sensitive, Mu^S cells and the lysate was used to transduce CAG16038 or CAG16040 to Kan^R. Only Kan^R insertions that conferred the $galK^-$ phenotype in CAG16038 (which carries *galK* under $E\sigma^E$ control) but not CAG16040 (which carries *galK* under $E\sigma^{70}$ control) were studied further. Approximately 22,500 Kan^R cells were screened (8000 at 20°C and 14,500 at 30°C); 3% of the Kan^R cells were white; 1% of the white cells were temperature sensitive; 50% of the temperature sensitive cells were Mu^S; 16 colonies had mini-Mu insertions that gave rise to white, Kan^R colonies when transduced into CAG16038 and red when transduced into CAG16040. The location of one insertion was mapped genetically using high frequency recombination F-factors (Hfrs) and linked Tn*10*tet (Singer et al. 1989). All of the other insertions were linked to the same Tn*10*. To identify the gene into which the mini-Mu inserted, chromosomal DNA from one insertion strain was isolated and digested with either *Hin*dIII or *Bam*HI. These enzymes do not cleave the kan^R gene allowing for cloning of adjacent chromosomal DNA. Digested DNA was ligated to pUC19 with either *Hin*dIII or *Bam*HI and transformed into DH5α. Plasmids from Kan^R colonies were analyzed with various restriction enzymes and compared with the restriction map of the *E. coli* chromosome between 74 and 76 min using the GeneScape program (Bouffard et al. 1992). The restriction pattern matched that of the *ompB* locus and showed that the *Mud*II1734 transposon had inserted 100 bp after the translational start site of the *ompR* gene.

Mapping pISE inserts and location of dse

A nylon membrane containing a subset of 466 phage from the collection of Kohara (Kohara et al. 1987) was obtained from Takara Biochemical. The plasmids were nick-translated, and hybridization, washing, and stripping of the membrane were done as recommended by Takara Biochemical.

β-Galactosidase assays

β-Galactosidase levels were measured as described (Miller 1972). Overnight cultures were diluted 1/1000 or to an OD_{450} of 0.004 in the appropriate medium and grown at 30°C. Cells (0.5 ml) were lysed with chloroform and one drop of 0.1% SDS. For line graphs, β-galactosidase activity was calculated by

$([A_{420(final)} - A_{420(initial)}]/(\text{min of reaction with ONPG}) \times 1000$ and plotted as a function of OD_{450} of the cells. Values in Figure 1 were divided by the A_{450} and normalized to the wild-type case. Although the level of β-galactosidase activity varied as much as threefold from day to day, all strains were affected equally; thus, the relative values remained constant. This unusually large variation could be the result of subtle changes in culture conditions that influence $E\sigma^E$ activity.

S1 mapping

pHtrA was used as the source of probe for the S1 mapping experiments as described (Lipinska et al. 1988). RNA harvesting (Barry et al. 1980) and S1 mapping (Maniatis et al. 1982) were performed as described. A constant amount of RNA was used in each assay. To control for losses during manipulations in the S1 assay, we used a second probe, pRL385, to a portion of the *rpoB* gene (Landick et al. 1990). Samples were electrophoresed on a 5% polyacrylamide–50% urea gel, and the gel was dried. To quantitate transcription from the *degP* promoter, radioactivity in the *degP* and *rpoB* transcripts was counted using an Ambis scanner. Changes in transcription from the *degP* promoter were calculated by [cpm $degP_{(mutant)}$/cpm $rpoB_{(mutant)}$]/[cpm $degP_{(wild\ type)}$/cpm $rpoB_{(wild\ type)}$].

Fractionation of cellular components

Protein composition of the inner membrane and OM was analyzed using their differential solubility in sarcosyl. Ten milliliters of cells growing in LB medium were harvested at an OD_{600} of 1. Cell pellets were resuspended in 500 μl of 100 mM Tris-HCl (pH 8.0), 10 mM EDTA and transferred into an Eppendorf tube. Cell walls were digested with lysozyme (100 μg/ml) on ice for 10 min. $MgCl_2$ (10 mM) and DNase I (50 μg/ml) were added, and the spheroplasts were lysed by three freeze–thaw cycles. The lysate was centrifuged for 10 min at 15,000*g*. The supernatant (cytoplasmic and periplasmic fractions) was removed, and the pellet (crude membrane fraction) was washed with 500 μl of 20 mM $NaPO_4$ (pH 7.0). The inner membrane and OM were resolved by solubilizing specifically the inner membrane with 100 μl of 0.5% sarcosyl in 20 mM $NaPO_4$ for 30 min at room temperature. The insoluble OMs were pelleted by centrifugation for 10 min at 15,000*g*, washed with 100 μl of sarcosyl, centrifuged again, resuspended in SDS sample buffer, and boiled for 5 min. Proteins were analyzed on a 12% polyacrylamide–SDS gel. For Figure 3, the gel also contained 50% urea to resolve OmpC and OmpF. When periplasm and cytoplasm were separated, the cell wall was digested in 20% sucrose to prevent lysis of the spheroplasts (Brissette and Russel 1990).

Acknowledgments

We thank Jori Vetzner for excellent technical assistance. We thank J. Bardwell, J. Beckwith, J. Brandner, M. Inouye, J. Klena, M. Lundrigan, S. Raina, G. Ramakrishnan, C. Schnaitman, T. Silhavy, and K. Strauch for providing plasmids and/or strains. We thank J. Bardwell, P. Danese, and T. Silhavy for discussing results before publication. We thank E. Craig, L. Heisler, A. Kamath-Loeb, and P. Rossmeissl for valuable discussions and critical reading of the manuscript. This research was supported by National Institutes of Health (NIH) grant GM36278-06 to C.A. Gross, and NIH grant GM37509 to T.J. Donohue.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Alfano, C. and R. McMacken. 1989. Heat shock protein-mediated disassembly of nucleoprotein structures is required for the initiation of bacteriophage λ DNA replication. *J. Biol. Chem.* **264:** 10709–10718.

Bardwell, J.C.A., K. McGovern, and J. Beckwell. 1991. Identification of a protein required for disulfide bond formation in vivo. *Cell* **67:** 581–589.

Barry, G., C. Squires, and C.L. Squires. 1980. Attenuation and process of RNA from the rplJL-rpoBC transcription unit of RNA polymerase. *Proc. Natl. Acad. Sci.* **77:** 3331–3335.

Bayer, M.H., G.P. Costello, and M.E. Bayer. 1982. Isolation and partial characterization of membrane vesicles carrying markers of the membrane adhesion sites. *J. Bacteriol.* **149:** 758–767.

Bouffard, G., J. Ostell, and K.E. Rudd. 1992. GeneScape: A relational data base of Escherichia coli genomic map data for MacIntosh computers. *Comput. Applic. Biosci.* **8:** 563–567.

Brandner, J.P., E.V. Stabb, R. Temme, and T.J. Donohue. 1991. Regions of *Rhodobacter sphaeroides* cytochrome c2 required for export, heme attachment and function. *J. Bacteriol.* **173:** 3958–3965.

Brissette, J.L. and M. Russel. 1990. Secretion and membrane integration of a filamentous phage-encoded morphogenetic protein. *J. Mol. Biol.* **211:** 565–580.

Burgess, R.R., A.A. Travers, J.J. Dunn, and E.K.F. Bautz. 1969. Factor stimulating transcription by RNA polymerase. *Nature* **221:** 43–46.

Buttner, M.J., A.M. Smith, and M.J. Bibb. 1988. At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). *Cell* **52:** 599–607.

Castilho, B.A., P. Olfson, and M.J. Casadaban. 1984. Plasmid insertion mutagenesis and *lac* gene fusion with mini-Mu bacteriophage transposons. *J. Bacteriol.* **158:** 488–495.

Catron, K.M. and C.A. Schnaitman. 1987. Export of protein in *Escherichia coli*: A novel mutation in *ompC* affects expression of other major outer membrane proteins. *J. Bacteriol.* **169:** 4327–4334.

Cavard, D., C. Lazdunski, and S.P. Howard. 1989. The acylated precursor form of the colicin A lysis protein is a natural substrate of the *degP* protease. *J. Bacteriol.* **171:** 6316–6322.

Click, E.M., G.A. McDonald, and C.A. Schnaitman. 1988. Translational control of exported proteins that results from OmpC porin overexpression. *J. Bacteriol.* **170:** 2005–2011.

Craig, E.A. and C.A. Gross. 1991. Is *hsp70* the cellular thermometer. *Trends Biochem. Sci.* **16:** 135–140.

Diedrich, D.L. and J.A. Fralick. 1982. Relationship between the OmpC and LamB proteins of *Escherichia coli* and its influence on the protein mass of the outer membrane. *J. Bacteriol.* **149:** 156–160.

Dodson, M., R. McMacken, and H. Echols. 1989. Specialized nucleoprotein structures at the origin of replication of bacteriophage λ. *J. Biol. Chem.* **264:** 10719–10725.

Erickson, J.W. and C.A. Gross. 1989. Identification of the γ^E subunit of *Escherichia coli* RNA polymerase: A second alternate γ factor involved in high-temperature gene expression. *Genes & Dev.* **3:** 1462–1471.

Erickson, J.W., V. Vaughn, W.A. Walter, F.C. Neidhart, and C.A. Gross. 1987. Regulation of the promoters and transcripts of *rpoH*, the *Escherichia coli* heat shock regulatory gene. *Genes & Dev.* **1:** 419–432.

Forst, S., J. Delgado, G. Ramakrishnan, and M. Inouye. 1988. Regulation of *ompC* and *ompF* expression in *Escherichia coli* in the absence of *envZ*. *J. Bacteriol.* **170:** 5080–5085.

Givskov, M., L. Olsen, and S. Molin. 1988. Cloning and expression in *Escherichia coli* of the gene for extracellular Phospholipase A1 from *Serratia liquefaciens*. *J. Bacteriol.* **170:** 5855–5862.

Grodberg, J., M.D. Lundrigan, D.L. Toledo, W.F. Mangel, and J.J. Dunn. 1988. Complete nucleotide sequence and deduced amino acid sequence of the *ompT* gene of *Escherichia coli* K12. *Nucleic Acids Res.* **16:** 1209.

Groisman, E.A. and M.J. Casadaban. 1986. Mini-Mu bacteriophage with plasmid replicons for in vivo cloning and *lac* gene fusion. *J. Bacteriol.* **168:** 357–364.

Gross, C.A., M. Lonetto, and R. Losick. 1992. Bacterial sigma-factors. In *Transcriptional regulation* (ed. S. McKnight), pp. 129–176. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Grossman, A.D., J.W. Erickson, and C.A. Gross. 1984. The *htpR* gene product of *E. coli* is a sigma-factor for heat-shock promoters. *Cell* **38:** 383–390.

Hirota, Y., H. Sukuki, Y. Nishimura, and S. Yasuda. 1977. On the process of cellular division in *Escherichia coli*: A mutant of *E. coli* lacking a murein-lipoprotein. *Proc. Natl. Acad. Sci.* **74:** 1417–1420.

Hoffmann, H.J., S.K. Lyman, C. Lu, M.-A. Petit, and H. Echols. 1992. Activity of the Hsp70 chaperone complex—DnaK, DnaJ, and GrpE—in initiating phage λ DNA replication by sequestering and releasing λ P protein. *Proc. Natl. Acad. Sci.* **89:** 12108–12111.

Johnson, K., I. Charles, G. Dougan, D. Pickard, P. O'Gaota, T. Ali, I. Miller, and C. Hormaecha. 1991. The role of a stress-response protein in *Salmonella typhimurium* virulence. *Mol. Microbiol.* **5:** 401–407.

Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *Escherichia coli* chromosome: Application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50:** 495–508.

Kozutsumi, Y., M. Segal, K. Normington, M.-J. Gething, and J. Sambrook. 1988. The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins. *Nature* **332:** 462–464.

Kumamoto, C.A. 1991. Molecular chaperones and protein translocation across the *Escherichia coli* inner membrane. *Mol. Microbiol.* **5:** 19–22.

Kumamoto, C.A. and J. Beckwith. 1985. Evidence for specificity at an early step in protein export in *Escherichia coli*. *J. Bacteriol.* **163:** 267–274.

Landick, R., J. Stewart, and D.N. Lee. 1990. Amino acid changes in conserved regions of the β-subunit of *Escherichia coli* RNA polymerase alter transcription pausing and termination. *Genes & Dev.* **4:** 1623–1636.

Lipinska, B., S. Sharma, and C. Georgopoulos. 1988. Sequence analysis and regulation of the *htrA* gene of *Escherichia coli*: A σ^{32}-independent mechanism of heat-inducible transcription. *Nucleic Acids Res.* **16:** 10053–10067.

Lipinska, B., O. Fayet, L. Baird, and C. Georgopoulos. 1989. Identification, characterization, and mapping of the *Escherichia coli htrA* gene, whose product is essential for bacterial growth only at elevated temperatures. *J. Bacteriol.* **171:** 1574–1584.

Lipinska, B., M. Zylicz, and C. Georgopoulos. 1990. The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase. *J. Bacteriol.* **172:** 1791–1797.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martin, J., A.L. Horwich, and F.U. Harlt. 1992. Prevention of protein denaturation under heat stress by the chaperonin Hsp60. *Science* **258:** 995–998.

Miller, J.H. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Mizuno, M.-Y., T. Chou, and M. Inouye. 1983. A comparative study on the genes for three porins of the *Escherichia coli* outer membrane. *J. Biol. Chem.* **258:** 6932–6940.

Mizuno, Z. and S. Mizushima. 1990. Signal transduction and gene regulation through the phosphorylation of the two regulatory components: The molecular basis for the osmotic regulation of the porin genes. *Mol. Microbiol.* **4:** 1077–1082.

Mori, K., A. Sant, K. Kohno, K. Normington, M.-J. Gething, and J.F. Sambrook. 1992. A 22 bp *cis*-acting element is necessary and sufficient for the induction of the yeast *kar2* (BiP) gene by unfolded proteins. *EMBO J.* **11:** 2583–2593.

Neidhardt, F.C. and R.A. VanBogelen. 1987. The heat shock response in *Escherichia coli* and *Salmonella typhimurium*. American Society for Microbiology, Washington, D.C.

Pugsley, A.P. 1993. The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* **7:** 50–108.

Ramakrishnan, G., K. Ikenaka, and M. Inouye. 1985. Uncoupling of osmoregulation of the *Escherichia coli* K12 *ompF* gene from *ompB*-dependent transcription. *J. Bacteriol.* **163:** 82–87.

Sen, K. and H. Nikaido. 1990. *In vitro* trimerization of OmpF porin secreted by spheroplasts of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **787:** 743–747.

Simons, R.W., F. Houman, and N. Kleckner. 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene* **53:** 85–96.

Singer, M., T.A. Baker, G. Schnitzler, S.M. Deischel, M. Goel, W. Dove, K.J. Jaacks, A.D. Grossman, J.W. Erickson, and C.A. Gross. 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of Escherichia coli. *Microbiol. Rev.* **53:** 1–24.

Strauch, K.L. and J. Beckwith. 1988. An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. *Proc. Natl. Acad. Sci.* **85:** 1576–1580.

Strauch, K.L., K. Johnson, and J. Beckwith. 1989. Characterization of *degP*, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperature. *J. Bacteriol.* **171:** 2689–2696.

Straus, D.B., W.A. Walter, and C.A. Gross. 1987. The heat-shock response of *E. coli* is regulated by changes in the concentration of σ^{32}. *Nature* **329:** 348–351.

———. 1988. *Escherichia coli* heat shock gene mutants are defective in proteolysis. *Genes & Dev.* **2:** 1851–1858.

———. 1989. The activity of σ^{32} is reduced under conditions of excess heat shock protein production in *Escherichia coli*. *Genes & Dev.* **3:** 2003–2010.

Wang, Q.P. and J.M. Kaguni. 1989. A novel sigma-factor is involved in expression of the *rpoH* gene of *Escherichia coli*. *J. Bacteriol.* **171:** 4248–4253.

Wild, J., E. Altman, T. Yura, and C.A. Gross. 1992a. DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli*. *Genes & Dev.* **6:** 1165–1172.

Wild, J., A. Kamath-Loeb, E. Ziegelhoffer, M. Lonetto, Y. Kawasaki, and C.A. Gross. 1992b. Partial loss of function mutations in DnaK, the Escherichia coli homologue of the 70-kDa heat shock proteins, affect highly conserved amino acids implicated in ATP binding and hydrolysis. *Proc. Natl. Acad. Sci.* **89:** 7139–7143.

Zylicz, M., D. Ang, K. Liberek, and C. Georgopoulos. 1989. Initiation of λ DNA replication with purified host- and bacteriophage-encoded proteins: The role of the *dnaK*, *dnaJ* and *grpE* heat shock proteins. *EMBO J.* **8:** 1601–1608.

competent for retention of McpA. Although the presence of the McpA chemoreceptors at the stalked pole is due to lack of turnover, it may be that under normal circumstances the newly synthesized McpA is targeted to the stalked pole as well as to the flagellated pole, but is degraded at the stalked pole soon after synthesis. We know that in *E. coli* the chemoreceptors can be targeted to both cell poles (*13*). Thus, proteolysis could play a role in the spatial distribution of McpA in C. *crescentus* by removing McpA from the stalked pole of the predivisional cell. Perhaps the presence of protease at the stalked cell pole prevents the deposition of other proteins that are used for the assembly of the flagellum and pili. There is evidence for spatially restricted proteolysis in eukaryotic cells. Localized proteolysis has been shown to be involved in setting up restricted protein distribution in polarized epithelial cells, resulting in the removal of proteins from one membrane domain and not the other upon induction of polarization (*14*).

The specific degradation of McpA could be mediated by a localized activity that modifies the protein, rendering the polypeptide susceptible to degradation by a protease that is present in all cells. Alternatively, the protease could be present or specifically activated only in the stalked cell. A stalked cell-specific protease is likely to be cytoplasmic or possibly associated with the inner membrane, because the COOH-terminus of the chemoreceptor is in the cytoplasm. There is evidence that the cytoplasmic Lon protease is involved in *Myxococcus xanthus* fruiting body formation (*15*). It has been shown that the C. *crescentus* homologue of the Lon protease preferentially segregates to the stalked cell upon division of the predivisional cell (*16*). The fact that Lon segregates to the stalked cell, and not to the swarmer cell, suggests that it might be involved in degradation of any McpA that ends up in the stalked cell portion of the predivisional cell. It is not yet known whether Lon recognizes McpA and whether Lon is specifically targeted to the stalk pole.

REFERENCES AND NOTES

1. H. R. Horvitz and I. Herskowitz, *Cell* **68**, 237 (1992).
2. M. R. K. Alley, S. L. Gomes, W. Alexander, L. Shapiro, *Genetics* **129**, 333 (1991).
3. M. R. K. Alley, J. R. Maddock, L. Shapiro, *Genes Dev.* **6**, 825 (1992).
4. S. L. Gomes and L. Shapiro, *J. Mol. Biol.* **178**, 551 (1984).
5. The half-life of McpA was determined in a mixed population of cells (M. R. K. Alley and L. Shapiro, unpublished data).
6. T. P. Hopp *et al.*, *BioTechnology* **6**, 1204 (1988).
7. Deletions of *mcpA* were obtained by digestion with exonuclease III and S1 nuclease (*3*) and were then ligated to the M2 epitope (*6*) and an in-frame stop codon present in the vectors pJM21, pJM22, and pJM23. These vectors have the M2 epitope in three different reading frames with respect to the polylinker. They also contain an in-frame stop codon with respect to the M2 epitope, and this codon is followed by the site for the restriction enzyme Spe I. These constructs were transferred on Eco RI–Spe I fragments from the M2 epitope vectors into a plasmid capable of replication in *C. crescentus*, pRK290KS1 (*2*). There was no effect attributable to the copy number of the plasmid, because the McpA protein derived from the plasmid-borne *mcpA* gene on pRCH9 (three to five copies per cell) showed a pattern of cell-cycle turnover similar to that of a single chromosomal copy of *mcpA* (Fig. 1).
8. A. Boyd, K. Kendall, M. I. Simon, *Nature* **301**, 623 (1983).
9. A. Krikos, N. Mutoh, A. Boyd, M. I. Simon, *Cell* **33**, 615 (1983).
10. The Tsr sequence originally submitted to GenBank (*9*) is incorrect due to a frameshift near the COOH-terminus. The corrected COOH-terminus of Tsr resembles that of Tar (J. S. Parkinson, personal communication).
11. P. Frederikse and L. Shapiro, unpublished data.
12. J. R. Maddock, M. R. K. Alley, L. Shapiro, unpublished data.
13. J. R. Maddock and L. Shapiro, *Science* **259**, 1717 (1993).
14. R. W. Hammerton *et al.*, *ibid.* **254**, 847 (1991); D. A. Wollner, K. A. Krzeminski, W. J. Nelson, *J. Cell Biol.* **116**, 889 (1992).
15. R. E. Gill, personal communication.
16. S. H. Reuter and L. Shapiro, *J. Mol. Biol.* **194**, 653 (1987).
17. R. C. Johnson and B. Ely, *Genetics* **86**, 25 (1977).
18. M. Evinger and N. Agabian, *J. Bacteriol.* **132**, 294 (1977).
19. McpA is not synthesized in swarmer cells but is synthesized later in the cell cycle (*2*). Total protein synthesis continues during transition from the swarmer to stalked cell. Therefore the ratio of McpA to total protein will decrease prior to McpA synthesis later in the cell cycle. By loading an equal volume of culture in each lane, we avoided observing a decrease in McpA because of the lack of its synthesis during the early part of the cell cycle.
20. U. K. Laemmli, *Nature* **227**, 680 (1970).
21. H. T. Towbin, T. Staehelin, J. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350 (1979).
22. The parental strain used in these experiments was SC1130N (*2*). This strain has a Tn5 insertion in *mcpA* and therefore has no McpA present, and therefore the antiserum to McpA could be used. The strain SC1130N is unable to carry out chemotaxis because the Tn5 insertion is in the first gene in the *mcpA* operon and thus is polar on the downstream genes that are required for chemotaxis (*2*). Although this strain is incapable of methylating McpA, it still degrades McpA, when it is provided in trans.
23. The *mcpA* deletion carried on plasmid pRCM223 (Fig. 3) was introduced into a synchronizable derivative of *C. crescentus* CB15, NA1000, and cell cycle immunoblots were performed on this strain with antiserum to McpA, as described in the legend of Fig. 1. Therefore, a direct comparison of the stability of the wild-type and deleted proteins could be performed in the same immunoblot.

24. Supported by Public Health Service grant GM 32506 (L.S.) and an NIH postdoctoral fellowship GM13929 (J.R.M.). We would like to thank Dale Kaiser and colleagues from the Shapiro laboratory for critical reading of the manuscript.

4 November 1992; accepted 8 February 1993

Microbial Competition: *Escherichia coli* Mutants That Take Over Stationary Phase Cultures

María Mercedes Zambrano, Deborah A. Siegele,* Marta Almirón, Antonio Tormo,† Roberto Kolter‡

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115.

*Present address: Department of Biology, Texas A&M University, College Station, TX 77843.

†Present address: Universidad Complutense de Madrid, Madrid, Spain.

‡To whom correspondence should be addressed.

Many microorganisms, including *Escherichia coli*, can survive extended periods of starvation. The properties of cells that survived prolonged incubation in stationary phase were studied by mixture of 10-day-old (aged) cultures with 1-day-old (young) cultures of the same strain of *Escherichia coli*. Mutants from the aged cultures that could grow eventually took over the population, which resulted in the death of the cells from the young cultures. This phenotype was conferred by mutations in *rpoS*, which encodes a putative stationary phase–specific sigma factor. These rapid population shifts have implications for the studies of microbial evolution and ecology.

Bacteria can remain viable under conditions of poor nutrient availability. Many microorganisms respond to starvation by forming dormant spores, which are generally resistant to extreme environments (*1*). Nonsporulating Gram-negative bacteria, among them *Escherichia coli*, remain metabolically active but also develop increased resistance to a variety of environmental stresses after exponential growth has stopped and cells enter stationary phase (*2*). In Gram-negative bacteria, the overall rate of protein synthesis decreases, but distinct sets of proteins are induced upon entry into stationary phase (*3*, *4*). Some of these proteins protect the cell against environmental challenges such as oxidative damage; others are necessary to maintain viability (*5*, *6*). The molecular mechanism of this response involves the induction of at least one regulon, defined by the genes whose expression depends on the putative stationary phase–specific sigma factor σ^S, the product of the *rpoS* gene also known as *katF* (*4*, *7*). In this report we show that mutations in *rpoS* can have profound effects on the ability of cells to compete and

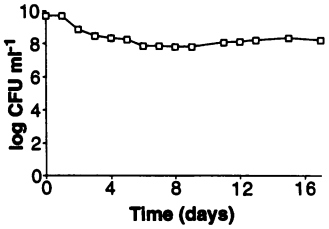


Fig. 1. Viability of *E. coli* ZK126 in LB. A 3-ml culture was kept aerated on a roller at 37°C.

survive under certain conditions in stationary phase cultures.

We studied *E. coli* survival in stationary phase. Our standard wild-type *E. coli* strain (ZK126) saturates at ~1.0×10^9 to 2.0×10^9 colony-forming units (CFU) ml^{-1} when grown in M63 minimal medium with 0.2% glucose. The number of viable counts remains constant for many days and even weeks under these conditions (*8*). The same strain saturates at a density of ~0.5×10^{10} to 1.0×10^{10} CFU ml^{-1} when grown in rich LB medium (Fig. 1). Viability over the first week decreased by one or two orders of magnitude, after which the viable counts stabilized at ~10^8 CFU ml^{-1}.

Cells of the ZK126 strain that were cultured in LB were examined microscopically at different times during 10 days of incubation. To distinguish between live and dead cells, methanol-fixed samples were stained with acridine orange (*8*). Once stained, LB-grown ZK126 cells fluoresce orange if alive and green if nonviable. As expected, exponential phase cells were rod-shaped and fluoresced orange; cells that entered stationary phase became spherical (*9*) and fluoresced orange (Fig. 2A). After 3 days of incubation, most of the cells had lost viability when assayed for CFU (Fig. 1) and fluoresced green (Fig. 2B). By day 10, the fraction of surviving cells had once again become elongated (Fig. 2C). To determine if these cells were undergoing cell division, 10-day-old cultures were treated with the antibiotic aztreonam, which inhibits septation but not cell elongation (*10*). Long filaments were observed in a 10-day-old, but not a 1-day-old, culture treated with aztreonam (Fig. 2D), which indicates that cell division was indeed taking place in these cultures.

To determine how the survivors from a 10-day-old (aged) culture behaved when reintroduced into a 1-day-old (young) culture, we mixed cells from both cultures and determined the numbers of each over several days. Cells from young and aged cultures were distinguished by either nalidixic acid (Nal^R) or streptomycin (Sm^R) resistance markers. To avoid any possible detrimental effect of the 10-day-old medium on the young culture, a small sample (3 μl) from an aged culture of ZK126 Sm^R was mixed with a young culture of ZK126 Nal^R (3 ml). The mixed culture was incubated for 2 weeks, and we determined the viable counts of each population at various times by plating on appropriate media (Fig. 3A). In these mixed cultures, cells from the aged culture grew and took over the population, with a concomitant loss of viable cells from the young culture. When cells from young and aged cultures were mixed in equal numbers (~10^6 CFU ml^{-1} each) in fresh LB liquid medium, both cell populations saturated at ~5×10^9 CFU ml^{-1}. Again, cells from the young culture lost viability whereas cells from the aged culture survived (Fig. 3B). Thus cells from aged *E. coli* cultures had a competitive advantage in stationary phase because they could both grow and cause the death of cells from a young culture. This phenotype was observed regardless of which population carried the Nal^R or Sm^R marker, which indicates that this phenomenon is not due to the presence of a particular antibiotic resistance marker. A strain that lacks any antibiotic resistance marker gave similar results (*12*). This phenotype was expressed only in stationary phase and did not affect exponentially growing cells (Fig. 3B). When cells from two young cultures were mixed, the cells in the minority did not grow and often died out slowly (Fig. 3C).

One possible explanation for the death of cells from the young culture could be that the cells from the aged culture release a stable toxic product or an antibacterial agent that accumulates in stationary phase. However, this hypothesis was disproved because cells from a young culture that were resuspended in filter-sterilized medium from a 10-day-old culture remained viable. Alternatively, death of the young culture cells could result indirectly from competition with cells better able to survive under these particular starvation conditions.

After repeated cycles of exponential growth in liquid or solid medium, cells from aged cultures could still take over those from young cultures (*13*). We obtained several strains that always expressed the phenotype by isolating cells directly from aged cultures or after they had been mixed with and taken over the young population in mixed cultures. This stable inheritance suggests that the phenotype was due to a mutation or mutations and not to a reversible physiological adaptation.

Because *rpoS* participates in regulating stationary phase phenomena, we tested the strains with a growth advantage in stationary phase for mutations in that gene. One of the genes in the σ^S regulon is *katE*, which encodes the enzyme hydrogen peroxidase II (HPII) (6). The allelic state of *rpoS* can be examined indirectly by analysis of HPII activity. This can be done semiqualitatively by addition of H_2O_2 to bacterial colonies grown on LB plates. HPII breaks down H_2O_2, and the concomitant release of O_2 results in bubbling of the colony. Strains bearing the wild-type *rpoS* allele bubble vigorously, whereas strains bearing null alleles (for example, *rpoS::kan*) bubble only

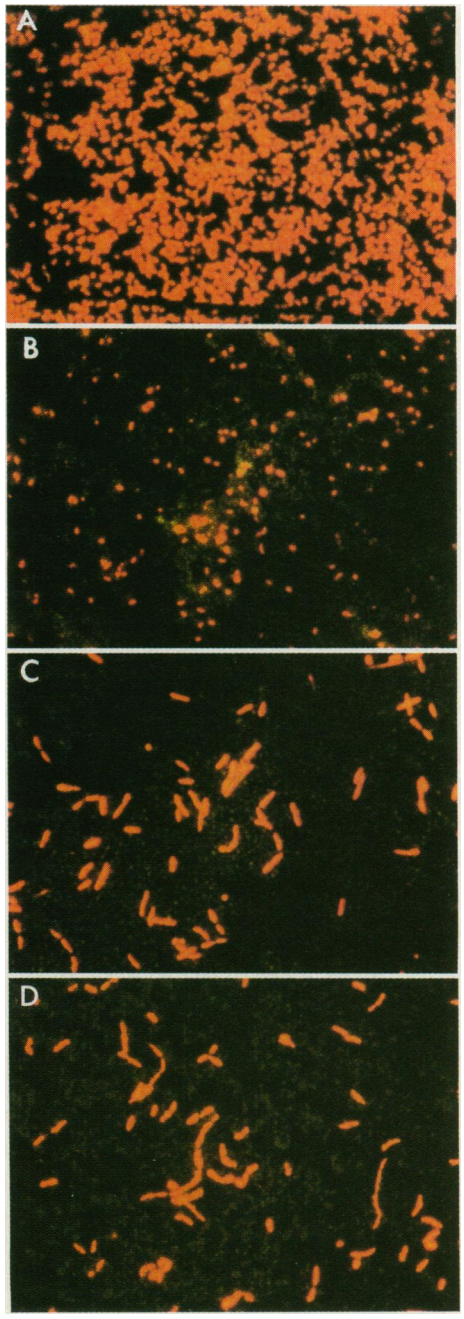


Fig. 2. Acridine orange–stained samples from LB cultures. Samples taken from (**A**) 1-day-old, (**B**) 3-day-old, and (**C**) 10-day-old cultures were spotted onto microscope slides, fixed with methanol, and stained with acridine orange (*8*). Stained cells were viewed under a Zeiss fluorescence microscope with a 487709 filter. (**D**) Cultures (10-day-old) were treated with aztreonam (0.1 μg ml^{-1}) (*10*) for 48 hours and then stained with acridine orange.

slightly (*14*). When the strains that expressed the phenotype of stationary phase growth advantage were treated with H_2O_2 many, but not all, displayed an intermediate bubbling phenotype, which suggests the presence of a mutation in *rpoS*.

Genetic mapping experiments with P1 transduction showed that the linkage between several mutations that result in reduced bubbling and *cysC* was similar to that between *rpoS* and *cysC* [45% cotransducible (*14*)]. The mutations also led to a reduction in the expression during stationary phase of the *rpoS*-dependent *bolA::lacZ* fusion (*14*). These results were all consistent with the hypothesis that many strains that expressed the stationary phase growth advantage phenotype harbored a mutation in *rpoS*. The location of one such *rpoS* mutation (designated with the allele number *rpoS819*) was confirmed by marker rescue with a fragment that contained the 3′ half of the *rpoS* gene and by DNA sequencing.

The wild-type *rpoS* and *rpoS819* alleles were cloned and sequenced (*15*), revealing a 46–base pair duplication at the 3′ end of the *rpoS819* gene (*16*). The mutant and wild-type proteins are identical up to four amino acids from the end. At this point the duplication in the *rpoS819* gene causes a frame shift that replaces the final four amino acids with 39 new residues (Fig. 4). The additional amino acids lie close to the helix that is thought to recognize the "−35" region of promoters (*17*). Polymerase chain reaction amplification of the 3′ region of mutated *rpoS* alleles from different strains revealed that not all of these putative mutations have the same sequence change (*18*).

To determine whether the *rpoS* mutation alone caused the growth advantage phenotype, we transduced the *rpoS819* allele, by way of its linkage to *cysC*, from a strain expressing the phenotype into wild-type cells. We determined the presence of the *rpos819* allele by assaying for bubbling of cells upon treatment with H_2O_2. The resulting strain was grown and, after 1 day in stationary phase, tested by mixture both as a minority in a stationary phase culture (Fig. 3D) or in equal numbers in fresh medium (Fig. 3E) with cells bearing the wild-type *rpoS* allele. The *rpoS819* allele conferred the stationary phase growth advantage phenotype on an otherwise wild-type strain. Similar results were obtained with another *rpoS* allele, *rpoS58* (*19*). In contrast, strains that contained *rpoS* null alleles did not express the stationary phase growth advantage phenotype but died rapidly when mixed with strains that retained *rpoS* function.

Having determined that transduction of certain *rpoS* mutant alleles into wild-type cells confers a growth advantage in stationary phase over unchanged wild-type cells, we tested whether additional mutations could confer a similar growth advantage over cells that contained the mutant *rpoS* alleles. Cells from young and aged cultures of *rpoS819* strains were mixed and cells from aged cultures again grew and caused the death of the young culture (Fig. 3F). This second cycle of aging led to a growth advantage in stationary phase that resulted from a second, unlinked mutation. This mutation requires the presence of the *rpoS819* allele to express its growth advantage phenotype. We have obtained a transposon insertion linked to this second mutant locus, near minute 73 of the *E. coli* chromosome, but the mutant gene has not yet been identified (*20*). We have also obtained a mutation, near minute 27, that renders cells resistant to killing by cells from an aged culture.

During prolonged incubation, mutants with a competitive advantage replace the original population under the strong selective pressure imposed by starvation. These population takeovers, which occurred more rapidly than the population shifts reported for growing *E. coli* cultures (*21*), have implications for the study of the origin of mutations in starved microorganisms. Several reports suggest that, in stationary phase cultures, mutations occur more often when advantageous and so are a direct response to particular environmental challenges (*22*). However, it has often been assumed that stationary phase colonies or cultures are static or nearly static. The observation that stationary phase cultures are dynamic raises the possibility that many of the "post-selection" mutations that have been reported could have arisen from a minority of preexisting mutant cells that were able to grow in the presence of starved cells without changing the overall bacterial counts. Interpretations regarding the appearance of mutations in stationary phase cultures should therefore take this possibility into consideration (*23*).

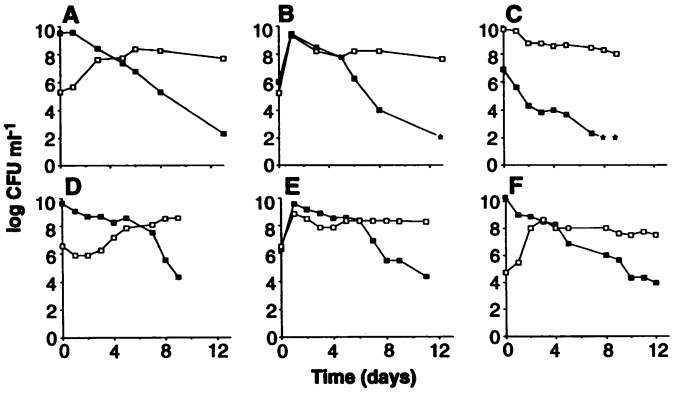


Fig. 3. Mixed culture experiments (*11*) conducted with (**A**) cells from an aged culture (□) in the minority and cells from a young culture (■) in the majority, (**B**) cells from aged (□) and young (■) cultures in equal concentrations, (**C**) cells from two young cultures, (**D** and **E**) *rpoS819* (□) and wild-type *rpoS* cells (■), and (**F**) two *rpoS819* strains in which cells from an aged culture (□) are in the minority and cells from a young culture (■) are in the majority. Asterisks indicate that no colonies were detected at the lowest dilution plated (10 μl directly from the culture).

Fig. 4. Comparison of COOH-terminal amino acid sequences of the wild-type and *rpoS819* gene products. A 46–base-pair duplication in the *rpoS819* gene resulted in the replacement of the last four residues in σ^S with 39 new amino acids. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

σ^S ... IEA | LFRE
σ^{S819} ... IEA | PFARNPANAGAEYRSAVPRVSKHLSERPVSSEAGLFCAQ

REFERENCES AND NOTES

1. R. Losick, P. Youngman, P. J. Piggot, *Annu. Rev. Genet.* **20**, 625 (1986).
2. S. Kjelleberg, M. Hermansson, P. Mårdén, G. W. Jones, *Annu. Rev. Microbiol.* **41**, 25 (1987); A. Matin, E. A. Auger, P. H. Blum, J. E. Schultz, *ibid.* **43**, 293 (1989); A. Matin, *Mol. Microbiol.* **5**, 3 (1991); D. A. Siegele and R. Kolter, *J. Bacteriol.* **174**, 345 (1992).
3. R. G. Groat, J. E. Schultz, E. Zychlinsky, A. Bockman, A. Matin, *J. Bacteriol.* **168**, 486 (1986); M. P. Spector, Z. Aliabadi, T. Gonzalez, J. W. Foster, *ibid.*, p. 420.
4. R. Lange and R. Hengge-Aronis, *Mol. Microbiol.* **5**, 49 (1991).
5. R. Hengge-Aronis, W. Klein, R. Lange, M. Rimmele, W. Boos, *J. Bacteriol.* **173**, 7918 (1991); M. P. McCann, J. P. Kidwell, A. Matin, *ibid.*, p. 4188.
6. P. C. Loewen, J. Switala, B. L. Triggs-Raine, *Arch. Biochem. Biophys.* **243**, 144 (1985).
7. M. R. Mulvey and P. C. Loewen, *Nucleic Acids Res.* **17**, 9979 (1989).
8. D. A. Siegele, M. Almirón, R. Kolter, in *Starvation in Bacteria*, S. Kjelleberg, Ed. (Plenum, New York, in press).
9. R. Lange and R. Hengge-Aronis, *J. Bacteriol.* **173**, 4474 (1991).
10. R. B. Sykes, D. P. Bonner, K. Bush, N. H. Georgopapadakou, *Antimicrob. Agents Chemother.* **21**, 85 (1982).
11. LB cultures (3 ml) were incubated in 18 mm by 150 mm glass test tubes and kept aerated by rotation in a New Brunswick roller at 37°C. Mixes (Fig. 3, A, C, D, and F) were done by transfer of 3 μl of the culture as a minority into the young culture. Cocultures (Fig. 3, B and E) were done by mixture of approximately equal numbers of CFU in fresh LB medium (3 ml). We determined viable cell counts by making serial dilutions in M63 salts

and plating onto both LB-Sm and LB-Nal plates. The media we used have been described [J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1972)].
12. Cells from an aged ZK126 culture were mixed with a young ZK126 Nal^R culture. Dilutions made in M63 were plated on LB and LB Nal media to determine the total number of viable organisms and the number containing the Nal^R marker, respectively. After 8 days, the number of Nal^R CFU per milliliter dropped from 10^9 to 10^6, which indicates that cells from the aged ZK126 culture were taking over the population.
13. Aged cultures and cells that had overtaken young cells in a mixed culture were streaked out on LB plates with the appropriate antibiotic. Isolated colonies were then grown in liquid LB for 1 day. We then tested these cultures by mixing them as a minority with young cultures.
14. D. E. Bohannon *et al.*, *J. Bacteriol.* **173**, 4482 (1991).
15. To clone the wt *rpoS* and *rpoS819*, Kpn I–digested chromosomal DNA was ligated to pUC19 and used to transform a strain bearing the *rpoS::kan* mutation. Because cells develop an *rpoS*-dependent resistance to low pH [P. L. C. Small and S. Falkow, *ASM Abstr.* **B74**, 38 (1992)], ampicillin-resistant transformants (selected at 30°C) were incubated for 30 min in LB (pH 2.5) to select for plasmids that harbor the *rpoS* gene. Survivors were then screened for the *rpoS* gene by restriction enzyme analysis of plasmid DNA. Plasmid DNA was sequenced with Sequenase (U.S. Biochemical) and primers were synthesized on the basis of the published sequence of *rpoS* (*7*).
16. Our sequences of the wt and mutant alleles of *rpoS* have been submitted to GenBank (accession number X16400). The sequence of the wt allele differed slightly from a previously published sequence (*7*); the differences have been noted in the GenBank entry. The most significant change is the absence of a base near the end of the coding region (position 1020), which shortens the predicted protein product by 20 amino acids from its originally reported length. When we sequenced this region from both the *rpoS* gene obtained by Mulvey and Loewen (*7*) and the same gene from our strain, ZK126, we found them to be identical.
17. J. D. Helmann and M. J. Chamberlin, *Annu. Rev. Biochem.* **57**, 839 (1987).
18. DNA was amplified from whole-cell extracts with the use of AmpliTaq polymerase (Perkin-Elmer) and the following primers: 5′-GTTAACGACCATTCTCG-3′ and 5′-TCACCCGTGAACGTGTTC-3′.
19. Many laboratory strains of *E. coli* K-12 show reduced amounts of *rpoS*-regulated genes, which suggests that these strains may have inadvertently undergone selections similar to those of our experiments.
20. The defective transposon mini-Tn*10kan* [J. C. Way, M. A. Davis, D. Morisato, D. E. Roberts, N. Kleckner, *Gene* **32**, 369 (1984)] was used to generate random transposition events in the chromosome of *rpoS819* cells that were isolated from an aged culture. Phage P1 was grown on a pool of the kanamycin-resistant (Km^R) cells, and the lysate was used to transduce *rpoS819* cells to Km^R. Km^R transductants were grown in LB for 1 day and mixed as a minority with a young *rpoS819* culture. Cells that grew from the minority population were isolated and used to determine the linkage between the stationary phase growth advantage phenotype and the Km^R marker (50% cotransducible). We mapped the mini-Tn*10* insertion by cloning the Km^R marker in pUC19 and using this plasmid as a hybridization probe against filters that were blotted with Kohara's ordered phage library of the *E. coli* chromosome [Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987)].
21. D. E. Dykhuizen and D. L. Hartl, *Microbiol. Rev.* **47**, 150 (1983); K. C. Atwood, L. K. Schneider, F. J. Ryan, *Proc. Natl. Acad. Sci. U.S.A.* **37**, 146 (1951); A. Novick and L. Szilard, *ibid.* **36**, 708 (1950); A. F. Bennet, K. M. Dao, R. E. Lenski, *Nature* **346**, 79 (1990).
22. J. Cairns, J. Overbaugh, S. Miller, *Nature* **335**, 142 (1988); B. G. Hall, *Genetics* **120**, 887 (1988); B. G. Hall, *New Biol.* **3**, 729 (1991).
23. J. Mittler and R. Lenski [*Nature* **356**, 446 (1992)] have reported that single mutations in the *bgl* operon allow the slow growth of cells on salicin-containing medium, which can account for the observed high frequency of Sal^+ double mutants.
24. We thank D. Bohannon for suggestions, N. Hochberg for technical assistance, and M. Fox and L. Sonenshein for editorial comments. Supported by grants from the National Science Foundation and the American Cancer Society (R.K.), the Ryan Foundation (M.M.Z.), and Public Health Service (D.A.S.).

29 September 1992; accepted 21 December 1992

An Osmosensing Signal Transduction Pathway in Yeast

Jay L. Brewster, Tamsen de Valoir, Noelle D. Dwyer, Edward Winter, Michael C. Gustin*

Yeast genes were isolated that are required for restoring the osmotic gradient across the cell membrane in response to increased external osmolarity. Two of these genes, *HOG1* and *PBS2*, encode members of the mitogen-activated protein kinase (MAP kinase) and MAP kinase kinase gene families, respectively. MAP kinases are activated by extracellular ligands such as growth factors and function as intermediate kinases in protein phosphorylation cascades. A rapid, *PBS2*-dependent tyrosine phosphorylation of *HOG1* protein occurred in response to increases in extracellular osmolarity. These data define a signal transduction pathway that is activated by changes in the osmolarity of the extracellular environment.

Cell growth requires the uptake of water, driven by an osmotic gradient across the plasma membrane. When the external osmolarity increases, many eukaryotic cells are capable of osmoregulation by increasing their internal osmolarity (*1*). The molecular mechanisms used by eukaryotic cells to sense changes in external osmolarity and transduce that information into an osmoregulatory response are poorly understood. The yeast *Saccharomyces cerevisiae* responds to increases in external osmolarity by increasing glycerol synthesis and decreasing glycerol permeability, thereby accumulating cytoplasmic glycerol up to molar concentrations (*2*, *3*).

We isolated osmoregulation-defective mutants of yeast (*4*) by first screening mutagenized cells for the failure to grow on high-osmolarity medium [YEPD (1% yeast extract, 2% bactopeptone, 2% dextrose) supplemented with 0.9 M NaCl or 1.5 M sorbitol]. Mutants that grew well on YEPD but not on high-osmolarity medium (Osm^S) were then assayed for cellular glycerol accumulation 1 hour after the addition of 0.4 M NaCl to the medium (*3*). Osm^S mutants with a reduction in the glycerol response were all recessive and fell into one of four complementation groups, identifying four *HOG* (high osmolarity glycerol response) genes, *HOG1* to *HOG4*. Of this collection of mutants, we further analyzed two mutants, *hog1-1* and *hog4-1*. The reduced glycerol response and Osm^S of *hog1-1* and *hog4-1* cosegregated 2:2 in tetrads resulting from a backcross to wild type and are thus the result of a single mutation.

Genomic DNA fragments were cloned (*5*) that complemented the Osm^S phenotype of *hog1-1* and *hog4-1*, respectively. To locate *HOG1* and *HOG4* on each genomic clone, we generated subclones and tested for complementation of the Osm^S phenotype of the respective *hog* mutant (*6*). The chromosomal locus of each clone was marked with a selectable marker and shown to be tightly linked to the original *hog* mutation (*7*), demonstrating that *HOG1* and *HOG4* (or closely linked genes) had been cloned.

The nucleotide sequence of the *hog1-1*–complementing DNA (*8*) revealed that *HOG1* (GenBank accession number L06279) is a member of the MAP (mitogen-activated protein) kinase family (*9*). The *HOG1* sequence contains a single, large open reading frame of 1.2 kb encoding a 416–amino acid protein with a molecular size of 47 kD. Northern (RNA) blot hybridization with the cloned *HOG1* gene as probe revealed a 1.4-kb transcript whose abundance was unaffected by exposure of the cells to increased osmolarity. Near the NH_2-terminus of the predicted amino acid sequence of *HOG1*, a stretch of 300 amino acids contains each of the strongly conserved amino acids found in protein kinases (*10*). This sequence is most similar to that of MAP kinase family members (*11*, *12*),

J. L. Brewster, T. de Valoir, N. D. Dwyer, M. C. Gustin, Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77251.
E. Winter, Department of Biochemistry and Molecular Biology and The Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Philadelphia, PA 19107.

*To whom correspondence should be addressed.

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 4023–4027, March 1999
Microbiology, Evolution

Evolution of microbial diversity during prolonged starvation

STEVEN E. FINKEL AND ROBERTO KOLTER*

Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115

Communicated by Stephen Jay Gould, Harvard University, Cambridge, MA, February 8, 1999 (received for review June 15, 1998)

ABSTRACT **Models of evolutionary processes postulate that new alleles appear in populations through random spontaneous mutation. Alleles that confer a competitive advantage in particular environments are selected and populations can be taken over by individuals expressing these advantageous mutations. We have studied the evolutionary process by using *Escherichia coli* cultures incubated for prolonged periods of time in stationary phase. The populations of surviving cells were shown to be highly dynamic, even after many months of incubation. Evolution proceeded along different paths even when the initial conditions were identical. As cultures aged, the takeovers by fitter mutants were incomplete, resulting in the coexistence of multiple mutant forms and increased microbial diversity. Thus, the study of bacterial populations in stationary phase provides a model system for understanding the evolution of diversity in natural populations.**

In studies of microbial evolution, population takeovers have been observed in chemostats (1, 2) and by using serial transfer of batch cultures (3, 4). In both the chemostat and serial transfer systems, cells are continuously incubated in defined medium supplemented with limiting amounts of an essential nutrient to restrict growth rates and population sizes. These model systems create culture environments that are essentially constant, leading to selection of specific phenotypes (reviewed in ref. 5). Occasionally, a mutation arises that confers a selective advantage on a cell. This cell can outcompete its siblings and replace the population with its own progeny, in a process sometimes referred to as "periodic" selection (1, 5, 6). It is generally assumed that the result of these periodic selections is a complete takeover of the culture by cells of a single genotype.

We have chosen to study evolution of the bacterium *Escherichia coli* during extended stationary phase incubation by using a constant batch culture system in which no nutrients are added after the initial inoculation and no cells are removed from the system either by dilution (as in a chemostat) or by transfer to fresh medium (as in serial passage systems). This mode of incubation ensures that there is no loss of genetic diversity due to the arbitrary discarding of cells, removing the "bottlenecks" that can be incurred in other systems. By using this type of long-term batch culture system we had previously shown that cultures of a standard laboratory wild-type strain (ZK126) grown in rich medium saturate at $\approx 5 \times 10^9$ colony forming units (cfu) ml^{-1}. Viability over the next few days decreases by about two orders of magnitude, stabilizing at $\approx 5 \times 10^7$ cfu ml^{-1}. However, these surviving cells are different from their parents (7–10). Cells aged for 10 days in Luria–Bertani (LB) medium will take over a fresh overnight culture of the parental strain when the aged cells are introduced as a minority, in a phenomenon referred to as the growth advantage in stationary phase (GASP) phenotype (7–10). It was shown that the GASP phenotype is conferred by the acquisition of increased fitness mutations, as was the case in the chemostat and serial transfer experiments (1–5). Here again, it was assumed that a single mutant, which arose during periods of rapid growth, completely took over the population. Although all of the previous studies showed population takeovers by fitter mutants, none of them directly addressed two important questions in microbial evolution: (*i*) Given identical initial conditions does evolution always proceed along the same path? and (*ii*) Does the appearance of advantageous mutations always result in a single genotype dominating the population? By analyzing cultures incubated in stationary phase for much longer periods of time we have directly addressed these questions.

MATERIALS AND METHODS

Bacterial Strains, Culture Media, and Cell Growth Conditions. Strains ZK1142 (Nal^R) and ZK1143 (Str^R) are *E. coli* K-12 F^- (nonmating) derivatives and have been described elsewhere (7). The nalidixic acid-resistance (Nal^R) and streptomycin-resistance (Str^R) markers (encoded by *gyrA* and *rpsL,* respectively) are effectively neutral in the absence of drug selection, having no effect on competitive advantage under these growth conditions (7). In addition, we routinely assay for the spontaneous acquisition of both drug markers in either strain and have not observed a Nal^R strain become Str^R, or vice versa. All bacterial cultures were incubated in 5.0 ml of LB broth (11) in 18 × 150 mm glass test tubes with constant aeration in a New Brunswick roller at 37°C.

Long-Term Growth of Bacterial Cultures. Cultures of ZK1142 and ZK1143 have continued incubating for more than 1 year (data not shown) and titers remain at $\approx 10^6$ cfu ml^{-1}. Sterile distilled water is added to each culture monthly to adjust for the loss of volume because of evaporation.

Mixing of Cells from Cultures Aged 1–30 Days. LB cultures (5 ml) of strains ZK1142 (Nal^R) and ZK1143 (Str^R) were incubated at 37°C with aeration. Samples of each culture were taken after 1, 10, 20, and 30 days of constant incubation and stored as frozen LB/20% glycerol stocks at −80°C. For GASP competition experiments, fresh overnight LB cultures of each strain (from 1-, 10-, 20-, and 30-day-old cultures) were inoculated from frozen stocks. Mixes were done by transfer of 5 μl of cells from an overnight culture of the aged cells (as the minority) into a 5.0 ml overnight culture inoculated with cells from the younger culture. Viable counts were determined by serial dilution of cells removed periodically from the culture followed by plating on LB agar containing either nalidixic acid or streptomycin at 20 or 25 μg/ml, respectively (11).

Mixing Experiments Showing Population Dynamism. Ten pairs of 5.0 ml cultures of ZK1142 (Nal^R) and ZK1143 (Str^R) were incubated for 30 days at 37°C with aeration and then 2.5 ml each mixed to generate 10 new pairs of cultures. The time of mixing became Week 0, and mixed cultures were allowed to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: LB, Luria–Bertani; GASP, growth advantage in stationary phase; cfu, colony forming units.

*To whom reprint requests should be addressed: e-mail: kolter@mbcrr.harvard.edu.

continue incubating. Viable counts of each subpopulation were determined biweekly as described above.

Mixing Experiments Showing Evolution Along Different Paths. One culture of ZK1142 (Nal^R) and two cultures of ZK1143 (Str^R-1 and Str^R-2) were incubated for 30 days at 37°C with aeration. After 30 days, the Nal^R culture was mixed 1:1 with each of the cultures from the pair carrying the Str^R drug marker. This resulted in two new cultures where each strain was mixed with two different, initially isogenic, aged cultures. Incubation of the mixed cultures was then continued and titers of Nal^R and Str^R subpopulations were determined weekly as described above. This experiment was done twice in pairs, resulting in eight mixed cultures being tested.

Observation of Cells with Altered Colony Morphotypes. Cells from long-term incubated batch cultures were plated periodically onto LB agar and incubated for 24–48 hr at 37°C. The number of colonies showing each morphotype was determined and individual colonies were picked and restreaked onto LB agar plates.

Preparation of Chromosomal DNA and Pulsed-Field Gel Electrophoresis. Chromosomal DNA in agarose microbeads was prepared as described (12). Restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's instructions. Pulsed-field gels were run using 1% agarose gels in a Bio-Rad CHEF apparatus following the manufacturer's instructions.

RESULTS

GASP Mutants Arise Constantly During Long-Term Incubation. New GASP mutants are continuously selected as cultures incubate for extended periods of time. For example, not only do cells from 10-day-old cultures outcompete cells in fresh overnight cultures (Fig. 1*A*), but cells isolated from 20-day-old cultures can outcompete those from 10-day-old cultures (Fig. 1*B*), and mutants from 30-day-old cultures express a competitive advantage when mixed with cells from 20-day-old cultures (Fig. 1*C*). The observation that cells from progressively aged cultures have a competitive advantage over previous generations suggests that cells incubated under these conditions are under constant selection for the appearance of fitter mutants and implies that these long-term stationary phase cultures are highly dynamic, despite the fact that viable cell numbers remain roughly constant. Therefore, we designed experiments to directly observe the dynamism of these cultures using a series of mixing experiments with initially isogenic strains carrying either nalidixic acid (Nal^R) or streptomycin (Str^R) resistance markers.

Prolonged Dynamism of Stationary Phase Cultures. The results in Fig. 2 show that long-term stationary phase cultures are highly dynamic. Ten pairs of initially isogenic "founder" cultures (marked with either Nal^R or Str^R) were aged for 30 days to allow for several cycles of GASP takeovers. These aged Nal^R and Str^R cultures were then mixed in equal numbers, in duplicate (Fig. 2*A*). The titers of Nal^R and Str^R cells were determined biweekly. In each of the 10 pairs of mixed cultures, the proportion of Nal^R and Str^R cells varied over time, where the total viable counts remained roughly constant ($\approx 10^6$–10^7 cfu ml^{-1}), directly demonstrating that these cultures are dynamic. An example of this population dynamism is shown in Fig. 2*B*. In this case, the number of Str^R cells increases after 2 weeks, with a concomitant decrease in the number of Nal^R cells. This suggests that on mixing, GASP mutants in the Str^R population were better competitors than the GASP mutants in the Nal^R culture. However, from weeks 4 to 6, the Nal^R subpopulation titer increased 100-fold with a concomitant decrease in the Str^R cell numbers, indicating that the next GASP mutants able to take over the culture came from the Nal^R subpopulation. (In some cultures subpopulation titers changed more than four orders of magnitude in two weeks.)

The patterns of GASP takeovers in all 10 pairs of cultures during the first 2 weeks of incubation occurred in parallel. As shown in Fig. 2*B*, the points of inflection in the subpopulation titers occurred at the same time. This initial parallelism was expected because the long-term cultures were prepared in duplicate due to the transfer of one-half of each aged Nal^R or Str^R culture into two newly mixed cultures (see Fig. 2*A*). The initial subpopulations in each pair of cultures should have been genetically identical. That is, the current best competitors comprising the majority of cells in the Nal^R and Str^R 30-day-old founder cultures are present in both mixed cultures. Therefore, we expected that the patterns of cell growth and death, as reflected by changes in the relative titers of each marked subpopulation, would be similar between each pair of cultures shortly after mixing. Most pairs of cultures diverged between 2 and 6 weeks after initial mixing (Table 1). Fig. 2*C* shows a mixed pair in which parallel growth patterns are lost after 2 weeks of incubation.

Quite surprisingly, however, the patterns of fluctuation of Nal^R and Str^R subpopulations remained parallel in some cultures for up to 10 weeks (see Fig. 2*D*.) In one case, the patterns of takeover remained parallel for almost 4 months (Table 1). The simplest interpretation of long-term parallelism is that the mutant cells that took over at later times were present on initial mixing (week = 0) and were able to persist in the culture until conditions arose that allowed the expression of their GASP phenotype. This implies that the culture environments in each pair of cultures remained similar. A very dramatic illustration of this ability to persist is shown in Fig. 2*D*. Str^R cells remained at titers of $\approx 5 \times 10^5$ cfu ml^{-1} until

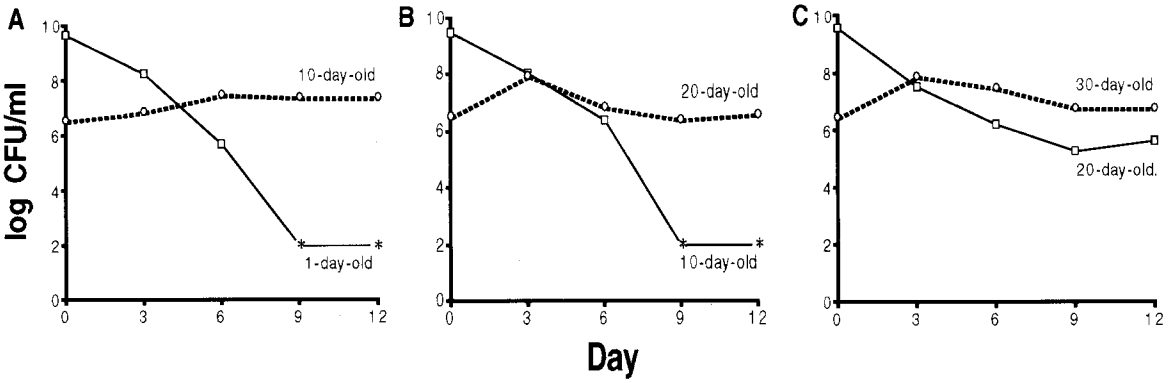


FIG. 1. Consecutive generations of GASP mutants arise in the same culture. Progressively aged cultures were mixed. (*A*) One-day-old in the majority (solid line) vs. 10-day-old in the minority (broken line). (*B*) Ten-day-old in the majority (solid line) vs. 20-day-old in the minority (broken line). (*C*) Twenty-day-old in the majority (solid line) vs. 30-day-old in the minority (broken line). Asterisks indicate that cfu ml^{-1} were below the limit of detection ($<10^2$ cfu ml^{-1}).

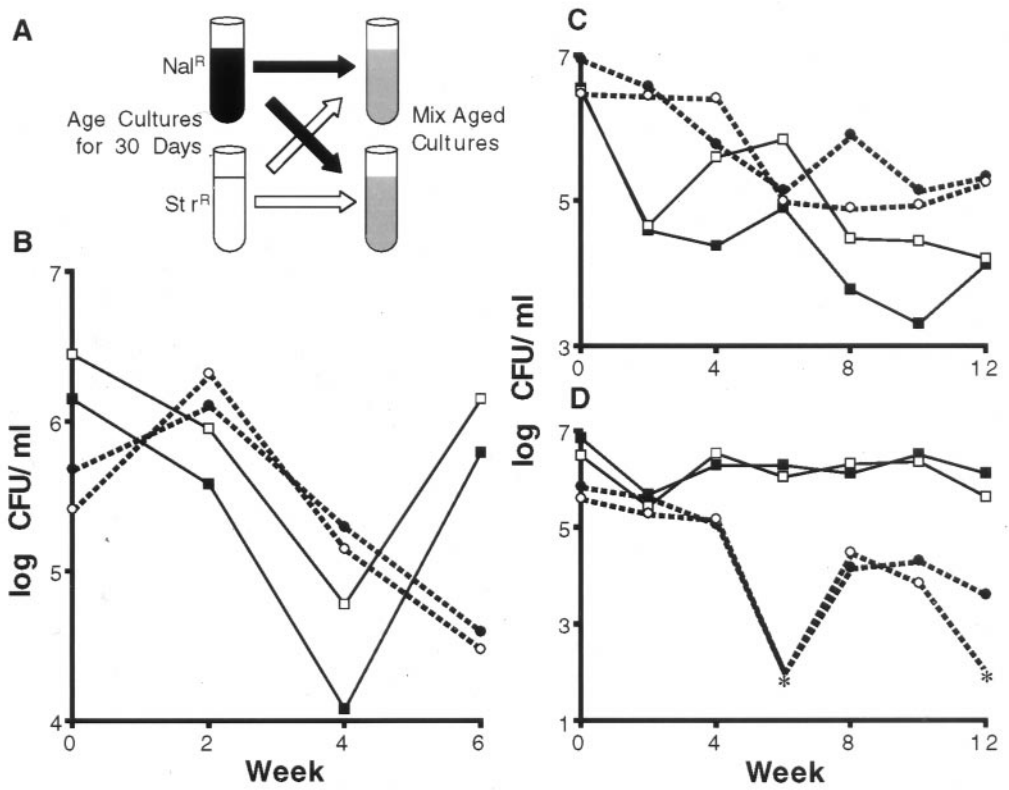


FIG. 2. Dynamism in stationary phase cultures. (*A*) Five milliliter cultures of either nalidixic acid (Nal^R)- or streptomycin (Str^R)-resistant bacteria were incubated for 30 days and then mixed 1:1 to generate two new cultures. (*B–D*) Three representative pairs of mixed Nal^R and Str^R cells. Each culture of the pair is represented by either the solid or open symbols. Nal^R cells are represented by solid lines with squares, Str^R cells are represented by broken lines with circles. Asterisks indicate that cfu ml^{-1} were below the limit of detection ($<10^2$ cfu ml^{-1}).

week 4. Two weeks later, their levels were below the level of detection ($<10^2$ cfu ml^{-1}), only to reappear after another 2 weeks at $\approx 5 \times 10^4$ cfu ml^{-1}. This pattern of disappearance and reappearance was identical in both cultures of the pair. We also confirmed that the reappearing strain had not become resistant to the second drug.

An alternative explanation for the observed long-term parallelism is that the particular mutant alleles conferring the GASP phenotype appeared after mixing and were selected in response to a particular environment. This explanation would require that any changes in the culture environments must happen identically in both cultures, and mutants with identical phenotypes would have to be selected simultaneously. As shown in Fig. 2*D*, the titers of the Str^R cells decreased to less than 10^2 cfu ml^{-1} and then, over 2 weeks time, increase at least 500-fold. It seems highly unlikely that two mutations could arise independently and simultaneously in separate subpopulations of such small size (fewer than 500 cells in the entire culture).

Table 1. Divergence of long-term cultures after 24 weeks of incubation

Culture pair (A and B)*	Time of divergence† (weeks)	Drug marker(s) remaining‡	
		Culture A	Culture B
1	2	Str	Nal and Str
2	2	Str	Nal and Str
3	2	Str	Str
4	4	Nal	Str
5	4	Nal	Nal
6	6	Nal	Nal and Str
7	6	Str	Str
8	8	Str	Str
9	10	Nal and Str	Nal
10	16	Nal	Nal

*See Fig. 2*A* for description of the culture pair.
†Divergence was observed after the week indicated.
‡Limit of direction is $<10^2$ cfu ml^{-1}.

Although the patterns of subpopulation fluctuation were parallel in all 10 pairs of cultures for some period of time, eventually these patterns diverged (Table 1; Fig. 2 *C* and *D*.) This suggests that over time new mutations arise allowing individual cells in these cultures to evolve along different paths. This divergence is expected because, as the mixed populations age, the chance of the same mutation arising in both cultures of each pair is very low. Cells within each pair of mixed cultures will acquire different GASP-conferring mutations and the relative numbers of Nal^R and Str^R cells within each pair of cultures will eventually diverge. Divergence was observed in all 10 pairs of cultures, although it occurred at different times (Table 1). After 24 weeks of continuous incubation, seven of the cultures contained only Nal^R cells, nine had only Str^R cells, and four cultures still contained both Nal^R and Str^R cells (Table 1).

Initially Isogenic Populations Evolve Along Different Paths. The fact that each of the 10 pairs behaved differently suggests that a high degree of genetic diversity was present in these aged cultures at the time of mixing, which allowed cells to evolve along different paths in waves of takeovers by different GASP mutants. We designed another mixing experiment to directly determine whether initially isogenic bacteria can evolve along different paths when placed in initially identical environments. In these experiments, two Str^R cultures (Str^R-1 and Str^R-2; derived from the same initial culture) and a single Nal^R culture were aged for 30 days (Fig. 3*A*). At that time (which becomes week 0 in the experiment) cells from the aged Nal^R culture were mixed separately with cells from either the Str^R-1 or Str^R-2 culture, incubation was continued, and their relative

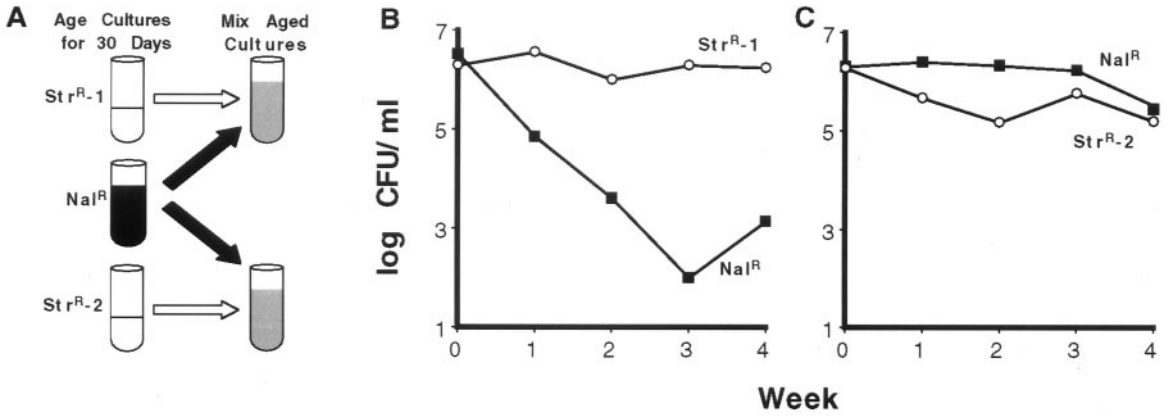


FIG. 3. Initially isogenic strains evolve along different paths. (*A*) One Nal^R and two Str^R cultures were incubated for 30 days. The Nal^R culture was mixed 1:1 with cells from either the Str^R-1 or Str^R-2 cultures, creating two new mixed cultures. The two Str^R cultures (Str^R-1 and Str^R-2) were derived from the same colony and were initially isogenic. (*B* and *C*) Representative pair of mixed cultures. Nal^R cells (■), Str^R cells (○).

titers were determined weekly. As shown in Fig. 3 *B* and *C*, cells from the Nal^R culture are quickly outcompeted by cells from culture Str^R-1, but are good competitors against cells from culture Str^R-2. This indicates that even though the same culture was used to initially inoculate the two Str^R cultures, during the 30 days of separate incubation different mutations were selected that gave cells in each new culture a distinct competitive advantage.

Different Subpopulations Coexist During Long-Term Incubation. Although we have observed that after the first 10 days of aging of bacterial cultures GASP takeovers seem complete, later takeovers are not complete, as shown in Fig. 1*C*. This suggests that on continued incubation different mutants may coexist. This diversity can be directly observed on plating of bacterial cells from an aging culture onto nutrient agar plates. As shown in Fig. 4*A*, cells plated from a 150-day-old culture fell into three distinct colony morphology classes (morphotypes): colonies of normal "cream" color (C), white colonies (W), and small or "mini" colonies (M). The colony morphologies of these cells are likely conferred by genetic factors, because they are stably inherited; on restreaking of a particular colony, the distinctive morphologies breed true. In addition, restriction fragment length polymorphism analysis demonstrates that these distinctive colony morphotypes are derived from the same parental strain (Fig. 4*B*). However, whereas the *Sfi*I digest patterns of these strains are identical for nearly every band, a few differences were observed. For example, the *Sfi*I digest showed that the C strain chromosome has an ≈50 kb deletion (located at ≈93 min), which causes the loss of two restriction fragments and the appearance of a new band (note the doublet in the top band of lane C in Fig. 4*B*); differences that distinguish the W and M morphotypes were also observed. Digests with *Not*I and *Xba*I revealed further genetic differences between the three colony morphotypes (data not shown). The appearance and disappearance of distinct colony morphotypes over time within a single culture directly demonstrates both that these aged cultures are dynamic and that over time GASP mutants coexist. On day 150 in this culture the composition was ≈75% C, ≈20% W, and ≈5% M (Fig. 4*C*). However, 2 weeks later the C morphotype appeared to almost completely take over the culture, only to decrease to about 50% of the population at day 180, and to less than 10% by day 240. Also, over this period two new morphotypes, "fried egg" (E) and very small or "nano" (N), appeared. Fluctuations in the proportion and class of colony morphotypes continued to be observed over many months in this particular culture (Fig. 4*C*), as well as in other initially isogenic cultures (data not shown). These experiments conclusively demonstrate that populations in these long-term cultures are continually dynamic and that takeovers in these environments are rarely complete, giving rise to the evolution of diversity along many different paths.

DISCUSSION

The consecutive takeovers we observe can be thought of as similar to the periodic selections described in chemostats and for long-term serially passaged cultures (5, 13), except that the selections described here arose during long-term stationary

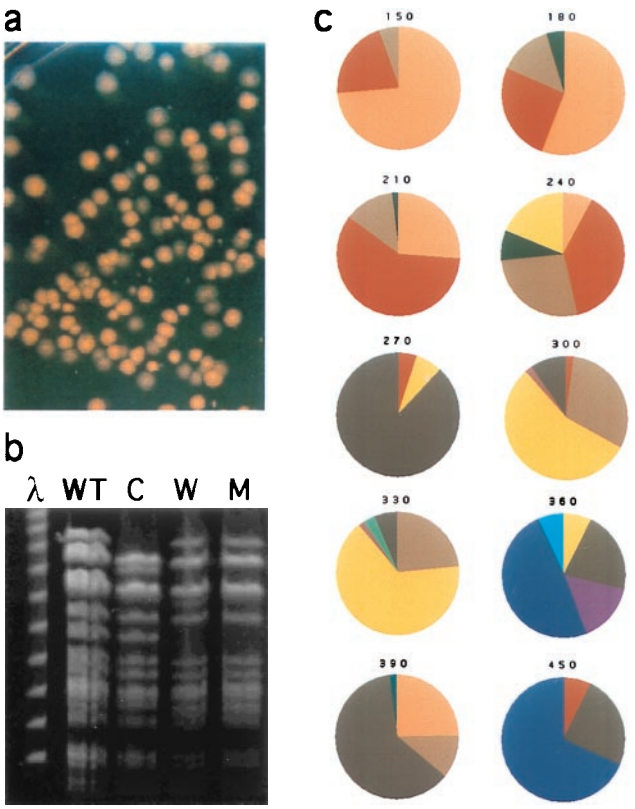


FIG. 4. Coexistence of different colony morphotypes during long-term incubation in a single culture. (*A*) Colonies plated from a 150-day-old culture: normal or cream-colored colony morphotype (C), white morphotype (W), and minicolony morphotype (M). (*B*) Pulsed-field gel electrophoresis of *Sfi*I digested chromosome of wild-type (WT) cells and the three new morphotypes (C, W, and M); a phage lambda (λ) size standard is at left. (*C*) Pie charts reflecting population composition (%) during different times of incubation (days) of each morphotype, each color represents a different colony morphotype (see text for descriptions).

phase incubation. The rapidity of takeover and the diversity generated in our system is striking, especially given the small population sizes. Long-term cultures that have been aged for as little as an additional week invariably contain cells that have a competitive advantage over cells from younger cultures. In addition, it is clear that there are many different mutations that can confer a GASP phenotype on these aging cultures. What are the possible mechanisms by which this diversity of mutant alleles is generated?

The starvation conditions encountered during stationary phase incubation may permit a transient increase in the mutation rate due to a variety of factors, including decreased fidelity during replication and reductions in repair activity (14–18). Although an overall increase in mutation frequency might be thought of as being deleterious during times of nutrient deprivation, several models of modulation of mutation frequency during stationary phase (18), as well as experimental evidence (14–17), suggest that in the face of extremely limiting nutrient resources and intense competition, an increase in mutation frequency is a potential mechanism to quickly generate new alleles. In addition, evidence that stationary phase cells may have more than one chromosome equivalent suggests that a transient increase in mutation frequency might not be so problematic (19). With several chromosome equivalents per cell, the effect of deleterious mutations will be masked by other copies of the mutated gene, whereas advantageous mutations will allow proliferation of cells that retain the beneficial allele. Two likely and not mutually exclusive mechanisms responsible for generating cells with more than one chromosome equivalent are (*i*) an uncoupling of the replication machinery from the cell division apparatus, such that all rounds of replication are completed once initiated, whether or not the cell has the ability to continue dividing (19) and (*ii*) *oriC*-independent replication referred to as "inducible stable DNA replication," where DNA replication is initiated in the absence of cell division (20).

Another source of new alleles may be the dead cells within the culture. For every cell that survives after 30 days of incubation, approximately 1,000 cells have died. Although there is little evidence for natural competence or horizontal gene transfer independent of conjugative mechanisms in *E. coli*, it is tempting to speculate that cells that find themselves surrounded by the detritus, including chromosomal DNA, of many dead sibling cells, may have an ability to "take up" and incorporate exogenously derived genetic information.

The work presented here makes it clear that as batch cultures age under starvation conditions a wide variety of beneficial mutations arise leading to the evolution of microbial diversity. In their natural setting, bacteria rapidly consume available nutrients and then most likely spend much of their existence in a starved state (21). When nutrients are abundant, selective pressures to increase genetic variation are low. However, upon onset of starvation, there is intense selective pressure for any mutation that confers a competitive advantage. An attractive model of bacterial evolution is one in which the bulk of new mutations is generated during periods of starvation. We would like to propose that our studies of bacterial populations evolving in the laboratory may reflect evolutionary processes acting on natural bacterial populations. When microbes occupy a new nutritional niche they can quickly consume available nutrients and, in all likelihood, enter a state of starvation akin to stationary phase. In these nutritionally poor environments, those bacteria able to scavenge nutrients trapped in the dead biomass will have a selective growth advantage. Over time, as the selective advantage conferred by new advantageous mutant alleles is less pronounced, different GASP mutants can coexist, eventually leading to the evolution of microbial diversity.

We thank members of the Kolter and Beckwith labs; Stephen Jay Gould, Michael Farrell, and Linc Sonenshein for insightful comments; and William Rosche and Patricia Foster for helpful discussions and assistance with pulsed-field gel electrophoresis. This work was supported by the National Science Foundation (MCB-9728736) and the National Institutes of Health (GM55199). S.E.F. is a fellow of the Helen Hay Whitney Foundation.

1. Novick, A. & Szilard, L. (1950) *Proc. Natl. Acad. Sci. USA* **36,** 708–719.
2. Helling, R. B., Vargas, C. N. & Adams, J. (1987) *Genetics* **116,** 349–358.
3. Lenski, R. E. & Travisano, M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6808–6814.
4. Elena, S. E. & Lenski, R. E. (1997) *Evolution* **51,** 1058–1067.
5. Dykhuizen, D. E. (1990) *Annu. Rev. Ecol. Syst.* **21,** 373–398.
6. Atwood, K. C., Schneider, L. K. & Ryan, F. J. (1951) *Proc. Natl. Acad. Sci. USA* **37,** 146–155.
7. Zambrano, M. M., Siegele, D. A., Almirón, M., Tormo, A. & Kolter, R. (1993) *Science* **259,** 1757–1759.
8. Zambrano, M. M. & Kolter, R. (1993) *J. Bacteriol.* **175,** 5642–5647.
9. Zambrano, M. M. & Kolter, R. (1996) *Cell* **86,** 181–184.
10. Finkel, S. E., Zinser, E., Gupta, S. & Kolter, R. (1997) in *Molecular Microbiology: NATO-ASI Series*, eds. Busby, S. J. W., Thomas, C. M. & Brown, N. L. (Springer, Berlin), pp. 3–16.
11. Miller, J. H. (1992) *A Short Course in Bacterial Genetics* (Cold Spring Harbor Lab. Press, Plainview, NY).
12. Koob, M. & Szybalski, W. (1992) *Methods Enzymol.* **216,** 13–20.
13. Elena, S. F., Cooper, V. S. & Lenski, R. E. (1996) *Science* **272,** 1802–1804.
14. Foster, P. L. (1997) *J. Bacteriol.* **179,** 1550–1554.
15. Torkelson, J., Harris, R. S., Lombardo, M. J., Thulin, C. & Rosenberg, S. M. (1997) *EMBO J.* **16,** 3303–3311.
16. Bridges, B. A. (1997) *Nature (London)* **387,** 557–558.
17. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. (1997) *Nature (London)* **387,** 703–705.
18. Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H. & Godelle, B. (1997) *Nature (London)* **387,** 700–702.
19. Åkerlund, T., Nordström, K. & Bernander, R. (1995) *J. Bacteriol.* **177,** 6791–6797.
20. Kogoma, T. (1997) *Microbiol. Mol. Rev.* **61,** 212–238.
21. Morita, R. M. (1993) in *Starvation in Bacteria*, ed. Kjellberg, S. (Plenum, New York), pp. 1–23.

JOURNAL OF BACTERIOLOGY, Sept. 1999, p. 5800–5807 Vol. 181, No. 18
0021-9193/99/$04.00+0

Mutations Enhancing Amino Acid Catabolism Confer a Growth Advantage in Stationary Phase

ERIK R. ZINSER AND ROBERTO KOLTER*

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115

Received 9 April 1999/Accepted 7 July 1999

Starved cultures of *Escherichia coli* undergo successive rounds of population takeovers by mutants of increasing fitness. These mutants express the growth advantage in stationary phase (GASP) phenotype. Previous work identified the *rpoS819* allele as a GASP mutation allowing cells to take over stationary-phase cultures after growth in rich media (M. M. Zambrano, D. A. Siegele, M. A. Almirón, A. Tormo, and R. Kolter, Science 259:1757–1760, 1993). Here we have identified three new GASP loci from an aged *rpoS819* strain: *sgaA*, *sgaB*, and *sgaC*. Each locus is capable of conferring GASP on the *rpoS819* parent, and they can provide successively higher fitnesses for the bacteria in the starved cultures. All four GASP mutations isolated thus far allow for faster growth on both individual and mixtures of amino acids. Each mutation confers a growth advantage on a different subset of amino acids, and these mutations act in concert to increase the overall catabolic capacity of the cell. We present a model whereby this enhanced ability to catabolize amino acids is responsible for the fitness gain during carbon starvation, as it may allow GASP mutants to outcompete the parental cells when growing on the amino acids released by dying cells.

In the natural environment, chemoorganotrophs such as *Escherichia coli* obtain both their carbon and energy from organic matter released by other cells. The mechanisms of organic nutrient release are variable, ranging from regulated extrusion of metabolic end products to release as a result of death and lysis of donor cells (3, 21–23, 34). However, the actual bioavailability of carbon in nature is low due to intense competition (22, 23). As a result, natural microbial populations spend the majority of their lives under starvation stress, interspersed with sporadic and short-lived periods of growth as nutrients become available.

Our laboratory uses carbon-starved cultures of *E. coli* as an experimental model to understand the processes of survival and evolution in natural microbial populations. *E. coli* can survive extended periods of starvation. In aerated rich medium (Luria-Bertani [LB] broth), *E. coli* ceases growth due to carbon limitation (37). During the first several days of starvation, the population loses 90 to 99% of the viable counts (40). However, the viable counts nearly level off after these first few days, and populations can survive for several years in this spent LB medium aerated at 37°C without further addition of carbon (7, 8). As the cultures consume exogenous carbon during exponential growth, the biomass is the most likely source of carbon during extended survival, which becomes available when the cells die.

While the overall population of stationary-phase *E. coli* cultures may be considered starved in that there is no net increase in biomass, there are subpopulations that are clearly not starved, as they are able to grow as a subculture and take over the population (8, 38–40). These subpopulations consist of mutants with enhanced fitness during starvation. The ability to grow during starvation has been termed the growth advantage in stationary phase (GASP) phenotype (38). Studies on cultures starved for extended periods demonstrate that the GASP phenomenon is continuous: multiple rounds of population takeovers occur throughout the starvation period (7, 40). Interestingly, as the cultures age, they increase in diversity, as several genetically distinct subpopulations coexist (7).

The first mutation conferring the GASP phenotype after growth in rich media was identified as an allele of *rpoS* (40), a gene whose product, σ^S, is responsible for the regulation of many genes during starvation stress (12). Transduction of the GASP allele of *rpoS* (*rpoS819*) into an otherwise wild-type strain was sufficient to confer the GASP phenotype (40). The *rpoS819* allele is a 46-bp duplication at the 3′ end of the gene, which results in a replacement of the last four residues in σ^S with 39 new amino acids. Expression of two σ^S-dependent genes, *katE* (25) and *bolA* (4, 15), are both reduced in the *rpoS819* strain (40), indicating a reduction of function in this allele. The physiological basis for the fitness gain of the *rpoS819* mutation is not yet known.

The purpose of our investigation was to understand how GASP mutations alter cell physiology to provide fitness gains in stationary phase. To this end, we sought to identify and characterize new GASP mutations. ZK1141, an isolate from an aged culture of the *rpoS819* strain, was capable of outcompeting its *rpoS819* parent, indicating that additional GASP mutations accumulated in this strain (38). In this study, we have demonstrated that the ZK1141 strain has acquired three new GASP mutations, each of which can confer the GASP phenotype on the *rpoS819* parent. Each of these newly identified GASP alleles, as well as the *rpoS819* allele, increased starvation survival fitness in an additive manner. Each of these four GASP alleles also conferred growth advantages on amino acids as the sole sources of carbon and energy. Similar to the competitive fitnesses, these growth phenotypes were additive.

MATERIALS AND METHODS

Bacterial strains. The *E. coli* strains used in this study are listed in Table 1.

Media and growth conditions. All experiments were performed at 37°C, except where noted. The media used in this study have been previously described (19). M63 minimal medium was supplemented with 1 μg of thiamine per ml and 1 mM $MgSO_4$. All amino acids used in this study were of the L configuration. Where appropriate, LB plates were supplemented with streptomycin (25 μg/ml), nali-

* Corresponding author. Mailing address: Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115. Phone: (617) 432-1776. Fax: (617) 738-7664. E-mail: kolter@mbcrr.harvard.edu.

TABLE 1. *E. coli* strains used in this study

Strain	Genotype or phenotype	Reference or source
GASP strain		
ZK126	W3110 *tna2* Δ*lacU169*; G_0	6
ZK819	ZK126 *rpoS819 rpsL* Sm^r; G_I	40
ZK820	ZK126 *rpoS819 gyrA* Nal^r; G_I	40
ZK1141	ZK819 *sgaA sgaB sgaC*; G_{II}	38
ZK2552	ZK819 *ilvGMEDA*$^+$ (spontaneous suppressor) Val^r	This study
ZK2553	ZK819 *bgl*$^+$ (spontaneous suppressor) Bgl^+	This study
ZK2554	ZK1141 *ilvGMEDA*$^+$ (same allele as in ZK2552) Val^r	This study
ZK2555	ZK1141 *bgl*$^+$ (spontaneous suppressor) Bgl^+	This study
ZK2557	ZK2553 *rpoS*$^+$	This study
ZK2559	ZK2553 *sgaB*	This study
ZK2561	ZK2553 *sgaA*	This study
ZK2563	ZK2553 *sgaC*	This study
ZK2564	ZK2553 *sgaA sgaB sgaC*	This study
ZK2618	ZK1141 *trpB*::Tn*10*Tc^r	This study
Other strains		
CAG12173	MG1655 *cysC95*::Tn*10*Tc^r	31
CAG18501	MG1655 *rbsD*::Tn*10*Tc^r	31
CAG18528	MG1655 *zbj-3110*::Tn*10*Kan^r	31
KER176	*rpsL lipA150*::Tn*1000*dKan^r	35
NU1107	VJS433 *serC*::mini-Mu*dI*194 Kan^r	14
ZK173	MB2 (Hfr Δ*gal trpB*::Tn*10* *proC*::Tn*5 lacU169*)	Lab stock
ZK1000	ZK126 Δ*rpoS*::Kan^r	4

dixic acid (20 μg/ml), tetracycline (15 μg/ml), kanamycin (50 μg/ml), or chloramphenicol (30 μg/ml). All chemicals were from Sigma. Optical density (OD) was monitored with a Spectronic 20D+ Spectrophotometer (Milton Roy).

Genetic techniques. Phage P1*vir* transduction and Hfr conjugation using the Singer et al. (31) strain collections were performed as described elsewhere (19). Insertional mutagenesis with mini-Tn*10* transposons from the vectors λNK1323 (Tc^r), λNK1316 (Kan^r), and λNK1324 (Cm^r) was performed as described previously (13).

Construction of GASP strains. Because incorporation of new genetic markers can alter the fitness of bacteria, we constructed our strains such that the final strains differed from the parental strains only by the allele(s) of the GASP loci. This was achieved by first bringing an auxotrophy mutation or the streptomycin-sensitive (Sm^s) allele of *rpsL* that mapped near the GASP loci (see Results) into the recipient by P1*vir* transduction. We could then cotransduce the GASP alleles with P1*vir* grown on ZK1141 or ZK126 into these strains by selecting for prototrophy or Sm^r, and then testing among those transductants for the cotransduction of the GASP allele, by assaying directly for the GASP phenotype or another physiological phenotype where appropriate (see below).

The *rpoS*$^+$ strains were constructed by using the *cysC95*::Tn*10*Tc^r mutation from CAG12173 and the Δ*rpoS*::Kan^r mutation from ZK1000 (4). Strains carrying the *sgaA* allele of ZK1141 were constructed with the *lipA150*::Tn*1000d*Kan^r marker from strain KER176 (35). Mutants with the *sgaA* GASP allele were identified by their larger colony sizes when grown on M63 glutamate (0.5%) plates (see Results). Strains carrying the *sgaB* allele of ZK1141 were constructed with the *serC*::mini-Mu*dI*194 allele from strain NU1107 (14). Mutants with the *sgaB* GASP allele were identified by their increased sensitivity to serine, determined by a filter disc technique and confirmed by assaying for mucoidy at 30°C (see Results). Strains carrying the *sgaC* GASP allele were constructed with the *rpsL*$^+$ (Sm^s) allele of ZK126. Mutants with the *sgaC* GASP allele were identified by scoring for the *sgaC* GASP phenotype (see below).

Stationary-phase competitions. Competition experiments were adapted from those of Zambrano et al. (40). For competitions in LB, initial cultures were inoculated from frozen glycerol stocks into 3 ml of LB and grown overnight. These were then subcultured 1:100 into fresh LB and incubated for 24 h before being mixed for the competitions. The two populations were monitored by serial dilution in M63 medium and plating on minimal salicin and minimal glucose-plus-valine plates. We verified that the majority of the population remained prototrophic (and hence detectable on the selection media) by comparing the counts on the selection media with those on LB. In no case did we observe a difference in total viable counts on minimal and rich media. For competitions in M63-serine (0.5%)-isoleucine (0.03%)-valine (0.03%)-leucine (0.03%)-NaCl (0.5%), colonies were inoculated from an LB plate into the defined medium, and the cultures were incubated until they reached stationary phase (1 to 4 days). Cultures of strain ZK2618 (*trpB*::Tn*10*) were also supplemented with 0.004% tryptophan to facilitate growth. The ZK1141 and ZK2618 cultures were washed in M63 medium before inoculating as a 1:10,000 minority into the ZK820 cultures. The two populations were monitored by serial dilution in M63 medium and plating on LB-streptomycin and LB-nalidixic acid.

Molecular techniques. The DNA flanking the mini-Tn*10*Cm^r transposons was determined by using an arbitrary PCR-based protocol (5), with the modifications described by Pratt and Kolter (28). PCR products were subjected to sequence analysis by the Micro Core Facility, Department of Microbiology and Molecular Genetics, Harvard Medical School, and the sequences were compared with the GenBank DNA database by using the BLAST program (1).

RESULTS

Isolation and mapping of three new GASP loci from an aged *rpoS819* strain. Our approach to investigate the physiology of the GASP phenomenon was to identify and characterize several mutations able to confer the GASP phenotype. That many rounds of GASP takeover occur in starved cultures of *E. coli* (7, 40) implies that the survivors have acquired multiple GASP mutations. As we were interested in exploring potential genetic interactions of accumulated GASP mutations, our strategy was to isolate and investigate the GASP mutations of a single mutant survivor (ZK1141) from a culture that had undergone several rounds of population takeover.

We have devised a nomenclature that describes the relationship of the GASP mutants within a single lineage (Table 2). G_n (short for $GASP_n$) denotes an isolate from an aged culture of strain G_{n-1} that is capable of outcompeting the G_{n-1} strain in a stationary-phase competition. In this study we analyzed the G_{II} strain ZK1141 (*rpoS819 sgaA sgaB sgaC*), which is descended from the G_I strain ZK819 (*rpoS819*), which is descended from the G_0 wild-type strain ZK126 (all G_n designations in this work refer to these strains).

ZK1141 was isolated as a Sm^r survivor of a mixed culture; a 10-day-old culture of ZK819 and a 1-day-old culture of the Sm^s nalidixic acid-resistant (Nal^r) version of ZK819 (ZK820) were grown together in fresh LB and then incubated for a week under starvation conditions (37). ZK1141 has the G_{II} phenotype: it is capable of completely taking over a 1-day-old population of G_I when inoculated as a 1-day-old minority (38).

The mutation responsible for the G_I GASP phenotype of ZK819 has been identified as an allele of *rpoS* called *rpoS819* (40). The G_{II} phenotype of ZK1141 was initially thought to be due to a single mutation, which was termed *sga*, for stationary-phase growth advantage (38). However, genetic analysis of ZK1141 (see below) has demonstrated that there are three GASP mutations that each contribute to the G_{II} GASP phenotype.

We identified the mutations responsible for the G_{II} phenotype of ZK1141 with a genetic selection technique adapted from Zambrano et al. (40): when introduced into the G_I strain, the G_{II} GASP alleles conferred a GASP phenotype versus the G_I parent. We made a pool of approximately 1,000 G_{II} mutants with randomly inserted mini-Tn*10*Tc^r or Kan^r transposons in

TABLE 2. GASP nomenclature

GASP mutant designation	Strain used in this study	Strain history	GASP mutations
G_0	ZK126	Initial strain	None (wild type)
G_I	ZK819	G_0 strain aged in LB	*rpoS819*
G_{II}	ZK1141	G_I strain aged in LB	*rpoS819 sgaA sgaB sgaC*

the chromosome. We then selected for linkage of the mini-Tn*10* to the GASP alleles by making a P1*vir* lysate of the pool and infecting G_I with this lysate to obtain a new pool of about 500 Tc^r or Kan^r transductants. These pools were then inoculated as a minority into 1-day-old ZK820 (the Nal^r Sm^s G_I strain) cultures, and the cultures were allowed to further starve to select from the pool those G_I transductants carrying G_{II} alleles: these transductants could grow in the starved culture, whereas the G_I transductants not carrying the G_{II} GASP alleles could not. The G_I transductants carrying the G_{II} GASP alleles were isolated from the culture several days after the pool was inoculated by titering the culture onto LB-streptomycin *Sm* plates. We then confirmed that the mutants isolated on the LB-streptomycin titer plate had the G_{II} GASP alleles by moving their mini-Tn*10* alleles into a fresh G_I background by P1*vir* transduction and testing among those transductants for cotransduction of the G_{II} GASP phenotype by competition versus the Nal^r G_I. In this manner, we identified three distinct mutations harbored by G_{II} that conferred a GASP phenotype to G_I. These mutations were designated *sgaA*, *sgaB*, and *sgaC*.

Mapping the G_{II} GASP loci. During our investigations of strain ZK1141, we discovered that it has two phenotypes that its ZK819 parent lacks: mucoid growth on glucose at 30°C (but not at 37°C) and an enhanced sensitivity to the amino acid serine. Serine is a competitive inhibitor of homoserine dehydrogenase I, and high levels of intracellular serine result in isoleucine starvation (9, 10). We scored relative serine sensitivities by streaking the strains on an M63-glucose plate toward a filter disc soaked with 10% serine placed in the center of the plate and determining the relative sizes of the growth inhibition zones. Instrumental in the mapping of the *sgaB* locus was the discovery that the *sgaB* GASP allele is responsible for both the mucoidy and serine sensitivity phenotypes of ZK1141. The *sgaB* mutation was mapped to min 20 by using the Hfr and P1 mapping sets (31); the *sgaB* mutation was 95% linked to the *zbj-3110*::Tn*10*Kan^r marker of CAG18528, located at min 19.8 (2, 31).

The *sgaA* and *sgaC* mutations were each mapped by determining the location of random mini-Tn*10*Cm^r insertions linked to the initial mini-Tn*10*Tc^r or mini-Tn*10*Kan^r insertions used to isolate the GASP alleles (see above). These mini-Tn*10*Cm^r markers were mapped by the arbitrarily primed PCR technique. The *sgaA*-linked mini-Tn*10*Cm^r was found to be in *ybdN* at min 13.7 and was 60% linked to the mini-Tn*10*Tc^r, which in turn was 10% linked to *sgaA*. The *sgaC*-linked mini-Tn*10*Cm^r was found to be in *gspA* at min 74.4, which was 66% linked to the mini-Tn*10*Kan^r, which in turn was 50% linked to *sgaA*.

The GASP phenotypes of three new GASP loci: *sgaA*, *sgaB*, and *sgaC*. Having isolated and mapped three GASP alleles of strain G_{II}, we wanted to test whether each of these alleles alone could confer a selective advantage over the G_I parent during stationary phase. We assayed for the GASP phenotype by mixing 24-h-old cultures of the two strains in question and monitoring changes in each population by viable count assay. The two populations were distinguished because they carry different neutral markers. Marker neutrality was confirmed empirically by switching the markers between the strains and performing all mixes reciprocally. Viable counts of each population are determined by titering the culture on the two relevant selection plates. Previous reports suggested that the *rpsL* allele conferring Sm^r and the *gyrA* allele conferring Nal^r are neutral in stationary-phase competitions in *E. coli* (7, 40). We confirmed that the Sm^r and Nal^r markers are neutral during extended starvation. However, during the first 4 days of competition of a 1:1 mix (see below), viable counts were consistently 2- to 10-fold higher for the Sm^r strain than for the Nal^r strain, although the Sm^r counts eventually dropped to equal those of the Nal^r population. Since the first 4 days of the competitions in this study were critical, we differentially marked the ZK819 (Sm^r) strain with two new markers that remain neutral throughout the competition: valine-resistant growth on glucose (Val^r) and the ability to grow on β-glucosides (Bgl^+). The Val^r and Bgl^+ markers were isolated as spontaneous mutations conferring the ability to grow on M63-glucose-valine or M63-salicin plates, respectively. Unless otherwise noted, this pair of markers was used for all competitions described below.

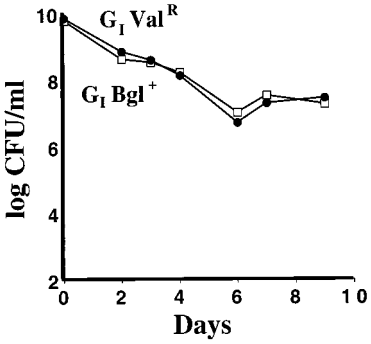


FIG. 1. The Val^r and Bgl^+ markers are selectively neutral in stationary phase. One-day-old cultures of the Val^r (ZK2552) (●) and Bgl^+ (ZK2553) (□) derivatives of the G_I mutant (ZK819) were mixed 1:1 and cocultured. Viable counts were assayed on M63-glucose-valine or M63-salicin plates. Neither strain was outcompeted in four competitions.

The Val^r mutation mapped to the *ilvGMEDA* operon. Since *E. coli* K-12 has a frameshift in *ilvG*, which prevents expression of the one Val^r isozyme of acetohydroxy acid synthase of *E. coli*, encoded by *ilvGM* (16), the Val^r mutation is most likely a suppressor of this frameshift. The Bgl^+ mutants arose at high frequency in our ZK819 and ZK1141 strains (about 1 in 10^7 plated cells), which prevented transduction of the same allele into either background. We therefore selected for Bgl^+ mutants in both ZK819 and ZK1141 backgrounds. Both Bgl^+ mutations mapped to the *bgl* operon and are likely to be insertions or point mutations in the *bglR* regulatory locus (30). The neutrality of the Val^r and Bgl^+ mutations throughout the starvation period was demonstrated by competitions of the Val^r and Bgl^+ derivatives of the ZK819 and ZK1141 strains. Neither marker conferred a competitive advantage or disadvantage, as determined by 1:1 competitions (Fig. 1). Furthermore, neither strain grew as a 1:1,000 minority versus the other strain (data not shown). Finally, neither of the two mutations altered the fitness of the strains versus the Val^s Bgl^- parent (data not shown). Each mix was performed at least four times.

The fact that we can isolate mutants with the GASP phenotype from starved bacterial cultures suggests that these GASP mutants, starting from a single mutant cell, are able to increase in number during the starvation period to establish themselves as the majority population. One assay for the GASP phenotype is thus the ability of the mutant, when placed as a minority population in a culture of the parent, to grow relative to the parent and eventually establish itself as the majority population. To demonstrate this aspect of GASP for each of the G_{II} GASP alleles, we constructed the G_I *sgaA* (the G_I strain that has the *sgaA* allele of G_{II}), G_I *sgaB*, and G_I *sgaC* strains with the Val^r or Bgl^+ selectable marker and competed them as a 1,000-fold minority with the G_I parent that had the other selectable marker. Figure 2 demonstrates that each of the three GASP alleles confers the ability to grow when inoculated

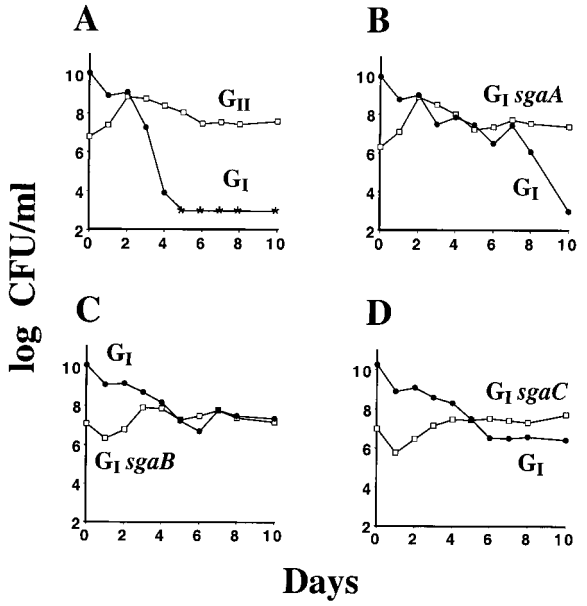


FIG. 2. The G_{II} alleles of ZK1141 confer the GASP phenotype. Into a 1-day-old culture of the G_I mutant (ZK2552) (●) was inoculated as a 1,000-fold minority of a 1-day-old culture of the G_{II} strain (ZK2555) (A), the G_I *sgaA* strain (ZK2561) (B), the G_I *sgaB* strain (ZK2559) (C), or the G_I *sgaC* strain (ZK2563) (□) (D). Asterisks indicate that viable counts fell below detectable levels ($<10^3$ CFU/ml). The patterns of GASP takeovers were identical in six replicate mixtures, including ones where the selectable markers were switched between competing strains.

as a minority and take over the population. Like the G_{II} mutant (Fig. 2A), the G_I *sgaA* mutant grew on the first day of the competition (Fig. 2B), while the G_I *sgaB* and G_I *sgaC* mutants experienced a 1-day lag period before growth (Fig. 2C and D).

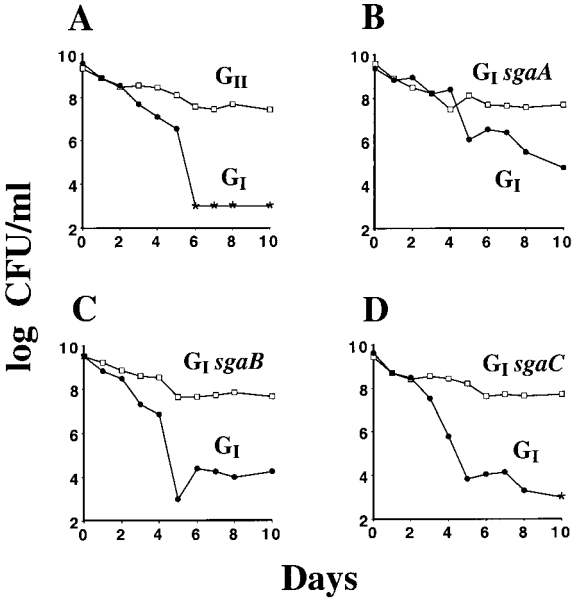


FIG. 3. The G_{II} alleles confer a competitive advantage to G_I cells. A 1-day-old culture of the G_I mutant (ZK2552) (●) was mixed 1:1 with a 1-day-old culture of the G_{II} strain (ZK2555) (A), the G_I *sgaA* strain (ZK2561) (B), the G_I *sgaB* strain (ZK2559) (C), or the G_I *sgaC* strain (ZK2563) (□) (D). Asterisks indicate that viable counts fell below detectable levels ($<10^3$ CFU/ml). The patterns of GASP takeovers were identical in six replicate mixtures, including ones where the selectable markers were switched between the competing strains.

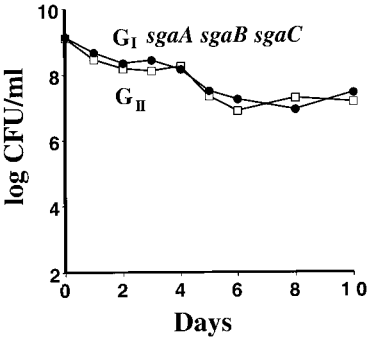


FIG. 4. The *sgaA*, *sgaB*, and *sgaC* alleles are sufficient for the G_{II} GASP phenotype of ZK1141. A 1-day-old culture of the reconstructed G_{II} strain G_I *sgaA sgaB sgaC* (ZK2564) (●) was mixed 1:1 with a 1-day-old culture of the G_{II} strain (ZK2554) (□). Neither strain was outcompeted in eight competitions.

However, none of the G_I mutants with a single G_{II} GASP allele was able to eliminate the majority population as rapidly or as completely as the G_{II} mutant, suggesting that their fitness advantages are additive, a possibility addressed below.

Another component of the GASP phenotype is the ability of the GASP mutant to directly outcompete the parent, resulting in the death of the parent. This ability is assayed by mixing the two cultures in a 1:1 ratio and looking for the decline of one of the two populations during the starvation period. Figure 3 demonstrates that all three of the G_I mutants with a G_{II} GASP allele are capable of outcompeting the G_I parent. Interestingly, during the first 2 days of the 1:1,000 and 1:1 mixes for both G_{II} and G_I *sgaA* mutants (Fig. 2A and B and 3A and B), the GASP mutant could grow as a minority population but did not yet outcompete the G_I parent when mixed in equal numbers. This finding implies that while there may be utilizable carbon for the strains available during the first 2 days, the competition for those nutrients does not become lethal for the parental strain until the environmental conditions change as a result of continued starvation.

The three G_{II} GASP mutations are necessary and sufficient for the G_{II} GASP phenotype. Our selection method identified three new distinct GASP loci on the chromosome of G_{II}: *sgaA*, *sgaB*, and *sgaC*. To determine whether there are additional GASP mutations in G_{II}, we examined whether the three GASP mutations identified so far were both necessary and sufficient for the G_{II} GASP phenotype. Our first approach was to compete the G_{II} mutant with a constructed G_I *sgaA sgaB sgaC* mutant. Neither strain had a competitive advantage when competed in a 1:1 mix (Fig. 4). Furthermore, the G_{II} mutant was unable to grow when inoculated as a 1:1,000 minority into a culture of the G_I *sgaA sgaB sgaC* (data not shown). These results demonstrate that the G_{II} strain lacks any additional GASP mutations that would confer a competitive advantage over the constructed strain. Additionally, the G_I *sgaA sgaB sgaC* strain grew immediately and completely displaced the G_I strain as the majority when inoculated as a 1:1,000 minority into the G_I culture (data not shown); this GASP phenotype was indistinguishable from that of the G_{II} strain (Fig. 2A). The results from these three different competition experiments thus demonstrate that the *sgaA*, *sgaB*, and *sgaC* mutations are sufficient for the G_{II} GASP phenotype.

We next asked if all three G_{II} GASP alleles were necessary for the G_{II} GASP phenotype. We competed 1:1 the G_{II} strain versus one of three constructed strains harboring two of the three G_{II} GASP mutations: G_I *sgaA*, *sgaB*, G_I *sgaA sgaC*, or G_I *sgaB sgaC*. In every case, the G_{II} strain outcompeted the con-

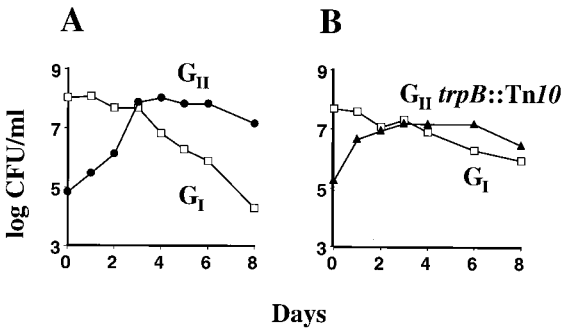


FIG. 5. GASP in a chemically defined medium. Stationary-phase cultures grown in M63-serine (0.5%)-isoleucine (0.03%)-valine (0.03%)-leucine (0.03%)-NaCl (0.5%) (plus tryptophan [0.004%] for the G_{II} *trpB*::Tn*10* strain [ZK2618]) were mixed 1:10,000 to assay the GASP phenotype. Both the G_{II} strain (ZK1141; ●) (A) and the tryptophan auxotrophic G_{II} *trpB*::Tn*10* strain (ZK2618; ▲) (B) express the GASP phenotype as a minority versus the G_I strain (ZK820; □). The patterns of GASP takeovers were identical in four competitions.

structed strain (data not shown). Hence, all three reconstructed strains, each lacking one of the three G_{II} alleles, were less fit than G_{II}, indicating that each of the three G_{II} alleles is necessary for the full fitness gain of the G_{II} strain. That each is necessary for the G_{II} GASP phenotype indicates that they act additively to confer higher and higher fitnesses in stationary phase.

GASP mutants obtain nutrients from dying cells in a chemically defined medium. Because growth in LB ceases due to carbon limitation (37), it has been assumed that the GASP mutants obtain the nutrients required for growth from the dying majority population. To demonstrate directly that the dying cells release nutrients which the GASP mutants can utilize during growth, we identified a chemically defined medium in which the G_{II} GASP strain could grow and outcompete the G_I strain. Figure 5A shows the GASP phenotype of the G_{II} strain versus G_I after both cultures were grown to stationary phase in M63 salts medium supplemented with serine (0.5%), isoleucine, valine, and leucine (0.03% each), and NaCl (0.5%). As in LB, the G_{II} strain grew rapidly as a 1:10,000 minority to take over the population. Like the prototroph, a tryptophan auxotrophic derivative of the G_{II} strain (*trpB*::Tn*10*) was able to grow and take over the G_I minority (Fig. 5B). This finding indicates that the G_{II} strain can scavenge enough tryptophan to meet its growth requirement during the population takeover. As no exogenous tryptophan was supplied by the medium, the only remaining source of tryptophan is the cells of the dying majority population. This is the first evidence that the GASP mutants can scavenge nutrients released by the dying cells during carbon starvation.

The three G_{II} GASP alleles and the *rpoS819* allele confer faster growth on amino acids. We reasoned that the primary selective force acting on carbon-starved cells is the ability to utilize the carbon released by the dying majority population. Therefore, we asked whether the GASP mutations confer faster growth on carbon sources that may resemble the nutrients released by dying cells. The three G_{II} GASP mutations were tested in the G_I background, as it was from this background that G_{II} was selected. We also tested the growth phenotypes of the G_I GASP mutation, *rpoS819*, by comparing its growth rate with that of the G_I strain carrying the *rpoS*$^+$ allele (this is essentially a G_0 strain, as the *rpoS819* mutation is sufficient to confer the G_I GASP phenotype of ZK819 [38]). We first assayed growth on LB, a rich medium containing many of the building blocks, vitamins, and energy sources necessary for growth. The growth rates on LB were indistinguishable between the different mutant strains (data not shown).

The composition of the medium during prolonged carbon starvation has not been characterized. However, we speculated that amino acids are the most abundant nutrients released by the dying cells, given that amino acids account for most of the dry weight of *E. coli* (26). Hence, we determined the relative growth rates of the GASP mutants on mixtures of amino acids, either as a combination of monomers and short peptides (tryptone) or only as monomers (Casamino Acids). While none of the individual GASP alleles had much of an effect, the G_{II} strain harboring all three G_{II} GASP mutations grew significantly faster than G_I on the monomer and peptide mixture (Table 3). In contrast, all four GASP mutations conferred significantly faster growth on the mixture of the amino acid monomers. Interestingly, in every case, the strains with more GASP mutations grew faster, indicating that the GASP mutations act additively to confer faster growth on mixtures of amino acids.

Since all four GASP mutations confer faster growth on mixtures of amino acids, we attempted to identify individual amino acids that the GASP mutants could catabolize more rapidly. We assayed growth in liquid media containing M63 salts and the amino acid in question at 0.5% (wt/vol). Because of the potential problem of amino acids such as serine and cysteine inhibiting isoleucine biosynthesis and thus preventing growth on single amino acids (9, 10), we supplemented all of the growth media with isoleucine (0.03%). We assayed for the ability of the GASP mutants to grow on each of the 20 amino acids singly as the sole source of carbon and energy. At least one of the six strains tested could grow on alanine, asparagine, aspartate, glutamate, glutamine, proline, serine, or threonine; none grew on the other 11 amino acids. While *E. coli* K-12 can grow on tryptophan (32), our strains were not expected to grow, as they lack tryptophanase activity due to the *tna2* mutation.

All four GASP alleles conferred growth advantages on several amino acids, manifested as a higher growth rate or, in some cases, a new ability to grow on the particular amino acid (Table 4). The *rpoS819* GASP allele conferred upon the cell the new ability to utilize asparagine and glutamine as sole sources of carbon and energy. However, the *rpoS819* mutants grew to an OD at 600 nm of only about 0.3 on glutamine. It is therefore uncertain whether the cells grew incompletely on

TABLE 3. Relative growth rates on amino acid mixtures

Carbon and energy source (1.0%)	Relative growth rate (±SD, $n = 3$)[a]					
	G_0 (ZK2557)	G_I (ZK2553)	G_I *sgaA* (ZK2561)	G_I *sgaB* (ZK2559)	G_I *sgaC* (ZK2563)	G_{II} (ZK2555)
Tryptone	1.00 (±0.01)	0.98 (±0.01)	1.00 (±0.03)	1.03 (±0.01)	1.03 (±0.02)	1.14 (±0.03)
Casamino Acids	1.00 (±0.02)	1.16 (±0.01)	1.34 (±0.01)	1.31 (±0.05)	1.21 (±0.02)	1.43 (±0.02)

[a] A relative growth rate of 1.00 indicates a generation time (in hours) of 0.59 on tryptone or 1.28 on Casamino Acids.

TABLE 4. Relative growth rates on single amino acids as carbon sources

Amino acid	Relative growth rate (±SD, $n = 2$)[a]					
	G_0 (ZK2557)	G_I (ZK2553)	G_I *sgaA* (ZK2561)	G_I *sgaB* (ZK2559)	G_I *sgaC* (ZK2563)	G_{II} (ZK2555)
Alanine	1.00 (±0.01)	1.04 (±0.02)	0.88 (±0.00)	1.25 (±0.01)	1.11 (±0.02)	1.21 (±0.01)
Asparagine	0	1.00 (±0.03)	1.90 (±0.04)	0.69 (±0.03)	0.72 (±0.04)	1.60 (±0.03)
Aspartate	0	0	1.00 (±0.01)	0	0	0.53 (±0.00)
Glutamate	1.00 (±0.03)	2.93 (±0.01)	9.14 (±0.00)	1.92 (±0.07)	2.17 (±0.03)	9.29 (±0.00)
Glutamine[b]	0	1.00 (±0.02)	0.87 (±0.04)	0.75 (±0.09)	0.76 (±0.03)	0.82 (±0.00)
Serine	1.00 (±0.05)	1.63 (±0.36)	0.86 (±0.06)	9.81 (±0.66)	2.30 (±0.14)	9.19 (±0.28)
Threonine	1.00 (±0.02)	2.17 (±0.11)	1.16 (±0.02)	3.66 (±0.00)	3.92 (±0.11)	5.01 (±0.04)
Proline[c]	1.00 (±0.1)	0.72 (±0.11)	2.1 (±0.1)	0.45 (±0.05)	1.6 (±0.1)	10.5 (±0.7)

[a] For all amino acids except proline (see footnote *c*), growth rates in liquid M63 plus 0.5% amino acid (plus 0.03% isoleucine) medium were determined. A relative growth rate of 1.00 indicates a generation time (in hours) of 4.50 on alanine, 14.0 on asparagine, 11.2 on aspartate, 58.5 on glutamate, 12.0 on glutamine, 61.2 on serine, or 304 on threonine; a relative growth rate of zero indicates no growth.

[b] Cultures stopped growing at an OD of 600 nm of 0.2 to 0.3. This was not due to alkalinization of the medium, as the pH remained below 7.5, nor was it due to end product inhibition, as the cultures grew readily when spiked with 0.2% glucose.

[c] Overnight LB cultures were diluted into M63 medium and plated on M63 proline (0.5%) plus isoleucine (0.03%). Values indicate relative colony areas (±SD, $n = 20$) after 4 days of incubation.

glutamine or grew on an impurity in the glutamine supply instead. The *rpoS819* allele alone conferred faster growth on glutamate, serine, threonine, and alanine. The *sgaA* allele conferred the new ability to grow on aspartate and conferred faster growth on asparagine and glutamate. The *sgaB* and *sgaC* alleles both conferred faster growth on alanine, threonine, and serine. Comparison of the growth rates for the G_0, G_I, and G_{II} strains indicates that both the repertoire of amino acids and the growth rates on the amino acids increase as more GASP alleles are added and demonstrates that these GASP alleles act additively to increase the overall capacity to catabolize amino acids.

We were not able to obtain relative growth rates on proline in liquid cultures, because the cultures were consistently taken over by faster-growing mutants. However, we were able to obtain estimates of relative growth rates by comparing the sizes (surface area) of colonies grown on M63-proline (plus isoleucine) plates (Table 4). The *sgaA* and *sgaC* alleles conferred faster growth on proline, and their effects on growth were strikingly additive, as the G_{II} colonies were significantly larger than those of the G_I mutants with either single allele alone.

Interestingly, the four GASP alleles conferred slower growth on several of the amino acids (Table 4). The *rpoS819* mutation conferred slower growth on proline; the *sgaA* mutation conferred slower growth on alanine, glutamine, and serine; the *sgaB* mutation conferred slower growth on asparagine, glutamate, glutamine, and proline; and the *sgaC* mutation conferred slower growth on asparagine, glutamate, and glutamine. This indicates that while the individual GASP alleles confer fitness gains on some amino acids, each also confers a fitness loss on other amino acids. We discuss the implications of these observations below.

DISCUSSION

Previous work identified an allele of *rpoS*, *rpoS819*, as a mutation that can confer the GASP phenotype on *E. coli* (40). Genetic analysis of an isolate (G_{II}) from a starved culture of the *rpoS819* GASP strain (G_I) has revealed three new GASP mutations: *sgaA*, *sgaB*, and *sgaC*. All four GASP mutations of this isolate map to different regions of the chromosome, suggesting that there are many loci that when mutated can confer fitness advantages in stationary phase.

As each of the three new GASP mutations acquired by the G_{II} strain can confer a GASP phenotype on the G_I parental strain, it is most likely that they were acquired as a result of three successive GASP takeover events. G_{II} was isolated from a culture starved for a total of about 2.5 weeks, which suggests that in our experimental system population takeovers during starvation can happen faster than once per week. The fact that the three G_{II} GASP mutations are necessary (and sufficient) for the G_{II} GASP phenotype demonstrates directly that the acquisition of multiple GASP mutations can provide successively higher fitnesses for starved bacteria. The continual accumulation of GASP mutations by cells within the surviving population can account for the multiple rounds of population takeovers observed in starved cultures (7, 40). Our results thus support the growing body of evidence that starved populations are highly dynamic and undergo frequent population takeovers as a result of fitness differences among the competing subpopulations (7, 8, 40).

A major purpose of our investigation was to understand how GASP mutations alter cell physiology to provide fitness gains in stationary phase. Previous studies have demonstrated that fitness gains during conditions of limited substrate availability in chemostats or selections for growth on novel substrates were manifested as increases in catabolic potential for those substrates (11, 17, 20, 24, 33, 36). In our system, we have observed that dying cells in stationary-phase cultures release nutrients (e.g., tryptophan) that can be utilized by the growing GASP mutants, and we hypothesized that the GASP mutants selected are those that outcompete their parents for these limited substrates. We observed a direct correlation between relative GASP fitness and relative ability to catabolize amino acids, which are likely to be the most abundant nutrients released by dying cells (26). The G_I GASP mutation *rpoS819* and the three G_{II} GASP mutations *sgaA*, *sgaB*, and *sgaC* all confer higher growth rates on an amino acid mixture (Casamino Acids). To our knowledge, this is the first report that mutations in *rpoS* can alter amino acid catabolism during exponential growth, extending previous observations that wild-type σ^S affects the physiology of both arrested and growing cells (12, 27).

All four of the GASP mutations examined in this study also increase the ability of the cell to utilize certain amino acids singly as the sole source of carbon and energy. These changes were manifested as either a higher growth rate or a new ability to grow on the particular amino acid. The GASP mutations are pleiotropic in this respect, as each affects growth on several amino acids. The subsets of amino acids on which they confer a growth advantage are overlapping but distinct. In general,

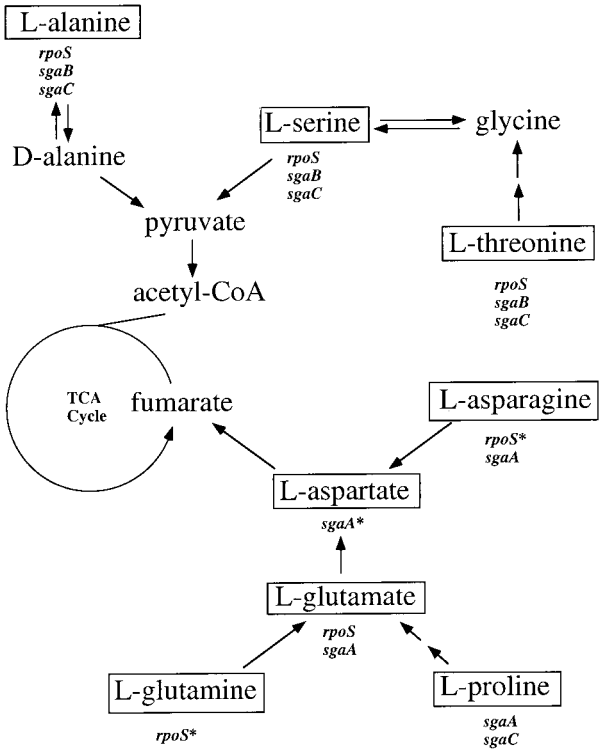


FIG. 6. The primary catabolic pathways for the amino acids on which the GASP mutants have a growth advantage (reviewed by McFall and Newman [18]). The arrows indicate enzymatic steps between metabolites. Boxed amino acids denote amino acids that can serve as sole sources of carbon and energy for at least one of the six strains tested. Listed under the boxed amino acids are the loci whose GASP alleles confer a growth advantage on the amino acid. An asterisk indicates the GASP allele that confers the novel ability to grow on the amino acid.

these subsets contain amino acids that enter catabolism along the same major degradative pathways (reviewed in reference 18) (Fig. 6). The *sgaA* allele confers a growth advantage on amino acids that enter the central metabolic pathway through aspartate and fumarate, while the *sgaB* and *sgaC* alleles confer growth advantages on amino acids that enter this pathway through pyruvate (the effect of *sgaC* on proline metabolism is the one exception). The *rpoS819* mutation, on the other hand, affected both of these degradative pathways. It is tempting to speculate that the GASP mutations, especially *rpoS819*, alter the regulation of the enzymes of these degradative pathways. We are currently investigating the mechanistic bases of these physiological changes.

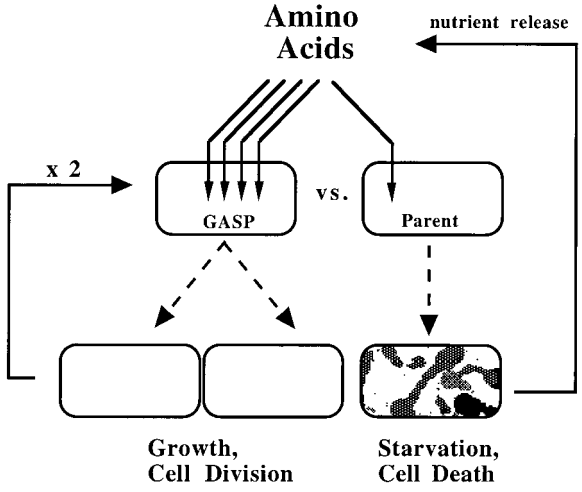


FIG. 7. A model for the physiological basis of the GASP phenotype.

Based on our findings, we propose a model for a physiological basis of GASP (Fig. 7). We hypothesize that the primary selective force acting on carbon-starved cells is the ability to scavenge carbon sources released by the dying cells for the purposes of cell maintenance and growth. Cells unable to obtain a sufficient supply of carbon and energy can no longer maintain activities essential for viability, and they die. The fact that in every case studied the GASP fitness correlates directly with the capacity to catabolize amino acids leads us to propose that the most significant nutrients released are catabolizable amino acids. At the onset of starvation, the population is composed almost entirely of cells of the parental genotype that compete with equal fitness for carbon sources. Hence, the death during the first few days of starvation is stochastic, as all parental cells have an equal chance of scavenging the nutrients and surviving. However, the rare mutants within the population expressing the GASP phenotype outcompete their parents for the carbon resources because of their enhanced catabolic capabilities. These advantages provide the GASP cells with enough resources not only to survive but to grow and divide during starvation conditions. Once the GASP mutants grow to a significant cell density, they effectively decrease the amount of carbon available to the parental cells. These parental cells die as a result and release their nutrients into the medium. This model involves two positive feedback loops which can account for both the rapid growth of the GASP mutant and the rapid death of the parent. Supporting our model is the finding that if the G_{II} strain ZK1141 lacks the respiratory enzyme NADH dehydrogenase I, essential for the utilization of several amino acids catabolized by G_{II} (29), it loses the GASP phenotype versus the G_{I} parental strain, ZK819 (38).

Interestingly, we observed that while G_{I} mutants with a single G_{II} allele grow faster than G_{I} with certain amino acids as sole carbon sources, they grow slower with others. This result may seem inconsistent with our model for the physiological basis of GASP. However, growth on mixed amino acids may more accurately reflect the growth of the GASP mutants during starvation, since all amino acids are likely to be released at similar rates from the dying cells. We have demonstrated that all four GASP alleles confer an overall advantage when growing on mixed amino acids (Casamino Acids). Hence, our results suggest that when the GASP mutants are competing with their parents for the complex mixtures of nutrients released by the dying cells, their faster growth on some amino acids outweighs their slower growth on others.

ACKNOWLEDGMENTS

We thank J. E. Cronan, Jr., M. E. Winkler, and A. Wright for providing strains. We thank S. E. Finkel, G. A. O'Toole, and L. A. Pratt for critical reading of the manuscript and members of the Kolter lab for helpful comments.

This work was supported by grants from the National Science Foundation (MCB9728936) and the National Institutes of Health (GM55199).

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Berlyn, M. K., K. B. Low, and K. E. Rudd.** 1996. Linkage map of *Escherichia coli* K-12, edition 9, p. 1715–1902. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley,

M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. ASM Press, Washington, D.C.
3. **Blackburn, N., T. Fenchel, and J. Mitchell.** 1998. Microscale nutrient patches in planktonic habitats shown by chemotactic bacteria. Science **282:**2254–2256.
4. **Bohannon, D. E., N. Connell, J. Keener, A. Tormo, M. Espinosa-Urgel, M. M. Zambrano, and R. Kolter.** 1991. Stationary-phase-inducible "gearbox" promoters: differential effects of *katF* mutations and the role of σ^{70}. J. Bacteriol. **173:**4482–4492.
5. **Caetano-Annoles, G.** 1993. Amplifying DNA with arbitrary oligonucleotide primers. PCR Methods Appl. **3:**85–92.
6. **Connell, N., Z. Han, F. Moreno, and R. Kolter.** 1987. An *E. coli* promoter induced by the cessation of growth. Mol. Microbiol. **1:**195–201.
7. **Finkel, S. E., and R. Kolter.** 1999. Evolution of microbial diversity during prolonged starvation. Proc. Natl. Acad. Sci. USA **96:**4023–4027.
8. **Finkel, S. E., E. Zinser, S. Gupta, and R. Kolter.** 1997. Life and death in stationary phase. Mol. Microbiol. **103:**3–16.
9. **Hama, H., T. Kayahara, M. Tsuda, and T. Tsuchiya.** 1991. Inhibition of homoserine dehydrogenase I by L-serine in *Escherichia coli*. J. Biochem. **109:**604–608.
10. **Hama, H., Y. Sumita, Y. Kakutani, M. Tsuda, and T. Tsuchiya.** 1990. Target of serine inhibition in *Escherichia coli*. Biochem. Biophys. Res. Commun. **168:**1211–1216.
11. **Helling, R. B., C. N. Vargas, and J. Adams.** 1987. Evolution of *Escherichia coli* during growth in a constant environment. Genetics **116:**349–358.
12. **Hengge-Aronis, R.** 1996. Regulation of gene expression during entry into stationary phase, p. 1497–1512. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
13. **Kleckner, N., J. Bender, and S. Gottesman.** 1991. Uses of transposons with emphasis on Tn*10*. Methods Enzymol. **204:**139–180.
14. **Lam, H. M., and M. E. Winkler.** 1990. Metabolic relationships between pyridoxine (vitamin B_6) and serine biosynthesis in *Escherichia coli*. J. Bacteriol. **172:**6518–6528.
15. **Lange, R., and R. Hengge-Aronis.** 1991. Growth phase-regulated expression of *bolA* and morphology of *Escherichia coli* cells is controlled by the novel sigma factor, σ^S (*rpoS*). J. Bacteriol. **173:**4474–4481.
16. **Lawther, R. P., D. H. Calhoun, C. W. Adams, C. A. Hauser, J. Gray, and G. W. Hatfield.** 1981. Molecular basis of valine resistance in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **78:**922–925.
17. **Lenski, R. E., J. A. Mongold, P. D. Sniegowski, M. Travisano, F. Vasi, P. J. Gerrish, and T. M. Schmidt.** 1998. Evolution of competitive fitness in experimental populations of *E. coli*: what makes one genotype a better competitor than another? Antonie Leeuwenhoek **73:**35–47.
18. **McFall, E., and E. B. Newman.** 1996. Amino acids as carbon sources, p. 358–379. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
19. **Miller, J. H.** 1992. A short course in bacterial genetics. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
20. **Miller, R. D., D. E. Dykhuizen, L. Green, and D. L. Hartl.** 1984. Specific deletion occurring in the directed evolution of 6-phosphogluconate dehydrogenase in *Escherichia coli*. Genetics **108:**765–772.
21. **Moriarty, D. J. W., and R. T. Bell.** 1993. Bacterial growth and starvation in aquatic environments, p. 25–53. *In* S. J. Kjelleberg (ed.), Starvation in bacteria. Plenum Press, New York, N.Y.
22. **Morita, R. Y.** 1988. Bioavailability of energy and its relationship to growth and starvation survival in nature. Can. J. Microbiol. **34:**436–441.
23. **Morita, R. Y.** 1993. Bioavailability of energy and the starvation state, p. 1–23. *In* S. Kjelleberg (ed.), Starvation in bacteria. Plenum Press, New York, N.Y.
24. **Mortlock, R. P.** 1982. Metabolic acquisitions through laboratory selection. Annu. Rev. Microbiol. **36:**259–284.
25. **Mulvey, M. R., P. A. Sorby, B. L. Triggs-Raine, and P. C. Loewen.** 1988. Cloning and physical characterization of *katE* and *katF* required for catalase HPII expression in *Escherichia coli*. Gene **73:**337–345.
26. **Neidhardt, F. C., and H. E. Umbarger.** 1996. Chemical composition of *Escherichia coli*, p. 13–28. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
27. **Notley, L., and T. Ferenci.** 1996. Induction of RpoS-dependent functions in glucose-limited continuous culture: what level of nutrient limitation induces the stationary phase of *Escherichia coli*? J. Bacteriol. **178:**1465–1468.
28. **Pratt, L. A., and R. Kolter.** 1998. Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. Mol. Microbiol. **30:**285–293.
29. **Prüß, B. M., J. M. Nelms, C. Park, and A. J. Wolfe.** 1994. Mutations in NADH:ubiquinone oxidoreductase of *Escherichia coli* affect growth on mixed amino acids. J. Bacteriol. **176:**2143–2150.
30. **Reynolds, A. E., J. Felton, and A. Wright.** 1981. Insertion of DNA activates the cryptic *bgl* operon of *E. coli*. Nature **293:**625–629.
31. **Singer, M., T. A. Baker, G. Schnitzler, S. M. Deischel, M. Goel, W. Dove, K. J. Jaacks, A. D. Grossman, J. W. Erickson, and C. A. Gross.** 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. Microbiol. Rev. **53:**1–24.
32. **Snell, E. E.** 1975. Tryptophanase: structure, catalytic activities and mechanisms of action. Adv. Enzymol. **42:**287–333.
33. **Sonti, R. V., and J. R. Roth.** 1989. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. Genetics **123:**19–28.
34. **Taylor, B. L.** 1983. How do bacteria find the optimal concentration of oxygen? Trends Biochem. Sci. **8:**438–441.
35. **Vanden Boom, T. J., K. E. Reed, and J. E. Cronan, Jr.** 1991. Lipoic acid metabolism in *Escherichia coli*: isolation of null mutants defective in lipoic acid biosynthesis, molecular cloning and characterization of the *E. coli lip* locus, and identification of the lipoylated protein of the glycine cleavage system. J. Bacteriol. **173:**6411–6420.
36. **Weikert, C., U. Sauer, and J. E. Bailey.** 1997. Use of a glycerol-limited, long-term chemostat for isolation of *Escherichia coli* mutants with improved physiological properties. Microbiology **143:**1567–1574.
37. **Zambrano, M. M.** 1993. Ph.D. thesis. Harvard University, Cambridge, Mass.
38. **Zambrano, M. M., and R. Kolter.** 1993. *Escherichia coli* mutants lacking NADH dehydrogenase-I have a competitive disadvantage in stationary phase. J. Bacteriol. **175:**5642–5647.
39. **Zambrano, M. M., and R. Kolter.** 1996. GASPing for life in stationary phase. Cell **86:**181–184.
40. **Zambrano, M. M., D. A. Siegele, M. Almirón, A. Tormo, and R. Kolter.** 1993. Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. Science **259:**1757–1760.

Week 6

Sporulation

Secretion

Cell, Vol. 121, 401–409, May 6, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.02.032

Developmental Commitment in a Bacterium

Jonathan Dworkin[1,2,*] and Richard Losick[1,*]
[1]Department of Molecular and Cellular Biology
Harvard University
Cambridge, Massachusetts 02138
[2]Department of Microbiology
College of Physicians and Surgeons
Columbia University
New York, New York 10032

Summary

We investigated developmental commitment during sporulation in *Bacillus subtilis*. Sporulation is initiated by nutrient limitation and involves division of the developing cell into two progeny, the forespore and the mother cell, with different fates. Differentiation becomes irreversible following division when neither the forespore nor the mother cell can resume growth when provided with nutrients. We show that commitment is governed by the transcription factors σ^F and σ^E, which are activated in the forespore and the mother cell, respectively. We further show that commitment involves *spoIIQ*, which is under the control of σ^F, and *spoIIP*, which is under the control of both σ^F and σ^E. In the presence of nutrients, the forespore can exhibit rodlike, longitudinal growth when SpoIIQ and SpoIIP are absent, whereas the mother cell can do so when SpoIIP alone is absent. Thus, developmental commitment of this single-celled organism, like that of the cells of complex, multicellular organisms, ensures that differentiation is maintained despite changes in the extracellular milieu.

Introduction

A fundamental challenge in developmental biology is understanding the mechanisms that cause differentiation to become irreversible (Gilbert, 2000). As Hans Spemann (1918) observed almost a century ago, transplantation of prospective epidermal cells from an early newt gastrula to another region of the embryo changes their fate in accordance with their new location. In contrast, transplanted prospective epidermal cells from a later-stage gastrula retain their original fate (Spemann, 1918). Thus, as development progresses, the cells become committed to becoming epidermal cells despite the change in their environment. Similarly, more recent experiments in other systems, such as zebrafish, show that when single cells taken from an embryo at early stage of gastrulation are transplanted to a different location, their fate changes, but when cells at a later developmental stage are transplanted, they remain committed to their original fate (Ho and Kimmel, 1993). Whereas regulatory proteins that are involved in commitment are known in several systems (Tam et al., 2003), the identity of the target genes that directly act to prevent the reversal of cell fate have largely remained elusive.

An attractive developmental system in which to attempt to identify genes that mediate commitment is the process of sporulation in the bacterium *Bacillus subtilis*. In response to conditions of nutrient limitation, cells of *B. subtilis* cease growing and instead enter a developmental pathway that culminates in the formation of a spore. A hallmark of sporulation is the formation of an asymmetrically positioned (polar) septum that divides the developing cell into dissimilarly sized progeny called the forespore (the smaller cell) and the mother cell, which follow different pathways of differentiation (left hand of Figure 1A) (Errington, 2003; Piggot and Losick, 2002). The forespore is eventually engulfed by the mother cell, where it proceeds to develop into the spore. The mother cell facilitates the conversion of the forespore into a spore but ultimately undergoes lysis, releasing the spore, when development is complete. Cells that have started to sporulate but have not yet formed the polar septum are capable of resuming vegetative growth when provided with nutrients, that is, they are not committed. On the other hand, developing cells that have passed the point of asymmetric division (postdivisional sporangia) are obliged to complete spore formation even when transferred to nutrient-rich medium (Parker et al., 1996).

Shortly after the formation of the polar septum, the transcription factor σ^F is activated in the forespore (Margolis et al., 1991), where it switches on the expression of about 16 genes (Piggot and Losick, 2002; S. Wang, P. Eichenberger, T. Sato, and R.L., unpublished data). Among the genes turned on by σ^F is that (*spoIIR*) for an intercellular signaling protein (Londono-Vallejo and Stragier, 1995) that causes the activation in the mother cell of the transcription factor σ^E (Figure 1A). The σ^E factor, in turn, switches on an unusually large regulon consisting of at least 262 genes (Eichenberger et al., 2003; Eichenberger et al., 2004). Here we report that σ^F and σ^E are separately responsible for rendering differentiation in the forespore and the mother cell irreversible. We show that σ^F can commit the forespore to its fate by switching on genes, including *spoIIP* and *spoIIQ*, that block growth and division in the forespore. We further show that *spoIIP* is under the control of both σ^F and σ^E, allowing it to be expressed in both the forespore and the mother cell, and that the action of *spoIIP* commits the mother cell to its fate.

Results

Commitment in the Forespore Is Governed by σ^F

Bacteria lacking either σ^F or σ^E are unable to complete differentiation and instead undergo a second round of asymmetric division in which an additional polar septum is formed at the opposite pole (right hand of Figure 1A). This second division creates an aberrant three-chamber sporangium with two forespore-like compart-

*Correspondence: dworkin@microbiology.columbia.edu (J.D.); losick@mcb.harvard.edu (R.L.)

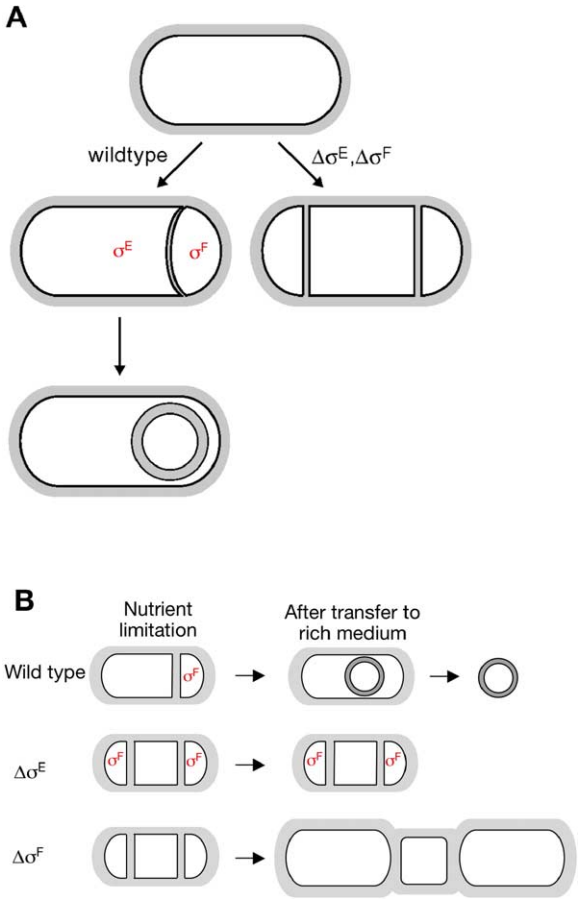


Figure 1. Cell Fate and Polar Division

(A) During sporulation of the wild-type (left), a single septum forms near one pole of the sporangium, creating a forespore (to the right of the septum) and a mother cell (to the left). In subsequent development, the forespore is engulfed by the mother cell and pinched off to create a cell within a cell. During sporulation of cells lacking σ^F or σ^E (due to a deletion of the *sigF* or *sigE* genes, respectively), a second polar septum is formed, creating a disporic sporangium that is blocked from further development (right).

(B) During sporulation of the wild-type (upper row), transfer to rich medium after a polar septum has formed does not prevent sporulation from continuing to completion. Sporulation of cells lacking σ^E ($\Delta\sigma^E$; middle row) yields disporic sporangia that are blocked in subsequent development and are unable to resume growth when transferred to rich medium. Finally, sporulation of cells lacking σ^F ($\Delta\sigma^F$; bottom row) yields disporic sporangia that are able to reinitiate longitudinal growth when transferred to rich medium.

ments at the poles, which each receive a chromosome (leaving the middle compartment devoid of DNA) (Setlow et al., 1991). We found that when such "disporic" sporangia of a σ^F mutant were transferred by dilution into rich medium, the polar compartments were able to reinitiate longitudinal growth (Figure 1B and Figure 2A, upper panels; Table 1, rows 2 and 4) followed by medial division. By contrast, disporic sporangia of a σ^E mutant failed to grow or divide when transferred to rich medium (Figure 1B and Figure 2A, lower panels; Table 1, rows 1 and 3). To confirm that the vegetative-like outgrowth observed in the σ^F mutants derived from fore-

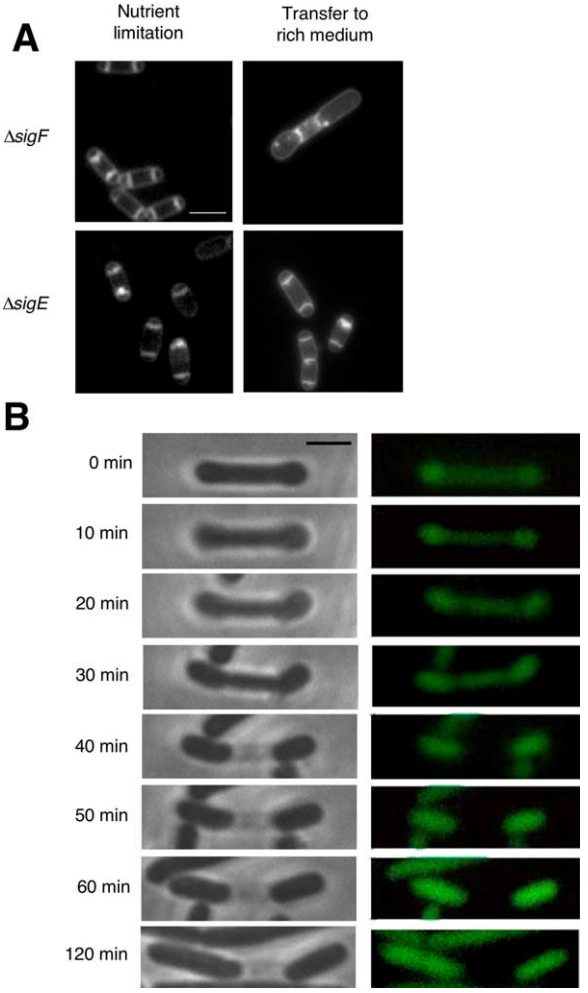


Figure 2. Disporic Sporangia Lacking σ^F Are Able to Reinitiate Growth in the Presence of Excess Nutrients

(A) Shown are sporangia that have resumed rodlike growth from a strain (RL1265) lacking σ^F (Δ*sigF*) or a strain (RL1061) lacking σ^E (Δ*sigE*). Samples were collected for fluorescence microscopy after 3 hr in minimal (sporulation) medium or after an additional 2 hr following dilution into rich (growth) medium and treated with the vital membrane stain TMA-DPH. Scale bar = 3 μm.

(B) Shown are time-lapse phase (left) and fluorescent (right) images of a sporangium from a strain (MF2279) lacking σ^F and carrying a fusion of *gfp* to the promoter P_{abrB}. The *abrB* promoter is active during growth and repressed during sporulation; thus, lower GFP fluorescence is indicative of a cell in sporulation, and higher GFP fluorescence is indicative of a cell that has resumed vegetative growth. Cells were grown for 3 hr in liquid minimal sporulation medium and then diluted 1:10 into rich medium and grown at 37°C for 45 min. Cells were concentrated by centrifugation, placed on an agarose pad (1% agarose in rich medium) maintained at 37°C, and visualized at regular intervals by time-lapse phase contrast and epifluorescence microscopy. As can be seen (left images), both polar forespores grew out into elongated cells, whereas the central mother-cell compartment, which in disporic sporangia lacks a chromosome, underwent lysis. At the time of dilution into rich medium, the disporic cell contained a comparatively fainter GFP fluorescence indicative of entry into sporulation (right images). At subsequent time points, the fluorescence grew brighter, indicative of a return to vegetative growth. Scale bar = 2 μm.

Table 1. Quantitative Analysis of Commitment

Strain	Genotype	Time in Rich Medium (hr)	Disporic	Vegetative	Outgrown
RL1061	$\Delta\sigma^E$	0	92	36	0
RL1265	$\Delta\sigma^F$	0	130	28	0
RL1061	$\Delta\sigma^E$	1.5	40	39	0
RL1265	$\Delta\sigma^F$	1.5	2	46	24
RL1061	$\Delta\sigma^E$	0	61	22	0
JDB967	$\Delta\sigma^E$Δ*spoIIQ*Δ*spoIIP*	0	53	21	0
RL1061	$\Delta\sigma^E$	2.5	30	102	0
JDB967	$\Delta\sigma^E$Δ*spoIIQ*Δ*spoIIP*	2.5	16	198	33

Samples were collected for fluorescence microscopy after 3 hr in minimal sporulation medium and after an additional 1.5 hr or 2.5 hr following dilution into rich growth medium; they were then treated with the vital membrane stain TMA-DPH. Cells that contained two recognizable, complete, asymmetric septa were categorized as "disporic," cells with medial septa were categorized as "vegetative," and cells with at least one polar compartment that had become rodlike were categorized as "outgrown."

spores, we conducted time-lapse microscopy. When a disporic σ^F-mutant sporangium placed on an agarose pad composed of a rich medium was visualized at regular intervals over 2 hr, longitudinal outgrowth of the forespore compartments was observed (Figure 2B, left panels). These cells contained, in addition, a fusion of *gfp* to a promoter (*abrB*) that is active during vegetative growth and repressed during sporulation (Strauch et al., 1990). Consistent with our interpretation that the polar compartments were forespores that were resuming vegetative growth, fluorescence from GFP in the two polar compartments was comparatively faint at the start of the time-lapse sequence and became progressively brighter as the polar cells elongated and returned to vegetative growth (Figure 2B, right panels).

spoIIQ and *spoIIP* Contribute to Blocking Growth and Division of the Forespore

Although σ^E-mutant sporangia are similar in appearance to σ^F-mutant sporangia, σ^E-mutant sporangia are able to activate σ^F in both polar compartments (Figure 1B). This suggested that the difference between σ^F and σ^E mutants in their potential to resume growth was due to a gene(s) under the control of the forespore-specific transcription factor that prevented growth. To identify this gene(s), we constructed a series of strains that were mutant for σ^E and that also carried a mutation in one of the previously known (Piggot and Losick, 2002) members of the σ^F regulon (*spoIIQ*, *spoIIIG*, *dacF*, *csfA*, *csfB*, *spoIIR*, *sspA*, *sspB*, *sspE*, and *rsfA*) or in one of the genes that were recently assigned to the regulon through microarray analysis (*yffL*, *yisN*, *ywmF*, *ykwF*, *yqhH*, and *yphA* [S. Wang, P. Eichenberger, T. Sato, and R.L., unpublished data]). *spoIIQ* was a particularly attractive candidate because its product, a membrane protein with similarity to metallo-endopeptidases, plays an important role in remodeling of the cell wall of the forespore (Londono-Vallejo et al., 1997; Rubio and Pogliano, 2004). However, neither a *spoIIQ* mutation (Figure 3A, upper panels) nor a mutation in any of the other σ^F-controlled genes tested allowed any disporic sporangia to undergo longitudinal growth when placed in rich growth medium.

Another attractive candidate was *spoIIP*, which encodes a membrane protein that is involved in preventing a second round of asymmetric division in wild-type sporulating cells. Strains lacking *spoIIP* exhibit a significantly higher frequency of bipolar and partially bipolar septa, and premature expression of *spoIIP* along with two other genes, *spoIID* and *spoIIM*, inhibits formation of the initial sporulation septum (Eichenberger et al., 2001; Pogliano et al., 1999). Whereas *spoIIP* is normally considered to be a σ^E-controlled gene (Frandsen and Stragier, 1995), and its transcription in the mother cell contributes to preventing the disporic phenotype (Eichenberger et al., 2001; Pogliano et al., 1999), it lies immediately downstream (62 bp) of a gene under σ^F control (*gpr*), and there are no obvious transcriptional terminators in the intergenic region between the two genes (Figure 4A). Also, prior genetic evidence (Frandsen and Stragier, 1995) as well as gene microarray analysis (S. Wang, P. Eichenberger, T. Sato, and R.L., unpublished data) suggested that *spoIIP* might exhibit a second mode of regulation in which it is transcribed under the control of the forespore-specific transcription factor by readthrough from *gpr*. Accordingly, we built a σ^E-mutant strain that also carried a *spoIIP* mutation, but the resulting sporangia were indistinguishable in their terminal differentiation phenotype from those of a strain that was mutant for σ^E alone (Figure 3A, middle panels).

We hypothesized that commitment could be due to the combined activity of more than one gene under σ^F control. We therefore constructed a series of σ^E-mutant strains that carried mutations in two σ^F-controlled genes. Whereas almost all of these strains failed to exhibit growth of their forespores when transferred into rich medium, in forespores of a σ^E-mutant strain that additionally carried mutations in both *spoIIP* and *spoIIQ*, swelling and/or pronounced rodlike, longitudinal elongation could be observed (Figure 3A, lower panels; Table 1, rows 6 and 8). In fact, some of the elongating forespores of the disporic sporangia eventually underwent medial division (Figure 3A, lower right panel), a hallmark of vegetative growth, and by 2.5 hr after transfer to rich medium, comparatively few disporic sporangia remained (Table 1, row 8). These findings suggest that the action of both genes contributes to preventing postdivisional sporangia from undergoing outgrowth. We note, however, that cells lacking *spoIIQ* and *spoIIP* were noticeably less efficient in exhibiting outgrowth than cells lacking σ^F (Table 1, rows 4 and 8),

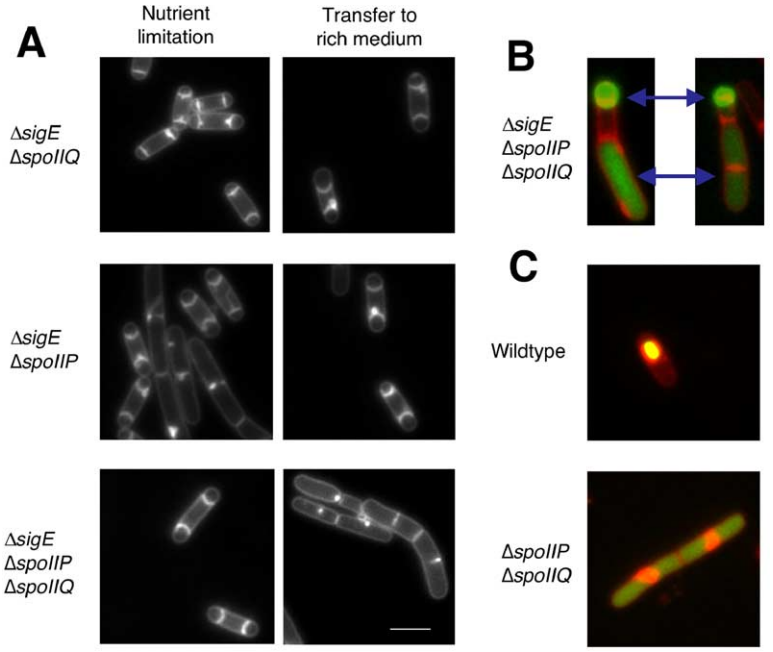


Figure 3. The *spoIIP* and *spoIIQ* Genes Contribute to Commitment in the Forespore

(A) Shown are sporangia that have resumed longitudinal growth from strains lacking σ^E and mutant for *spoIIQ* (Δ*spoIIQ*; JDB919), *spoIIP* (Δ*spoIIP*; JDB963), or doubly mutant for both *spoIIP* and *spoIIQ* (Δ*spoIIP* Δ*spoIIQ*; JDB967). Scale bar = 3μm.

(B) Shown are two sporangia after 2.5 hr in rich medium from a strain (JDB972) lacking σ^E, containing mutations of *spoIIP* and *spoIIQ*, and carrying a fusion of *gfp* to a promoter under σ^F control (P_{spoIIQ}-*gfp*). One forespore compartment of each of the disporic sporangia (bottom arrow) resumed longitudinal growth and contains GFP fluorescence that had been diluted during growth, whereas the other forespores (top arrow) had not undergone growth and therefore contain brighter GFP signals that have not been diluted. Note that the middle compartments lack GFP fluorescence, as would be expected for mother cells in which σ^F is not active. Scale bar = 3 μM. Of cells exhibiting green fluorescence observed in this experiment, 9 (33%) were forespores of disporic sporangia that had not grown, 16 (60%) were outgrowing forespores, and 2 (7%) were cases in which fluorescence was seen in the mother cell, which evidently arose from leakage from the forespore.

(C) Shown are a sporangium from a wild-type strain (RL2382) that had continued through engulfment in rich medium (top) and a sporangium from a strain (JDB1047) carrying mutations in *spoIIP* and *spoIIQ* (Δ*spoIIP* Δ*spoIIQ*) that had resumed vegetative growth after 2 hr in rich medium (bottom). Both strains carried P_{spoIIQ}-*gfp*. Of 152 wild-type cells examined, 122 were rodlike and lacked significant green fluorescence indicative of the progeny of vegetative cells that had not committed to sporulation at the time of transfer to rich medium and were able to resume (or continue) vegetative growth rapidly and undergo multiple rounds of division during the period of growth in rich medium. Thirty cells in the control population were undergoing or had undergone engulfment, as indicated by the green fluorescence signal in their forespore. Of 168 Δ*spoIIP* Δ*spoIIQ* cells examined, 80 were rodlike and contained green fluorescence above background, and 88 were rodlike and lacked green fluorescence. The latter is indicative of cells that had initiated sporulation but returned to vegetative growth after transfer to rich medium.

suggesting that one or more additional genes under σ^F control contribute to commitment.

Cytoplasmically Inherited Green Fluorescent Protein as a Lineage Reporter

To confirm that these elongating, rod-shaped cells arose from cells that had entered sporulation, we introduced a reporter containing the gene (*gfp*) for the green fluorescent protein (GFP) under control of σ^F, the forespore-specific transcription factor. Two examples of rod-shaped cells in which the presence of green fluorescence indicates that the cells originated from forespores can be seen in Figure 3B (bottom arrow). In each of these disporic sporangia, a forespore compartment (bottom arrow) resumed longitudinal growth (and in one case had undergone binary fission [right panel]). This growth diluted the GFP fluorescence so that it was weaker than that seen in the other forespores (top arrow) that had not grown out and whose GFP therefore had not been diluted. The middle compartments lacked GFP fluorescence, as would be expected for mother cells in which σ^F is not active. Although leakage of GFP from the forespore to mother cell has been observed in some sporulation mutants, this phenomenon typically only occurs following extended periods (>8 hr) in sporulation medium (Li et al., 2004), and while we could observe some sporangia with GFP fluorescence in the mother cell, such sporangia were rare. Thus, *spoIIQ spoIIP*-mutant cells exhibiting green fluorescence were most likely to have derived from a forespore cell that had entered sporulation and activated σ^F but had not proceeded further into sporulation. In other words, in this strain, GFP can be considered to be a cytoplasmically inherited lineage reporter that certifies that a growing cell arose from a cell that had σ^F activity.

We examined whether mutations in *spoIIP* and *spoIIQ* would prevent commitment in sporangia that were otherwise wild-type, that is, sporangia that were not lacking σ^E. When wild-type cells harboring a construct in which *gfp* was under the control of σ^F were transferred into rich medium at 2.25 hr after the start of sporulation, postdivisional sporangia in which σ^F was activated were observed to continue through further stages of sporulation, including engulfment of the forespore by the mother cell (note the fluorescence of the engulfed forespore in the upper panel of Figure 3C). In contrast, sporangia of a *spoIIQ spoIIP* double mutant (once again carrying the *gfp* construct) that were treated in the same way could revert to growth (note the fluorescence of the dividing cells in the lower panel of Figure 3C). Thus, otherwise wild-type cells lacking *spoIIP* and *spoIIQ* that have initiated sporulation and have activated σ^F are not necessarily committed to sporulation and appear to be able to resume rodlike elongation when exposed to excess nutrients.

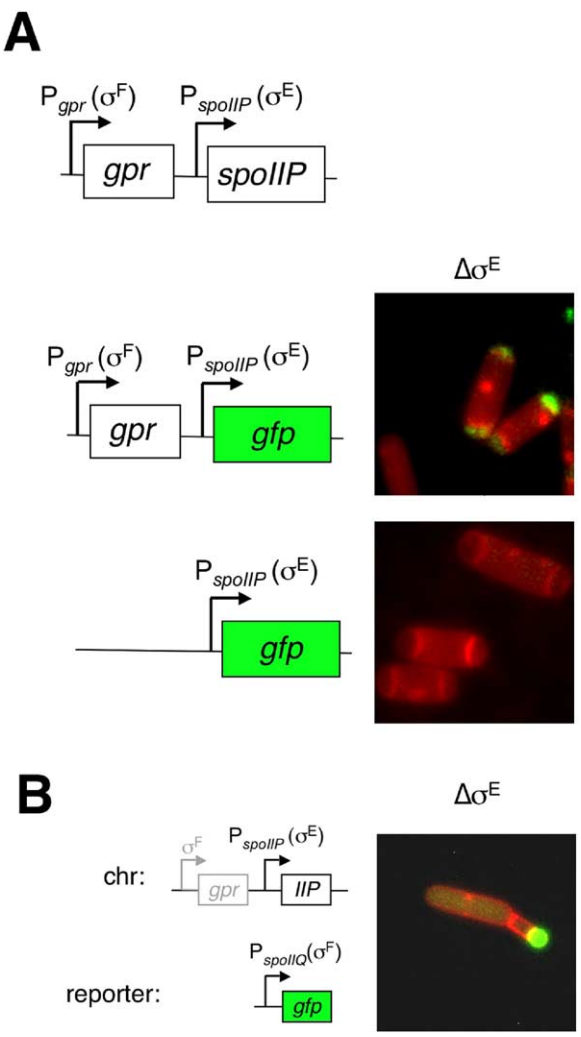


Figure 4. Readthrough from the *gpr* Promoter Drives Transcription of *spoIIP* in the Forespore

(A and B) Sporangia were processed for fluorescence microscopy after 3 hr in sporulation medium as in Figure 2.

(A) The *spoIIP* gene is transcribed from a σ^E-controlled promoter (P_{spoIIP}) located immediately upstream of *spoIIP* and a σ^F-controlled promoter (P_{gpr}) located just upstream of *gpr*. Shown are sporangia from strains lacking σ^E and carrying *gfp* fused to the 1462 bp of DNA upstream of *spoIIP* including the σ^F-dependent P_{gpr}, the *gpr* gene, and the σ^E-dependent P_{spoIIP} (JDB1080; upper) or carrying a fusion of *gfp* to the 183 bp of DNA upstream of *spoIIP* including the σ^E-dependent P_{spoIIP} (JDB1082; lower).

(B) Sporangium from a strain (JDB1025) lacking σ^E, *spoIIQ*, and *gpr* and containing a fusion of *gfp* to a promoter under σ^F control (P_{spoIIQ}-*gfp*).

Expression of *spoIIP* in the Forespore

The idea that *spoIIP* is involved in the developmental commitment of the forespore requires that *spoIIP* be in fact expressed in the forespore under the control of σ^F. We confirmed previous work (Frandsen and Stragier, 1995) suggesting that some transcription of *spoIIP* arises by readthrough from the upstream, σ^F-controlled *gpr* gene (data not shown) and extended this analysis further by fusing the chromosomal region upstream of *spoIIP*, including the entire *gpr* gene and its promoter, to a promoterless copy of *gfp*. We then introduced the *gfp*-containing construct into the chromosome (at a nonessential locus) of a σ^E mutant. The fusion was expressed in the forespore as would be expected if expression of the fusion were solely under control of σ^F (Figure 4A, middle panel). Consistent with this interpretation, when the construct was introduced into cells that were mutant for σ^F (and therefore also lacked σ^E activity since activation of σ^E is dependent on σ^F), little or no production of GFP was detected (data not shown). When *gfp* was fused to the intergenic region between *gpr* and *spoIIP* (and hence lacked the σ^F-dependent *gpr* promoter), once again little or no production of GFP was detected in the absence of σ^E (Figure 4A, lower panel). In toto, these results confirm that *spoIIP* is subject to two modes of expression: σ^E-directed transcription from a promoter located immediately upstream of the gene and σ^F-directed readthrough transcription from the promoter of the adjacent upstream gene.

If this readthrough transcription is critical for the contribution of *spoIIP* to commitment, then a deletion mutation that removes *gpr* and its promoter but leaves *spoIIP* intact should, when tested in combination with a *spoIIQ* mutation, result in a defect in commitment. *gpr* is dispensable for sporulation (Sussman and Setlow, 1991), and, indeed, a deletion that spans both *gpr* and its promoter had no measurable effect on sporulation, which indicates that development does not depend on expression of *spoIIP* in the forespore. However, when the *gpr* deletion was introduced into a strain that was mutant for *spoIIQ* and the resulting double-mutant strain allowed to sporulate, sporangia that had reached the stage of asymmetric division were capable of exhibiting rodlike elongation when introduced into rich medium (Figure 4B).

As a final test of the role of readthrough transcription in commitment, we examined a strain in which *spoIIP* (but not *gpr*) was moved to another site on the chromosome. At this ectopic location, *spoIIP* is fully functional in sporulation (Abanes-De Mello et al., 2002), but when a *spoIIQ* mutation was introduced, postdivisional sporangia from the resulting strain were able to reinitiate growth in the presence of excess nutrients despite the presence of a functional copy of *spoIIP* (data not shown). We conclude that transcription of *spoIIP* from its own promoter suffices for sporulation but that readthrough transcription from the *gpr* promoter is necessary for the role of *spoIIP* in commitment.

spoIIP Blocks Growth and Division in the Mother Cell

We then examined the requirements for commitment in the mother cell. To identify cells that had initiated the mother-cell program of gene expression, we introduced a reporter containing *gfp* under the control of σ^E, the mother-cell-specific transcription factor. When sporulating cells that had reached the stage at which this reporter had been activated were transferred to rich medium, they proceeded through the later stages of sporulation, such as engulfment (Figure 5, upper panels), and no sporangia that resumed vegetative growth were seen. By contrast, when postdivisional sporangia

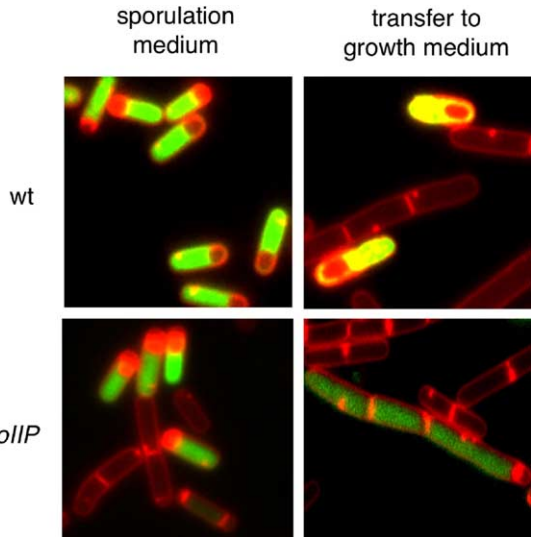


Figure 5. Role of *spoIIP* in Mother-Cell Commitment

Shown are sporangia from an otherwise wild-type strain (JDB869) and a strain (JDB973) mutant for *spoIIP* (Δ*spoIIP*), each carrying *gfp* fused to a promoter under the control of σ^E (P_{spoIID}-*gfp*). Sporangia were processed for fluorescence microscopy as in Figure 2. For quantitative analysis of the *spoIIP* mutant strain, 221 cells were examined microscopically at hour 2 of sporulation. Of these rodlike cells, 54 (or 25%) exhibited a GFP signal above background.

of a *spoIIP*-mutant strain (once again carrying the *gfp* reporter under σ^E control) were exposed to excess nutrients, vegetative-like cells with medial septa that exhibited green fluorescence (and hence were of mother-cell origin) were readily observed (Figure 5, lower panels). Whereas the absence of both *spoIIP* and *spoIIQ* was necessary to prevent forespore elongation, the absence of *spoIIP* alone was sufficient to disrupt commitment in the mother cell. Indeed, the absence of *spoIIP* was more effective in abrogating commitment in the mother cell than was the corresponding effect on the forespore of the absence of both *spoIIP* and *spoIIQ*.

Cell-Specific Expression of *spoIIP* Orthologs from *B. anthracis*

Interestingly, the related spore-forming bacterium *Bacillus anthracis* has two genes, BA2068 and BA3102, that are orthologous to *B. subtilis spoIIP*, and both are turned on during sporulation (Liu et al., 2004). There is extensive synteny between the *B. subtilis* and *B. anthracis* genomes (Read et al., 2003), but neither ortholog is located near *gpr* in *B. anthracis* (Figures 6Aa and 6Ab) even though the chromosomal regions around *gpr* are otherwise highly similar in both organisms (Figures 6Ac and 6Ad). We note, however, that the upstream regions of BA2068 and BA3102 contain sequences that match the consensus for σ^E (Eichenberger et al., 2003) and σ^F (S. Wang, P. Eichenberger, T. Sato, and R.L., unpublished data) controlled promoters, respectively. To investigate the regulation of BA2068 and BA3102 directly, we fused their putative promoter regions to *gfp* and introduced those constructs into a nonessential chromosomal locus of *B. subtilis*. In sporulating cells carrying a fusion of *gfp* to DNA corresponding to the region upstream of BA3102, a GFP signal was observed in the forespore in some cells (Figure 6B, right panel). In contrast, fusion of *gfp* to DNA corresponding to the region upstream of BA2068 resulted in mother-cell-specific production of GFP (Figure 6B, left panel). Thus, *B. anthracis* apparently has evolved a strategy different from that of *B. subtilis* to ensure that *spoIIP* expression occurs in both compartments: it uses two copies of the gene, one under σ^F and one under σ^E control. It remains to be seen, however, whether commitment in *B. anthracis* is governed by a mechanism similar to that described here for *B. subtilis*. In another related spore-forming bacterium, *Clostridium difficile*, the ortholog of the σ^F-controlled *spoIIQ* gene is located immediately adjacent to a gene (*spoIID*) that is known to be under σ^E control, and *spoIIP* is even closer to *gpr* (15 bp) than it is in *B. subtilis* (62 bp). Thus, in *C. difficile* both *spoIIQ* and *spoIIP* are likely to be expressed in both compartments of the sporangium (Stragier, 2002).

Discussion

Viewed in light of our results, σ^F is emerging as the master regulator for the establishment of cell fate during sporulation. The σ^F factor directs the expression of genes that drive differentiation of the forespore as well as the expression of the signaling gene that triggers the activation of σ^E in the mother cell and hence unleashes the mother-cell program of differentiation (Errington, 2003; Piggot and Losick, 2002). Now we see that two of the genes switched on by σ^F, *spoIIP* and *spoIIQ*, have a previously unrecognized role in causing the forespore program of differentiation to become irreversible (Figure 7). One of these genes, *spoIIP*, is also expressed in the mother cell, where it plays a parallel role in preventing the reversal of differentiation (Figure 7). Thus, *spoIIP* plays a critical role in commitment in both compartments, but its function in commitment of the forespore is masked by redundancy with *spoIIQ*. This requirement for an additional factor(s) in the forespore may reflect the weak level of expression of *spoIIP* in the forespore. In addition, the less-efficient outgrowth observed in a *spoIIP spoIIQ*-mutant strain as compared to a σ^F-mutant strain indicates that an additional, yet-to-be-identified, forespore-expressed gene(s) contributes to preventing outgrowth.

Although the precise biochemical nature of the growth and division block mediated by the *spoIIP* and *spoIIQ* gene products is not known, both are membrane proteins that play a direct role in the modification of the peptidoglycan that surrounds the forespore during engulfment (Abanes-De Mello et al., 2002; Rubio and Pogliano, 2004), and the *spoIIP* gene product is part of the mechanism that prevents formation of a second septum in the mother cell (Eichenberger et al., 2001; Pogliano et al., 1999). In fact, premature expression during sporulation of *spoIIP*, along with *spoIIM* and *spoIID*, inhibits formation of the initial polar septum and causes distension of the cell wall (Eichenberger et al., 2001; Pogliano et al., 1999). Conceivably, the effect of SpoIIP and SpoIIQ on growth results from an ability to directly inhibit enzymes involved in peptidoglycan syn-

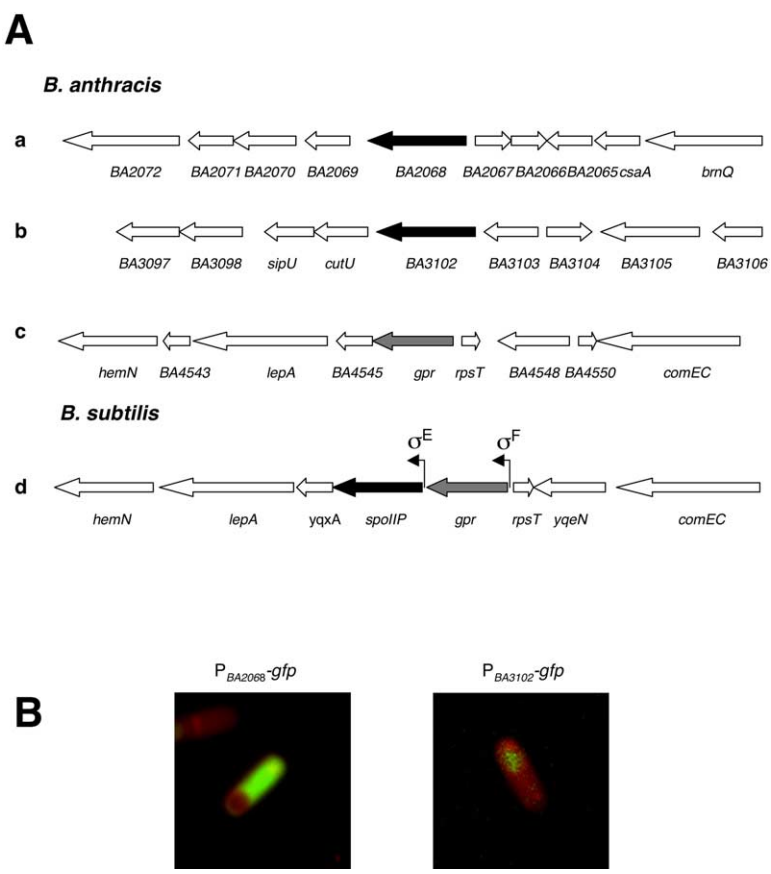


Figure 6. The Presence of Two *spoIIP* Genes in *B. anthracis*

(A) Organization of the chromosome in the vicinity of the *B. anthracis spoIIP* orthologs BA2068 (a) and BA3102 (b), near the *B. anthracis* ortholog of *gpr* (c), and near *B. subtilis gpr* (d).

(B) Shown are sporangia from *B. subtilis* strains carrying *gfp* fused to the promoters of the *B. anthracis spoIIP* orthologs BA3102 (JDB1104) and BA2068 (JDB1105) after 3 hr in sporulation medium. Sporangia were processed for fluorescence microscopy as in Figure 2.

thesis, as has been described for the E lysis protein of phage ϕX174 (Bernhardt et al., 2001).

Our findings take on added significance in view of the recent discovery that cells that have entered sporulation (predivisional sporangia) secrete factors that induce sibling cells that have not yet initiated sporulation to lyse (Gonzalez-Pastor et al., 2003). Nutrients released by lysis delay predivisional sporangia from proceeding further into sporulation. However, once the sporangia have reached the stage of asymmetric division and have activated σ^F and σ^E, they employ expression of *spoIIP* and *spoIIQ* as a means to ensure that they are committed to completing spore formation even in the presence of nutrients released by lysing sibling cells. Thus, developmental commitment of a single-celled organism, like that of the cells of complex, multicellular organisms, ensures that differentiation is maintained despite changes in the extracellular milieu.

Experimental Procedures

Strains

B. subtilis strains were derivatives of the wild-type strain PY79 and are listed in Table S1. JDB919 was constructed by transforming RL1061 with genomic DNA from RL2022 (*spoIIQΔ::spc*, laboratory stock). JDB963 was constructed by transforming RL1061 with genomic DNA from RL2373 (*spoIIPΔ::tet*, laboratory stock). JDB967 was constructed by transforming JDB919 with genomic DNA from RL2373. JDB972 was constructed by transforming JDB967 with genomic DNA from a strain derived from RL2382 where the spec^R was switched to a cm^R gene through the use of pCm::Sp (Steinmetz and Richter, 1994). JDB973 was constructed by transforming JDB869 with genomic DNA from RL2373. JDB1025 was constructed in several steps. First, long-flanking-homology PCR (Wach, 1996) was used to replace codons 1–324 of *gpr* as well as 248 bp upstream of the start codon with a spec^R gene. Genomic DNA from this strain was then used to transform JDB919. Finally, this strain was transformed with genomic DNA from a strain carrying P_{spoIIQ}-*gfp* where spec^R was switched to tet^R through the use of plasmid pCm::Tc (Steinmetz and Richter, 1994). JDB1047 was constructed by transforming RL2022 with genomic DNA from RL2373. JDB1080 was constructed by transforming RL1061 with pCB47. JDB1082 was constructed by transforming RL1061 with pCB48. JDB1104 was constructed by transforming PY79 with pCB52. JDB1105 was constructed by transforming PY79 with pCB54. MF2279 was constructed by transformation of RL1265 with chromosomal DNA from MF1179 (P_{abrB}-*gfp spc*) that was generated by introducing plasmid pMF175 into *amyE* by double recombination.

Plasmids

Plasmid constructions were performed in *E. coli* DH5α using standard methods. pCB47 contains P_{1462}-*gfp*, which is the 1462 bp upstream of the *spoIIP* ribosome binding site (RBS) fused to *gfp*, and was created by amplifying genomic DNA from PY79 using primers ojd787 (5′-GGCGCTAGCGATTGCAGTACGCATAGCAG-3′) and ojd790 (5′-GTCGCATGCGCGCTTGTCTAGTAATTACTC-3′). This PCR-amplified DNA was digested with NheI and SphI and cloned into pCB45 that contains NheI and SphI sites upstream of *gfpmut2* with an optimized RBS in the *amyE* integration vector pLD30 (Garsin et al., 1998). pCB48 contains P_{183}-*gfp*, which is the 183 bp upstream of the *spoIIP* RBS fused to *gfp*, and was created by amplifying genomic DNA from PY79 using primers ojd788 (5′-GGCGCTAGCGGC CACAACTTAATGGTTAC-3′) and ojd790. This PCR-amplified DNA was digested with NheI and SphI and cloned into pCB45. pCB52 contains the 580 bp upstream of the *B. anthracis* BA3102 RBS

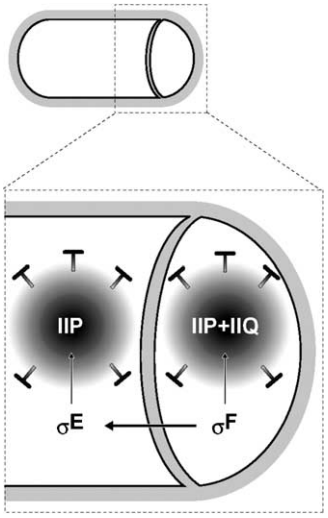


Figure 7. Role of *spoIIP* and *spoIIQ* in Establishment of Cell Fate during Sporulation

The σ^F factor turns on the synthesis of the *spoIIP* (IIP) and *spoIIQ* (IIQ) gene products in the forespore as well as the synthesis of an intercellular signaling protein (horizontal arrow) that causes the activation of σ^E in the mother cell. The σ^E factor, in turn, turns on the synthesis of the *spoIIP* (IIP) gene product in the mother cell. The *spoIIP* and *spoIIQ* gene products likely block cell-wall growth and septation.

fused to *gfp* and was created by amplifying genomic DNA from *B. anthracis* strain Sterne (kind gift of P. Hanna, University of Michigan) using primers ojd775 (5′-GGCGAATTCCCTTGTCTTGATTGG AGTG-3′) and ojd776 (5′-GGCAAGCTTCTTCATTTACTATATTCTT ACG-3′). This PCR-amplified DNA was digested with EcoRI and HinDIII and cloned into pCB44 that contains *gfpmut2* with an optimized RBS cloned into the HinDIII and BamHI sites of the *amyE* integration vector pLD30 (Garsin et al., 1998). pCB54 contains the 565 bp upstream of the *B. anthracis* BA2068 RBS fused to *gfp* and was created by amplifying genomic DNA from *B. anthracis* strain Ames using primers ojd777 (5′-GGCGAATTCGAATGCTACC AACAATCAGC-3′) and ojd778 (5′-GGCAAGCTTCTCTCTATTTCA AGTTATGTAC-3′). This PCR-amplified DNA was digested with EcoRI and HindIII and cloned into pCB44. To construct pMF175 (P_{abrB}-*gfp spc*), a larger DNA fragment derived from pMF35 (P_{spacC}-*gfp spc*; Fujita and Losick, 2002) digested with EcoRI and HindIII was ligated to a smaller DNA fragment derived from pMF172 (P_{abrB}-*lacZ spc*; Fujita et al., 2005) digested with the same restriction enzymes.

General Methods

All PCR reactions were performed with *pfu* DNA polymerase (Stratagene). Preparation and transformation of *B. subtilis* competent cells were as described (Harwood and Cutting, 1990). Sporulation and growth were carried out at 37°C. Cells were grown in hydrolyzed casein (CH) growth medium (Harwood and Cutting, 1990). Experimental cultures were inoculated with a portion of an overnight culture grown in CH at 25°C to an OD600 of 0.05, and when cultures reached an OD600 of ~0.6, they were resuspended in an equal volume of preheated resuspension medium (Sterlini and Mandelstam, 1969) and allowed to grow with aeration. To transfer sporulating cells to a rich medium, 1 ml of a sporulating culture was added to 9 ml of Luria Broth (LB) preheated to 37°C and then allowed to grow with aeration.

Fluorescence Microscopy

Fluorescence microscopy was performed as described previously (Dworkin and Losick, 2002).

Comparative Genomics

B. subtilis genome sequence was obtained from the SubtiList website (http://genolist.pasteur.fr/SubtiList/). *B. anthracis* strain Ames genome sequences were obtained from the NCBI website (http://www.ncbi.nlm.nih.gov/genomes/static/eub_g.html). To identify orthologs of *B. subtilis spoIIP* in related genomes, its sequence was blasted against the selected bacterial genome using the tblastn program on the NCBI website.

Supplemental Data

Supplemental Data include one table and are available with this article online at http://www.cell.com/cgi/content/full/121/3/401/DC1/.

Acknowledgments

We are very grateful to M. Fujita for contributing the experiment in Figure 2B; P. Hanna for *B. anthracis* DNA; Matthieu Piel for assistance with time-lapse microscopy; K. Pogliano and S. Wang for strains; S. Wang for access to unpublished data; L. Shapiro, P. Stragier, and P. Eichenberger for comments on an earlier version of the manuscript; and P. Eichenberger for helpful discussions regarding comparative genomics. This work was supported by NIH grant GM18568 to R.L. and by Institutional Startup Funds from the Columbia University Department of Microbiology to J.D.

Received: August 27, 2004
Revised: November 11, 2004
Accepted: February 17, 2005
Published: May 5, 2005

References

Abanes-De Mello, A., Sun, Y.L., Aung, S., and Pogliano, K. (2002). A cytoskeleton-like role for the bacterial cell wall during engulfment of the *Bacillus subtilis* forespore. Genes Dev. *16*, 3253–3264.

Bernhardt, T.G., Struck, D.K., and Young, R. (2001). The lysis protein E of phi X174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis. J. Biol. Chem. *276*, 6093–6097.

Dworkin, J., and Losick, R. (2002). Does RNA polymerase help drive chromosome segregation in bacteria? Proc. Natl. Acad. Sci. USA *99*, 14089–14094.

Eichenberger, P., Fawcett, P., and Losick, R. (2001). A three-protein inhibitor of polar septation during sporulation in *Bacillus subtilis*. Mol. Microbiol. *42*, 1147–1162.

Eichenberger, P., Jensen, S.T., Conlon, E.M., van Ooij, C., Silvaggi, J., Gonzalez-Pastor, J.E., Fujita, M., Ben-Yehuda, S., Stragier, P., Liu, J.S., and Losick, R. (2003). The sigmaE regulon and the identification of additional sporulation genes in *Bacillus subtilis*. J. Mol. Biol. *327*, 945–972.

Eichenberger, P., Fujita, M., Jensen, S.T., Conlon, E.M., Rudner, D.Z., Wang, S.T., Ferguson, C., Haga, K., Sato, T., Liu, J.S., and Losick, R. (2004). The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS Biol. *2*, e328. 10.1371/journal.pbio.0020328.

Errington, J. (2003). Regulation of endospore formation in *Bacillus subtilis*. Nat. Rev. Microbiol. *1*, 117–126.

Frandsen, N., and Stragier, P. (1995). Identification and characterization of the *Bacillus subtilis spoIIP* locus. J. Bacteriol. *177*, 716–722.

Fujita, M., and Losick, R. (2002). An investigation into the compartmentalization of the sporulation transcription factor sigmaE in *Bacillus subtilis*. Mol. Microbiol. *43*, 27–38.

Fujita, M., Gonzalez-Pastor, J.E., and Losick, R. (2005). High- and low-threshold genes in the spo0A regulon of *Bacillus subtilis*. J. Bacteriol. *187*, 1357–1368.

Garsin, D.A., Duncan, L., Paskowitz, D.M., and Losick, R. (1998). The kinase activity of the antisigma factor SpoIIAB is required for activation as well as inhibition of transcription factor sigmaF during sporulation in *Bacillus subtilis*. J. Mol. Biol. *284*, 569–578.

Gilbert, S.A. (2000). Developmental Biology, Sixth Edition (Sunderland, MA: Sinauer Associates).

Gonzalez-Pastor, J.E., Hobbs, E.C., and Losick, R. (2003). Cannibalism by sporulating bacteria. Science *301*, 510–513.

Harwood, C.R. and Cutting, S.M., eds. (1990). Molecular Biological Methods for Bacillus (New York: John Wiley & Sons).

Ho, R.K., and Kimmel, C.B. (1993). Commitment of cell fate in the early zebrafish embryo. Science *261*, 109–111.

Li, Z., Di Donato, F., and Piggot, P.J. (2004). Compartmentalization of gene expression during sporulation of *Bacillus subtilis* is compromised in mutants blocked at stage III of sporulation. J. Bacteriol. *186*, 2221–2223.

Liu, H., Bergman, N.H., Thomason, B., Shallom, S., Hazen, A., Crossno, J., Rasko, D.A., Ravel, J., Read, T.D., Peterson, S.N., et al. (2004). Formation and composition of the *Bacillus anthracis* endospore. J. Bacteriol. *186*, 164–178.

Londono-Vallejo, J.A., and Stragier, P. (1995). Cell-cell signaling pathway activating a developmental transcription factor in *Bacillus subtilis*. Genes Dev. *9*, 503–508.

Londono-Vallejo, J.A., Frehel, C., and Stragier, P. (1997). SpoIIQ, a forespore-expressed gene required for engulfment in *Bacillus subtilis*. Mol. Microbiol. *24*, 29–39.

Margolis, P., Driks, A., and Losick, R. (1991). Establishment of cell type by compartmentalized activation of a transcription factor. Science *254*, 562–565.

Parker, G.F., Daniel, R.A., and Errington, J. (1996). Timing and genetic regulation of commitment to sporulation in *Bacillus subtilis*. Microbiol. *142*, 3445–3452.

Piggot, P.J., and Losick, R. (2002). Sporulation genes and intercompartmental regulation. In Bacillus subtilis and Its Closest Relative: From Genes to Cells, A.L. Sonenshein, J.A. Hoch, and R. Losick, eds. (Washington, DC: ASM Press), pp. 483–518.

Pogliano, J., Osborne, N., Sharp, M.D., Abanes-De Mello, A., Perez, A., Sun, Y.L., and Pogliano, K. (1999). A vital stain for studying membrane dynamics in bacteria: a novel mechanism controlling septation during *Bacillus subtilis* sporulation. Mol. Microbiol. *31*, 1149–1159.

Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R., et al. (2003). The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature *423*, 81–86.

Rubio, A., and Pogliano, K. (2004). Septal localization of forespore membrane proteins during engulfment in *Bacillus subtilis*. EMBO J. *23*, 1636–1646.

Setlow, B., Magill, N., Febbroriello, P., Nakhimovsky, L., Koppel, D.E., and Setlow, P. (1991). Condensation of the forespore nucleoid early in sporulation of Bacillus species. J. Bacteriol. *173*, 6270–6278.

Spemann, H. (1918). Über die determination der ersten organanalagen der amphibienembryonen I–IV. Arch. F. Entw. Mech. *43*, 448–555.

Steinmetz, M., and Richter, R. (1994). Plasmids designed to alter the antibiotic resistance expressed by insertion mutations in *Bacillus subtilis*, through in vivo recombination. Gene *142*, 79–83.

Sterlini, J.M., and Mandelstam, J. (1969). Commitment to sporulation in *Bacillus subtilis* and its relationship to development of actinomycin resistance. Biochem. J. *113*, 29–37.

Stragier, P. (2002). A gene odyssey: exploring the genomes of endospore-forming bacteria. In Bacillus subtilis and Its Closest Relative: From Genes to Cells, A.L. Sonenshein, J.A. Hoch, and R. Losick, eds. (Washington, DC: ASM Press), pp. 519–526.

Strauch, M., Webb, V., Spiegelman, G., and Hoch, J.A. (1990). The SpoOA protein of *Bacillus subtilis* is a repressor of the *abrB* gene. Proc. Natl. Acad. Sci. USA *87*, 1801–1805.

Sussman, M.D., and Setlow, P. (1991). Cloning, nucleotide sequence, and regulation of the *Bacillus subtilis gpr* gene, which codes for the protease that initiates degradation of small, acid-soluble proteins during spore germination. J. Bacteriol. *173*, 291–300.

Tam, P.P., Kanai-Azuma, M., and Kanai, Y. (2003). Early endoderm development in vertebrates: lineage differentiation and morphogenetic function. Curr. Opin. Genet. Dev. *13*, 393–400.

Wach, A. (1996). PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. Yeast *12*, 259–265.

Cell, Vol. 88, 667–674, March 7, 1997, Copyright ©1997 by Cell Press

Bipolar Localization of the Replication Origin Regions of Chromosomes in Vegetative and Sporulating Cells of B. subtilis

Chris D. Webb,*# Aurelio Teleman,*# Scott Gordon,† Aaron Straight,‡ Andrew Belmont,§ Daniel Chi-Hong Lin,‖ Alan D. Grossman,‖ Andrew Wright,† and Richard Losick*
*Department of Molecular and Cellular Biology
Harvard University
Cambridge, Massachusetts 02138
†Department of Molecular Biology and Microbiology
Tufts University Health Sciences Campus
Boston, Massachusetts 02111
‡Department of Physiology
School of Medicine
University of California
San Francisco, California 94143
§Department of Cell and Structural Biology
University of Illinois, Urbana–Champaign
Urbana, Illinois 61801
‖Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

To investigate chromosome segregation in B. subtilis, we introduced tandem copies of the lactose operon operator into the chromosome near the replication origin or terminus. We then visualized the position of the operator cassettes with green fluorescent protein fused to the LacI repressor. In sporulating bacteria, which undergo asymmetric cell division, origins localized near each pole of the cell whereas termini were restricted to the middle. In growing cells, which undergo binary fission, origins were observed at various positions but preferentially toward the poles early in the cell cycle. In contrast, termini showed little preference for the poles. These results indicate the existence of a mitotic-like apparatus that is responsible for moving the origin regions of newly formed chromosomes toward opposite ends of the cell.

Introduction

The segregation of homologous chromosomes during the mitotic cycle of eukaryotic cells is mediated by a microtubule-based spindle that pulls chromosomes to opposite poles of the cell prior to cytokinesis (Hyman, 1995; Pluta et al., 1995; Barton and Goldstein, 1996). Bacteria, in contrast, lack a conspicuous apparatus for chromosome segregation, and the nature of the machinery for ensuring that newly duplicated chromosomes are faithfully distributed to daughter cells is obscure for prokaryotes (Wake and Errington, 1995). In formulating their replicon model for the regulation of DNA replication, Jacob et al. (1963) called attention to the DNA segregation problem in bacteria, envisioning that the attachment of newly duplicated chromosomal origins to the cell surface and growth between the points of attachment could progressively pull the two chromosomes apart. Lineage studies have demonstrated that DNA segregates in a nonrandom fashion (Eberle and Lark, 1966; Cooper, 1991), and evidence has been obtained for the association of the replication origin region with the cytoplasmic membrane (Bone et al., 1985; Sandler and Keynan, 1988). Nonetheless, zonal growth is not a general feature of cell wall growth, and it is unlikely that this is the basis for the distribution of newly formed chromosomes to daughter cells (Mobley et al., 1984; Woldringh et al., 1987).

Some progress has been made in addressing the chromosome segregation problem in bacteria through the identification of partition (*par*) genes, some of which govern the segregation of plasmid replicons whereas others influence the efficiency of chromosome segregation (Hiraga, 1992). Nonetheless, the problem of chromosome movement has resisted analysis in part because of the small size of bacteria and the absence of compact morphological structures equivalent to the highly condensed chromatids of eukaryotic cells that can be visualized by light microscopy.

To address the issue of chromosome movement in bacteria, we took advantage of a method recently devised by Straight et al. (1996) and Robinett et al. (1996) for visualizing specific chromosomal sites during mitosis in budding yeast and in Chinese Hamster ovary cells. The method is based on the insertion into the chromosome of a cassette consisting of multiple (256) tandem copies of the operator for the E. coli lactose operon (*lacO*) and the use of a fusion of the green fluorescent protein (GFP) from Aequorea victoria to the lactose operon repressor LacI. Binding of GFP-LacI to the tandem operators causes the hybrid repressor molecules to cluster in a small spot that can be visualized by fluorescence microscopy.

Our strategy for visualizing chromosome segregation in bacteria was to insert a copy of the *lacO* cassette either near the replication origin or near the terminus of the chromosome of Bacillus subtilis. This spore-forming bacterium was advantageous for use in these studies because previous genetic experiments had led to the inference that at the start of the process of spore formation, when the sporangium contains two newly formed chromosomes, the replication origins become attached near opposite poles of the cell (Wu and Errington, 1994). We asked whether we could directly visualize polar attachment of the replication origins at the start of sporulation. We also asked whether polar attachment was a normal feature of the cell cycle during binary fission. Here we report that the origins, but not the termini, of the two sporangial chromosomes do indeed localize near the ends of the sporangium. Moreover, we detect a population of cells at an early stage of the cell cycle in which the replication origin is preferentially located near the cell poles, a finding that indicates the existence of a mitotic-like apparatus that is responsible for moving the replication origin regions of newly duplicated chromosomes toward opposite ends of the cell.

# These authors contributed equally to this work.

Results and Discussion

Bipolar Localization of the Replication Origins during Sporulation

We first sought to visualize the replication origin and terminus in cells undergoing sporulation because the results of previous genetic studies had suggested that the region of the chromosome proximal to the origin is sequestered near the cell pole (Wu and Errington, 1994). Sporulating cells undergo a modified form of cell division in which a septum is formed at an extreme polar position to create a small forespore cell and a large mother cell. Unlike the process of binary fission in which DNA segregation precedes cytokinesis, the sporulation septum is formed prior to translocation of a complete chromosome into the forespore. Initially after polar septation, only a small portion of the chromosome is trapped in the forespore; the remainder is subsequently transported across the septum into the small sporangial compartment by a conjugation-like protein called SpoIIIE (Wu et al., 1995). Mutants defective in the *spoIIIE* gene are blocked in DNA translocation and hence their forespore compartment contains only about one-third of the normal complement of DNA (Wu and Errington, 1994; Sharpe and Errington, 1995). That this one-third corresponds to the origin-proximal region of the chromosome is inferred from the following chromosomal position effect on the expression of genes under the control of the forespore-specific transcription factor σ^F: in sporulating cells of a *spoIIIE* mutant, σ^F-controlled genes located proximal to the replication origin are actively transcribed whereas genes located distal to the origin are transcriptionally silent (Sun et al., 1991). This has been interpreted (Wu and Errington, 1994) to indicate that the replication origins are attached near the cell poles at the start of sporulation and that, in the absence of DNA translocation, only origin-proximal DNA is present in the compartment in which σ^F is active.

To observe the origin and the terminus in sporulating cells, we engineered the synthesis of the GFP-LacI fusion protein in sporulating cells using a promoter (*spoVG*) that is active at the onset of sporulation (Zuber and Losick, 1983). The GFP-LacI construct was then introduced into cells that contained the operator cassette located near the replication origin region of the chromosome (at *amyE* [25°] or *spo0J* [359°]) or near the terminus (*cgeD* [181°]).

GFP-LacI-producing bacteria that harbored the *lacO* cassette near the replication origin characteristically exhibited two green fluorescent dots near opposite poles of the cell (Figures 1A and 1E). Due to the low level of the fluorescent signal and a moderate level (30%–40%) of sporulation in our engineered strain, only a minority of the cells exhibited green dots (~11%). In other work involving the use of a cooled charge coupled device (CCD) camera, a higher proportion of sporangia exhibiting fluorescence was observed (~30%). Of those sporangia in which green dots were observed, a substantial proportion (61%, out of a total of 95 cells examined) exhibited a bipolar pattern of fluorescence (in many of the remaining sporangia only a single polarly localized dot was seen). When the *gfp-lacI* fusion was introduced into bacteria lacking the operator cassette, no green dots were observed above a uniform background of low level fluorescence.

Bacteria exhibiting green dots were mostly observed at an early stage of development, prior to the formation of the polar septum (Figure 1B). Fluorescent sporangia were harder to detect at later stages of development, although some postseptation sporangia were observed that exhibited a bipolar pattern of fluorescence (Figure 2B). Yet later, at the stage at which the forespore is wholly engulfed by the mother cell, the origins no longer appeared to be attached near the poles of either compartment (Figure 2C).

In contrast to the results obtained with the operator cassette near the origin, GFP-LacI-producing bacteria harboring the *lacO* cassette near the terminus of replication exhibited one or two green dots that were characteristically located approximately in the middle of the cell (Figure 1C and 1G). For sporangia in which only one dot was evident, the two chromosomes may not have fully completed replication. Alternatively, these sporangia may contain two fully duplicated chromosomes but the termini may be in such close proximity that they could not be resolved by fluorescence microscopy. Once again, fluorescence was principally observed in sporangia at an early stage of development (Figure 1D). Among cells with a signal (11% of the total), a high proportion (74% out of a total of 149 cells) exhibited green dots in the middle region of the cells, and rarely, if ever, near both poles.

In summary, our observations are consistent with the idea depicted in Figure 3A that at the start of sporulation, chromosomes are oriented with their origins near opposite poles and their termini in close proximity to each other near the middle of the cell (i). Next, after polar septation (ii), the origin-distal region of the forespore chromosome is translocated into the small sporangial compartment (iii). Finally, after engulfment of the forespore by the mother cell, the replication origin ceases to be attached near either end of the sporangium (iv).

Polyploidy in a Partition Mutant

Recent genetic experiments have implicated the *spo0J* gene in the anchoring of the replication origin region to the sporangial poles (Sharpe and Errington, 1996). Mutation of *spo0J*, which is related to the family of partition genes that are required for proper plasmid segregation in Gram-negative bacteria (Hiraga, 1992), causes a defect in chromosome segregation during growth and a block at the onset of sporulation (Ireton et al., 1994). This sporulation block can be overcome by a second mutation in *soj* (for *suppressor of spo0J*), which is also related to partition genes. Thus, a *soj spo0J* double mutant is capable of sporulating (Ireton et al., 1994). Interestingly, however, the strict chromosome position effect observed in a *spoIIIE* DNA translocase mutant is not observed in the presence of additional mutations in *soj* and *spo0J*. Thus, in a *soj spo0J spoIIIE* triple mutant, a low level of expression of σ^F-controlled genes located distal to the replication origin can be observed (Sharpe and Errington, 1996). These observations have been interpreted to indicate that *spo0J* is involved in polar attachment of the replication origins and that, in its absence, the chromosomes are not oriented and random

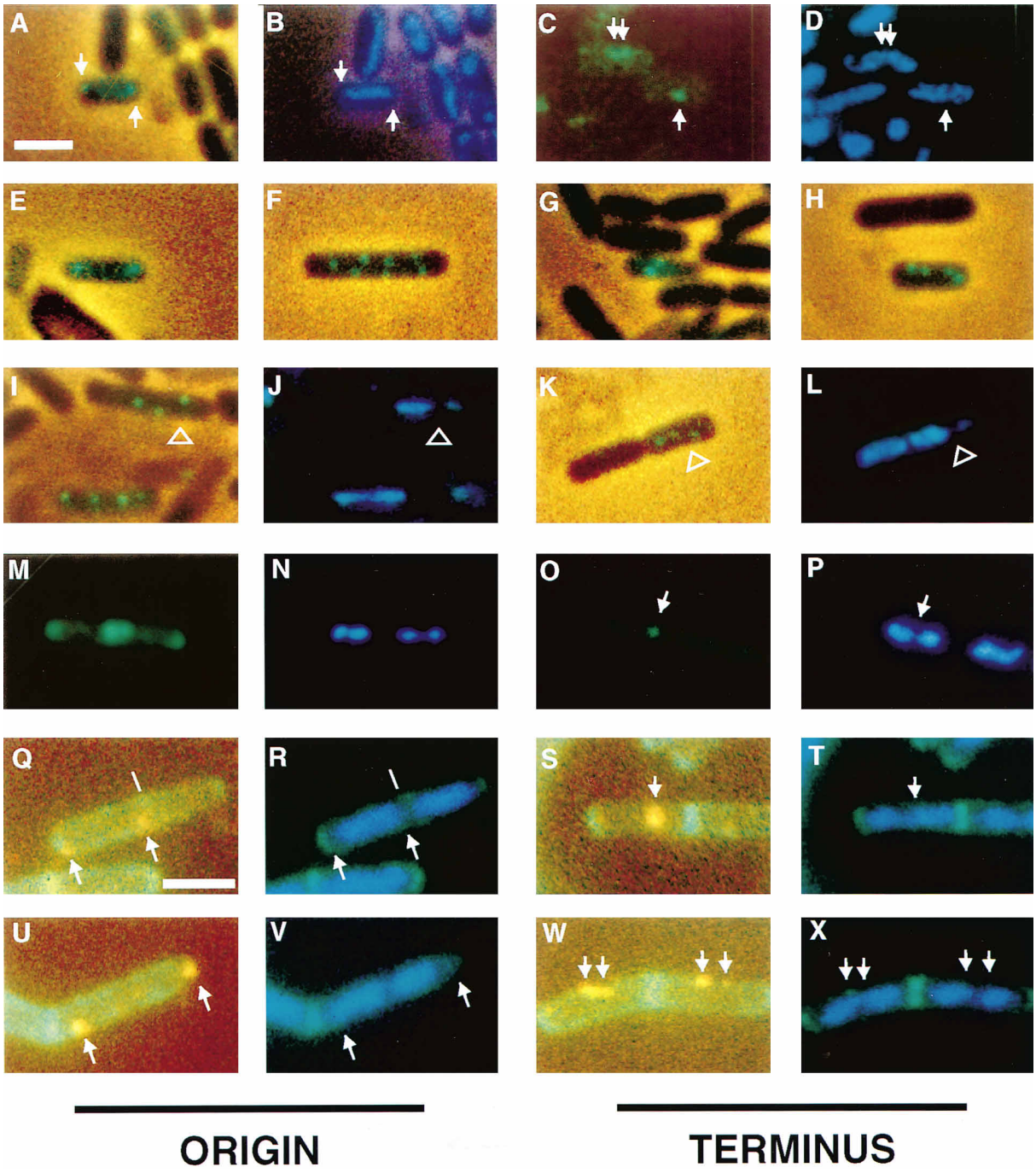


Figure 1. Fluorescence Micrographs of Sporulating and Vegetative Cells with the Operator Cassette Near the Origin or the Terminus

The operator cassette is near the origin (*amyE*) in (A), (B), (E), (F), (I), (J), (M), (N), (Q), (R), (U), and (V) and near the terminus (*cgeD*) in (C), (D), (G), (H), (K), (L), (O), (P), (S), (T), (W), and (X). Arrows indicate the location of dots from GFP-LacI. (A, B, and E) Sporangia (strain AT16) exhibiting bipolar dots from the origin region (green in [A] and [E]) and a predivisional pattern of unsegregated chromosomal DNA (blue in [B], corresponding to sporangium in [A]). (C, D and G) Sporangia (strain AT18) with one or two dots from the terminus region and a predivisional pattern of chromosomal DNA ([C] and [D] show the same field of cells). (F, H–L) Multiple dots in *soj spo0J* mutant sporangia, with the operator cassette near the origin (F, I–J; strain AT19) or near the terminus (H, K–L; strain AT20). Note the gap between the forespore and mother cell chromosome as indicated by open arrowheads in (J) and (L). (M) Bipolar pattern of dots from the origin region in a pair of vegetative cells (strain CW437), with separated nucleoids in each (N). (O) Single dot from the terminus region in the middle of a vegetative cell (strain CW438), with nucleoids in the process of separating, seen in (P). (Q–X) Immunofluorescence of vegetative cells stained with anti-GFP antibodies (orange), with lectin to stain the cell wall (green), and with DAPI to visualize DNA (blue). (Q and R) Bipolar pattern of dots from the origin region (strain CW437) seen early in the cell cycle, prior to chromosome segregation. Line indicates location of a lectin-stained septum, not clearly seen in photograph. (U and V) Bipolar pattern of origin region dots (strain CW437) later in the cell cycle when the chromosomes have separated into two nucleoids. (S and T) Single dot from the terminus in a cell (strain CW438) in which the nucleoids have not yet fully separated. (W and X) Two dots from the terminus (strain CW438) seen in adjacent cells in which the nucleoids have begun to separate (left-hand cell) or have fully separated (right-hand cell). The scale bar in (A), which represents 2 μm, pertains to (A)–(P), whereas the scale bar in (Q), which also represents 2 μm, pertains to (Q)–(X).

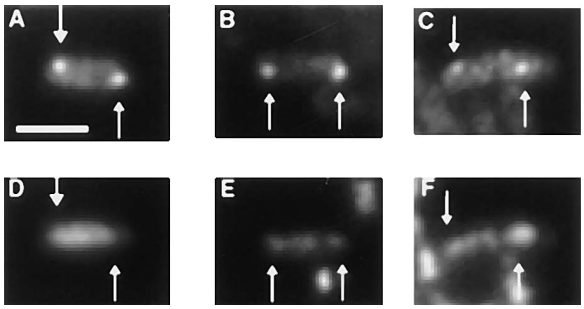


Figure 2. Fluorescence Micrographs of Sporangia with the Operator Cassette Inserted Near the Origin

Strain AT25 (with the operator cassette at 359°) is shown. The upper panels show fluorescence from the GFP-LacI fusion in a predivisional sporangium (A), in a postseptation sporangium (B) and in a postengulfment sporangium (C). The corresponding lower panels (D–F) show fluorescence from DNA stained with DAPI for each of the same sporangia. Arrows indicate the locations of dots from GFP-LacI.

regions of the chromosome become trapped in the forespore when the polar septum is formed. Indeed, consistent with the idea that Spo0J could be involved in sequestering the replication origin near the poles, other work (D. C.-H. L., P. Levin, A. G., unpublished data) has shown that Spo0J exhibits a bipolar distribution during sporulation (as well as during growth; see below).

To investigate the role of these partition genes in the polar attachment of the origin region during sporulation, we introduced the *gfp-lacI* construct and the *lac* operator cassette into a *soj spo0J* double mutant. Strikingly, multiple dots were observed when the operator cassette was located either near the origin or near the terminus (Figures 1F, 1H, 1I, and 1K). Among sporangia exhibiting a fluorescent signal, 53% (out of a total of 156 cells) had more than two dots. Moreover, sporangia were observed with up to eight dots when the cassette was located near the origin and up to four dots when the cassette was at the terminus. Because multiple origins and multiple termini were observed, some of the *soj spo0J* mutant sporangia must contain more than two complete copies of the chromosome. This is also evident from the size of the mutant cells: those cells with multiple dots were significantly larger than other cells (for example, compare the sporangium in Figure 1F with that in Figure 1E).

Most of the mutant sporangia exhibiting a fluorescent signal were at the predivisional stage of sporulation, but some sporangia with multiple chromosomes were observed that had appeared to reach the stage of polar division. These could be recognized by means of DAPI staining, which revealed a separated region of DNA, indicative of the presence of a forespore (Setlow et al., 1991; Resnekov et al., 1996) (Figures 1J and 1L). The existence of postseptation sporangia with multiple chromosomes suggests a possible alternative explanation for the absence of a strict chromosomal position effect in a *soj spo0J spoIIIE* triple mutant: polar septation in a sporangium containing multiple chromosomes could result in the trapping of random chromosomal regions within the newly formed forespore due to the presence of extra copies of the chromosome in the sporangium

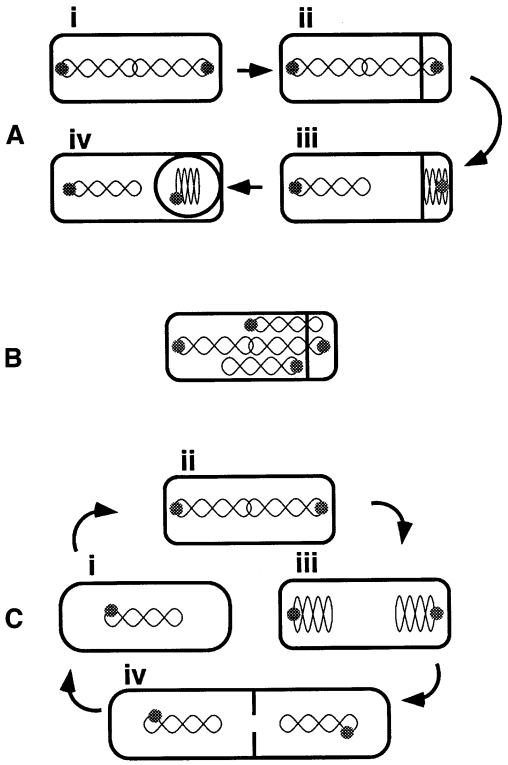


Figure 3. Model Depicting the Position of the Origin during Asymmetric and Symmetric Cell Division

Dots indicate the location of the origin on the chromosome.

(A) Movement of the origin during sporulation. (i) Prior to septation the two chromosomes appear as a single elongated filament and the origins are localized near opposite poles. (ii) Initially after polar septation, the origin proximal region of the forespore chromosome is trapped in the forespore. (iii) After DNA translocation, an entire chromosome is packed into the forespore. (iv) After engulfment of the forespore by the mother cell, the origins are detached from the poles.

(B) Multiple chromosomes in a *soj spo0J* double mutant sporangium. The presence of multiple chromosomes could allow the origin distal regions of an extra chromosome to become trapped in the forespore.

(C) Movement of the origin during binary fission. (i) Early in the cell cycle a single unreplicated chromosome is spread across the cells as an elongated filament. (ii) Next, following replication, the origins localize to opposite poles. (iii) Later, the chromosomes condense, creating a gap in which the septum can form. (iv) Cytokinesis.

(Figure 3B). According to this view, Spo0J need not be responsible for anchoring the origin region to the poles. Instead, it may serve to coordinate cell division with chromosome copy number in cells entering sporulation so that normally only two chromosomes are present in cells undergoing polar division. Alternatively, Spo0J could be involved both in chromosome anchoring as well as in coordinating chromosome copy number with cell division.

Polar Localization of the Origin during Vegetative Growth

Next, we asked whether polar attachment of the replication origin region was unique to cells entering sporulation or a general feature of chromosome segregation

during asymmetric and symmetric cell division. To visualize the origin and terminus in growing cells, we engineered the expression of *gfp-lacI* in vegetative cells by placing the gene (*spo0H*) for σ^H under the control of the xylose-inducible promoter (Dubnau et al., 1988; Gartner et al., 1988). Thus, in such cells, xylose would enhance the levels of σ^H during growth and, hence, stimulate transcription of *gfp-lacI* from the σ^H-controlled *spoVG* promoter. Nevertheless, only a low level of green fluorescence was observed in such engineered cells.

To increase the sensitivity at which GFP-LacI could be detected in growing cells, we turned to immunofluorescence using antibodies directed against GFP, an effective strategy that enabled us to observe signals for the GFP-LacI fusion in over 40% of the cells. In the following experiments, GFP-LacI appeared as orange dots whereas the cell wall and septa were stained green by the use of a lectin. Finally, nucleoids were visualized by the use of DAPI. The use of DAPI enabled us to distinguish cells at early stages of the cell cycle in which the nucleoid appears as a single filament along the length of the cell (Figures 1R and 1T) from cells at late stages of the cycle when the chromosomes have segregated into two distinct condensed nucleoids (Figures 1V and 1X).

First, we consider cells in exponential growth phase harboring the operator cassette near the origin. Among cells for which a signal could be seen, most (75%, out of a total of 140 cells) exhibited only a single dot, which was often located near a cell pole. The significance of such one-dot cells was difficult to interpret. They could represent cells in which the origin had not replicated or in which newly formed origins had not yet begun to segregate from each other. Alternatively, such cells could have contained two well-separated origins but, because of the low level of signal detection during growth, a second origin located out of the plane of focus might not have been detected. More informative was the case of cells exhibiting two dots (25%, out of a total of 140 cells). Among such cases, cells were often observed to have a dot close to each end. This bipolar pattern of dots was most prevalent at early stages of the cell cycle, when the chromosomes had not yet segregated (Figures 1Q and 1R), but could also be observed at later stages when the chromosomes had separated into two distinct nucleoids (Figures 1M, 1N, 1U, and 1V). A similar pattern of bipolar localization was observed in experiments designed to visualize Spo0J in growing cells, reinforcing the view that Spo0J binds to the replication origin region and providing independent evidence that chromosomes become oriented with their origins toward the poles (D. C.-H. L. et al., unpublished data). Recently, the partition homologs ParA and ParB of the gram-negative bacterium Caulobacter crescentus have also been found to exhibit a bipolar pattern of localization, an indication that the orientation of the chromosomes with the origin toward the poles of the cell could be a general feature of chromosome segregation in prokaryotes (Mohl and Gober, 1997).

Next, we consider growing cells harboring the operator cassette near the terminus. Such cells generally displayed one or two dots located approximately in the middle of the cell (Figures 1O, 1S, and 1W). Although cells exhibiting two dots were infrequent, a bipolar localization pattern of termini tagged with the operator cassette was rarely if ever observed. Thus, our results indicate the existence of a population of growing cells in which newly formed origins, but not termini, are sequestered at or near opposite ends of the cell.

The above experiments were carried out with cells that were grown in a rich medium, conditions of rapid growth in which multiple replication forks are to be expected. When the generation time is shorter than the time required for duplication of a chromosome, a second round of replication is initiated before the first round of replication is complete (Cooper, 1991). Despite this expectation, we were not able to observe more than two dots in cells in which the origin was tagged with the *lacO* cassette. Thus, if multiple replication forks were present in rich medium, we were not able to detect them under our conditions. However, using a newly constructed strain (see below) that exhibits brighter fluorescence, we do observe multiple dots in rapidly growing cells. In light of this issue, we also carried out experiments with cells grown in a minimal medium (with glucose as the sole carbon source) in which no more than two replication forks were expected to be present. In minimal-medium grown cells with the *lacO* cassette at the origin, we detected cells with two dots (data not shown). Among such cells were those in which the dots were located in a bipolar fashion, similar to that observed for cells grown in rich medium.

Finally, to improve the efficiency with which operator-bound GFP-LacI could be detected in growing cells, we created an additional fusion in which *gfp-lacI* was joined to a strong vegetative promoter (*veg*; Moran et al., 1982). Minimal-medium grown cells having the P_{veg}-*gfp-lacI* fusion exhibited much brighter signals than in the experiments described above (data not shown). Moreover, a high proportion ($>$90%) of the cells exhibited detectable fluorescence. As before, among cells harboring the P_{veg}-*gfp-lacI* fusion, a subpopulation could readily be observed that exhibited a bipolar pattern of dots when the operator cassette was at the origin but not when it was at the terminus (data not shown).

A Model for Chromosome Segregation

Our results suggest the following model for chromosome segregation during binary fission. As depicted in Figure 3Ci, the newborn cell initially contains a single chromosome and hence a single replication origin region. Next, after replication commences, the two newly formed origins are drawn toward opposite ends of the cell, eventually resulting in a cell (Figure 3Cii) with two complete chromosomes, each anchored near a cell pole from a site (centromere) located in the vicinity of its replication origin. At this stage, the two chromosomes are spread across the length of the cell in the form of a single filament, similar to that observed early in sporulation. Cells (Figure 3Cii) with a bipolar arrangement of replication origin regions could arise in one of two ways. One possibility is that before the initiation of replication, the origin region is not attached near either pole. Only after replication commences do the origins migrate toward the poles. Alternatively, the origin region

in a newborn cell could be attached near a pole prior to replication. Then after replication ensues, one of the two newly formed origin regions could be transported toward the opposite end of the cell. Our data do not allow us to distinguish between these possibilities decisively.

In the next stage of the cell cycle (Figure 3Ciii), the two chromosomes separate from each other, moving to opposite ends of the cells. A simple way in which this could occur would be by condensation of the chromosomes, each anchored near an opposite end of the cell. This condensation could be mediated by a protein similar to members of the SMC family of eukaryotic chromosome condensation proteins, homologs of which are known to exist in bacteria (Hirano et al., 1995). Finally, after each chromosome has coalesced into a distinct nucleoid, cytokinesis (Figure 3Civ) occurs by the formation of a septum in the gap between the separated chromosomes. If these ideas are correct, then a principal challenge for future investigations will be to define the nature of the centromere and the mitotic-like apparatus that effects centromere movement and chromosome segregation.

Experimental Procedures

Strain Constructions

To express the *gfp-lacI* fusion of Straight et al. (1996) in B. subtilis, the fusion was placed under the control of the *spoVG* promoter by being joined in-frame at the third codon of the *spoVG* open reading frame (ORF). DNA from the *spoVG* promoter region was obtained by PCR amplification using oligonucleotides SDMOL1 (5′-GCTGGC GAAAGGGGGATGTG-3′), which corresponds to a vector sequence near the site of insertion of a *spoVG*-containing HindIII fragment in the HindIII site of pBSKS+, and ATO3 (5′-TCTAGGATCCCCATCGAT GTAACTTCCACAGTAGTTCACC-3′), which created adjacent ClaI and BamHI restriction sites after the third codon of the *spoVG* ORF. The PCR product was digested with EcoRI and BamHI and ligated to similarly digested pDG795 to create pAT10. Plasmid pDG795 is an MLS-resistance gene-containing vector for creating insertions at the ThrC locus on the B. subtilis chromosome (a gift of P. Stragier).

Next, the *gfp-lacI* gene fusion was amplified using pAFS78 (Straight et al., 1996) with oligonucleotides ATO1 (5′-AAAAAAGATC TGATTAAGTTGGGTAA-3′), which created a BglII site downstream of the open reading frame, and ATO4 (5′-ATAGCATCGATGAGTAAA GGAGA-3′), which created a ClaI restriction site 5′ of the ATG start codon. The resulting PCR product, digested with BglII and ClaI, was ligated to BamHI-ClaI-digested pAT10, yielding pAT11. This created an extended ORF in which three codons of *spoVG* were joined via a two codon linker to the *gfp-lacI* gene fusion. The plasmid was linearized by digestion with XhoI and used to transform wild-type, prototrophic B. subtilis strain PY79 (Youngman et al., 1984), followed by selection for MLS resistance (5 μg/ml). This yielded strain AT15, which was confirmed to be auxotrophic for threonine.

To insert the *lacO*-containing plasmid pAFS52 (Straight et al., 1996), into the chromosome near the origin (25°), we introduced into the chromosome at *amyE* a segment of DNA corresponding to the yeast *trp1* gene that was present in pAFS52. This 1.4 kb segment of pAFS52 was introduced into the *amyE* locus using the *amyE*-integrating vector pDG364 to create strain AT10. Next, a kanamycin resistance gene contained on a 1.9 kb BamHI-SalI fragment from pER82 (Driks et al., 1994) was introduced into pAFS52 to create pAT7. Strain AT10 was transformed with pAT7 followed by selection for kanamycin resistance (5 μg/ml). The resulting strain was then transformed with chromosomal DNA from AT15 followed by selection for MLS resistance (5 μg/ml), yielding AT16 which had both the *lacO* cassette at *amyE* (25°) and the *spoVG-gfp-lacI* fusion. To generate a *soj spo0J* mutant derivative of AT16, the strain was transformed with chromosomal DNA from AG1505 (Ireton et al., 1994) followed by selection for spectinomycin resistance (5 μg/ml), which yielded strain AT19.

To select for introduction of the *lacO* cassette in other regions of the chromosome of AT15, a chloramphenicol resistance gene (*cat*) contained on a 1.3 kb HindIII-SalI fragment from pDG364 (Cutting and Vander Horn, 1990) was introduced into a HindIII-SalI-digested pAFS52. This created pAT12, which was used to insert the *lacO* cassette at the sites near the origin (359°) and near the terminus (181°).

A 1 kb HindIII fragment from pIK217 corresponding to the *soj-spo0J* region of the chromosome (359°) (Ireton et al., 1994) was cloned into the HindIII site of pAT12, yielding pAT15. Finally, pAT15 was used to transform AT15 followed by selection for chloramphenicol resistance (5 μg/ml), generating strain, AT25.

To place the operator cassette near the terminus (181°), a 1.2 kb AatII-BglII fragment from pSR129 corresponding to the *cgeD* gene (Roels and Losick, 1995) was ligated to pAT12 that had been digested with AatII and BamHI, resulting in plasmid pAT14. This plasmid was used to transform AT15 followed by selection for chloramphenicol resistance (5 μg/ml), generating strain AT18. To generate a *soj spo0J* mutant derivative of AT18, the strain was transformed with chromosomal DNA from AG1505 (Ireton et al., 1994) followed by selection for spectinomycin resistance (100 μg/ml), which yielded strain AT20.

To express the *spoVG-gfp-lacI* fusion at higher levels during vegetative growth, an inducible *spo0H* fusion was made. A 550 bp SmaI-EcoRI fragment from pSG20H (Fort and Errington, 1985) containing part of *spo0H*, with ends rendered flush with DNA Polymerase I Klenow fragment, was cloned into a SalI-digested pDAG8–1, yielding pCW75. Plasmid pDAG8-1 has a xylose-inducible promoter-containing HindIII fragment from pDG1832 (a gift of P. Stragier) in the HindIII site of pJL74 (LeDeaux and Grossman, 1995). Plasmid pCW75 was inserted into the chromosome by a Campbell-like recombination at *spo0H* by transforming PY79 followed by selection for spectinomycin resistance (100 μg/ml), yielding CW436. Strains AT16, AT18, and AT25 were transformed with the chromosomal DNA of CW436 followed by selection for spectinomycin resistance (100 μg/ml) resulting in strains CW437, CW438, and AT29, respectively.

Growth Conditions

For vegetative growth, cells were grown overnight in Luria-Bertani (LB) medium at 25°C and then diluted 1:100 in either LB or S7 defined minimal medium (Jaacks et al., 1989) grown at 30°C to an OD_{600} of ~0.6. For sporulation, cells were grown either in Difco sporulation medium (DSM) overnight at 25°C or on DSM plates overnight at 30°C (Nicholson and Setlow, 1990). The signal was brighter when cells were grown on plates. The number of copies of the *lacO* cassette in strains AT18 and AT29 was increased by growing the strains on higher concentrations of chloramphenicol, which also resulted in a brighter signal.

Microscopy and Photography

Immunofluorescence was carried out as previously described (Harry et al., 1995; Pogliano et al., 1995), with the following modifications: the fixation used contained 0.06% glutaraldehyde (v/v) in 16% paraformaldehyde (v/v), and the cells were treated with lysozyme for 3 min. The cells were not treated with methanol or acetone. The polyclonal rabbit anti-GFP antibodies (Clonetech) were used at 1:4000 dilution, and the incubation time was 1 hr at room temperature. Anti-rabbit Cy3 secondary antibodies were used at a 1:200 dilution and incubated for 1 hr at room temperature, with wheat germ agglutinin-fluorescein (10 ng/ml) (Molecular probes) and 4,6-Diamidino-2-phenyloindole (DAPI) (2 μg/ml) (Sigma). Photomicrographs were taken as previously described (Harry et al., 1995; Pogliano et al., 1995), with the exception of the use of Fujichrome Provia400 film. Images were processed as previously described (Resnekov et al., 1996). For scoring, photomicrographs were taken using a Photometrics SenSys1400 Cooled CCD camera and a PowerMac9500 with IPLab Spectrum image processing software (version 2.5.6).

For observation of GFP signals, cells were resuspended in phosphate-buffered saline (pH 7.4) or water and placed on microscope slides. For better DNA staining, cells were lightly fixed as described for immunofluorescence above. The fix did not appear to affect the GFP signal, as similar results were observed without fixation. DAPI

was added to a final concentration of 2 μg/ml. Color photomicrographs were taken as previously described (Webb et al., 1995), with the exception of the use of Kodak Ektachrome 400X. Black and white photomicrographs were taken with the Photometrics SenSys1400 camera and a PowerMac9500 with IPLab Spectrum image processing software (version 2.5.6). Images were processed as previously described (Resnekov et al., 1996).

Acknowledgments

We thank S. Meyer for plasmid pSM79, D. Garsin for plasmid pDAG8-1, P. Stragier for plasmids pDG795 and pDG1832, and O. Resnekov for invaluable discussions and help with immunofluorescence experiments. C. D. W. is a predoctoral fellow of the National Science Foundation. This work was supported by NIH grants GM38035 to A. W. and GM18568 to R. L.

Received December 9, 1996; revised January 22, 1997.

References

Barton, N.R., and Goldstein, L.S. (1996). Going mobile: microtubule motors and chromosome segregation. Proc. Natl. Acad. Sci. *93*, 1735–1742.

Bone, E.J., Todd, J.A., Ellar, D.J., Sargent, M.G., and Wyke, A.W. (1985). Membrane particles from *Escherichia coli* and *Bacillus subtilis*, containing penicillin-binding proteins and enriched for chromosomal-origin DNA. J. Bacteriol. *164*, 192–200.

Cooper, S. (1991). The segregation of DNA and the cell surface. In Bacterial Growth and Division (San Diego, California: Academic Press), pp. 279–312.

Cutting, S.M., and Vander Horn, P.B. (1990). Genetic analysis. In Molecular Biological Methods for *Bacillus*, C.R. Harwood and S.M. Cutting, eds. (New York: John Wiley & Sons), pp. 27–74.

Driks, A., Roels, S., Beall, B., Jr., Moran, C.P., and Losick, R. (1994). Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. Genes Dev. *8*, 234–244.

Dubnau, E., Weir, J., Nair, G., III, Carter, H.L., Moran, C.P., and Smith, I. (1988). *Bacillus* sporulation gene *spoOH* codes for σ^{30} (σ^{H}). J. Bacteriol. *170*, 1054–1062.

Eberle, H., and Lark, K.G. (1966). Chromosome segregation in *Bacillus subtilis*. J. Mol. Biol. *22*, 183–186.

Fort, P., and Errington, J. (1985). Nucleotide sequence and complementation analysis of a polycistronic sporulation operon, *spoVA*, in *Bacillus subtilis*. J. Gen. Microbiol. *131*, 1091–1105.

Gartner, D., Geissendorfer, M., and Hillen, W. (1988). Expression of the *Bacillus subtilis xyl* operon is repressed at the level of transcription and is induced by xylose. J. Bacteriol. *170*, 3102–3109.

Harry, E., Pogliano, K., and Losick, R. (1995). Use of immunofluorescence to visualize cell-specific gene expression during sporulation in *Bacillus subtilis*. J. Bacteriol. *177*, 3386–3393.

Hiraga, S. (1992). Chromosome and plasmid partition in *Escherichia coli*. Annu. Rev. Biochem. *61*, 283–306.

Hirano, T., Mitchison, T.J., and Swedlow, J.R. (1995). The SMC family: from chromosome condensation to dosage compensation. Curr. Opin. Cell Biol. *7*, 329–336.

Hyman, A.A. (1995). Microtubule dynamics. Kinetochores get a grip. Curr. Biol. *5*, 483–484.

Ireton, K., Gunther, N.W., and Grossman, A.D. (1994). spo0J is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. J. Bacteriol. *176*, 5320–5329.

Jaacks, K.J., Healy, J., Losick, R., and Grossman, A.D. (1989). Identification and characterization of genes controlled by the sporulation regulatory gene *spoOH* in *Bacillus subtilis*. J. Bacteriol. *171*, 4121–4129.

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the regulation of DNA replication in bacteria. Cold Spring Harbor Symp. Quant. Biol. *28*, 329–347.

LeDeaux, J.R., and Grossman, A.D. (1995). Isolation and characterization of *kinC*, a gene that encodes a sensor kinase homologous to the sporulation sensor kinases KinA and KinB of *Bacillus subtilis*. J. Bacteriol. *177*, 166–175.

Mobley, H.L.T., Koch, A.L., Doyle, R.J., and Streips, U.N. (1984). Insertion and fate of the cell wall in *Bacillus subtilis*. J. Bacteriol. *158*, 169–179.

Mohl, D.A., and Gober, J.W. (1997). Cell cycle-dependent polar localization of chromosome partitioning proteins in Caulobacter crescentus. Cell, this issue.

Moran, C.P., Lang, N., LeGrice, S.F., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J., and Losick, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in Bacillus subtilis. Mol. Gen. Genet. *186*, 339–346.

Nicholson, W.L., and Setlow, P. (1990). Sporulation, germination and outgrowth. In Molecular Biological Methods for *Bacillus*, C.R. Harwood and S.M. Cutting, eds. (New York: John Wiley & Sons), pp. 391–450.

Pluta, A.F., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G., and Earnshaw, W.C. (1995). The centromere: hub of chromosomal activities. Science *270*, 1591–1594.

Pogliano, K., Harry, E., and Losick, R. (1995). Visualization of the subcellular location of sporulation proteins in *Bacillus subtilis* using immunofluorescence microscopy. Mol. Microbiol. *18*, 459–470.

Resnekov, O., Alper, S., and Losick, R. (1996). Subcellular localization of proteins governing the proteolytic activation of a developmental transcription factor in *Bacillus subtilis*. Genes Cells *1*, 529–542.

Robinett, C.C., Straight, A.F., Li, G., Willhelm, C., Sudlow, G., Murray, A., and Belmont, A.S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using *lac* operator/repressor recognition. J. Cell Biol. *135*, 1685–1700.

Roels, S., and Losick, R. (1995). Adjacent and divergently oriented operons under the control of the sporulation regulatory protein GerE in *Bacillus subtilis*. J. Bacterol. *177*, 6263–6275.

Sandler, N., and Keynan, A. (1988). Membrane binding and release of *Bacillus subtilis* DNA as a function of the cell cycle. J. Gen. Microbiol. *134*, 1155–1163.

Setlow, B., Magill, N., Febbroriello, P., Nakhimovsky, L., Koppel, D.E., and Setlow, P. (1991). Condensation of the forespore nucleoid early in sporulation of *Bacillus* species. J. Bacteriol. *173*, 6270–6278.

Sharpe, M.E., and Errington, J. (1995). Postseptational chromosome partitioning in bacteria. Proc. Natl. Acad. Sci. USA *92*, 8630–8634.

Sharpe, M.E., and Errington, J. (1996). The Bacillus subtilis *soj-spo0J* locus is required for a centromere-like function involved in prespore chromosome partitioning. Mol. Microbiol. *21*, 501–509.

Straight, A.F., Belmont, A.S., Robinett, C.C., and Murray, A.W. (1996). GFP tagging of budding yeast chromosomes reveals that protein–protein interactions can mediate sister chromatid cohesion. Curr. Biol. *6*, 1599–1608.

Sun, D., Fajardo-Cavazos, P., Sussman, M.D., Tovar-Roja, F., Cabrera-Martinez, R.-M., and Setlow, P. (1991). Effect of chromosome location of *Bacillus subtilis* forespore genes on their *spo* gene dependence and transcription by Eσ^{F}: identification of features of good Eσ^{F}-dependent promoters. J. Bacteriol. *173*, 7867–7874.

Wake, R.G., and Errington, J. (1995). Chromosome partitioning in bacteria. Annu. Rev. Genetics *29*, 41–67.

Webb, C.D., Decatur, A., Teleman, A., and Losick, R. (1995). Use of green fluorescent protein for visualization of cell-specific gene expression and subcellular protein localization during sporulation in *Bacillus subtilis*. J. Bacteriol. *177*, 5906–5911.

Woldringh, C.L., Huls, P., Pas, E., Brakenhoff, G.J., and Nanninga, N. (1987). Topography of peptidoglycan synthesis during elongation and polar cap formation in a cell division mutant of *Escherichia coli* MC4100. J. Gen. Microbiol. *133*, 575–586.

Wu, L.J., and Errington, J. (1994). *Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division. Science *264*, 572–575.

Wu, L.J., Lewis, P.J., Allmansberger, R., Hauser, P.M., and Errington, J. (1995). A conjugation-like mechanism for prespore chromosome partitioning during sporulation in *Bacillus subtlis*. Genes Dev. *9*, 1316–1326.

Youngman, P., Perkins, J.B., Losick, R. (1984). Construction of a cloning site near one end of TN*917* into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. Plasmid *12*, 1–9.

Zuber, P., and Losick, R. (1983). Use of a *lacZ* fusion to study the role of the *spoO* genes of *Bacillus subtilis* in developmental regulation. Cell *35*, 275–283.

JOURNAL OF BACTERIOLOGY, May 1982, p. 686–691
0021-9193/82/050686-06$02.00/0

Vol. 150, No. 2

Identification of a New Gene (*secA*) and Gene Product Involved in the Secretion of Envelope Proteins in *Escherichia coli*

DONALD B. OLIVER* AND JON BECKWITH

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115

Received 8 September 1981/Accepted 10 December 1981

We have constructed lambda specialized transducing phages which carry an *Escherichia coli* gene (*secA*) involved in the secretion of certain envelope proteins. These phage have been used to show that *secA* is a new gene to the clockwise side of *envA*. The *secA* mutation previously described, *secA51*(Ts) (D. B. Oliver and J. Beckwith, Cell **25**:765–772, 1981), is recessive to the wild-type allele. We have also isolated Tn*5* insertions in the gene carried on the transducing phage to further define the gene. These phage were used to infect UV-irradiated cells to allow the identification of the *secA* gene product as a 92-kilodalton polypeptide and to show that transcription of *secA* is clockwise relative to the *E. coli* genetic map.

We have described a conditional lethal (temperature-sensitive) *Escherichia coli* mutant which is pleiotropically defective in the secretion of certain envelope proteins (11). This mutation maps in a new gene which lies at approximately 2.5 min on the bacterial chromosome and which we term *secA*. At high temperatures (37 or 42°C) the *secA* mutant strain accumulates cytoplasmic precursors of the maltose binding protein and a number of other envelope proteins. However, certain proteins still seem to be secreted normally. Sixteen additional temperature-sensitive mutants of this class map in this same gene (D. Oliver, M. Quinlan, and J. Beckwith, unpublished data). We have suggested that the *secA* mutations alter a component of the cell's secretion machinery so as to block normal protein export. If this explanation is correct, the characterization of the *secA* gene product and its function should contribute to an understanding of the secretion process.

In this paper we report a detailed analysis of the *secA* gene, including its precise location and orientation on the *E. coli* chromosome. We further show that the previously described *secA* mutation *secA51*(Ts) (11) is recessive to the wild-type allele. Last, we have used translational mapping to identify the *secA* gene product as a 92-kilodalton (92K) polypeptide.

MATERIALS AND METHODS

Media. Media for growth and plating of bacteria have been described (4, 9).

Phage strains. λ16-2 and λ16-25 carrying the *ftsA-envA* region were obtained from J. Lutkenhaus (6). The *Eco*RI cloning vector λ616 (10) was obtained from H. Revel.

Bacterial strains. Genetic nomenclature is from Bachmann and Low (1). *E. coli* K-12 MC4100 (F⁻ *araD* Δ*lacU169 relA rpsL*) was used for growth and plating of the λ phage. MM52 [MC4100, *secA51*(Ts)] was used as the standard *secA* mutant (11). AB2463/KLF4 (*thi-1 thr-1 leu-6 argE3 his-4 proA2 recA13 mtl-1 xyl-5 ara-14 galK2 lacY1 str-31 tsx-33 supE44*/F'104) was obtained from A. Wright. This strain, MM59 [MC4100, *secA51*(Ts) *recA1 srl*::Tn*10* φ (*malE-lacZ*)*72-47*(Hyb)], and MM60 (MC4100, *leu*::Tn*10*) were used in the diploid analysis. MM61 (MC4100, *leu*::Tn*10 ftsA12*), MM62 (MC4100, *leu*::Tn*10 ftsZ84*), and MM63 (MC4100, *leu*::Tn*10 envA*) were used in the complementation experiments. *E. coli* K-12 I7023 [F⁻ Δ*lacU169* Δ(*gal-bio*) *rpsL*] lacking the primary λ attachment site was used to obtain lysogens of λ16-25 integrated into the *ftsA-envA* region. I10.00 [F⁻ *araD* Δ(*lac-pro*) *metB argE*(Am) Nal^r *Rif^r su6* (P2)] was used to select for λ *spi*⁻ phage (12). Strain 294 (*endoI*⁻, *hsdR*⁻ *hsdM*⁺ *su*II) was used for transfection. MM64 (MC4100, *leu*::Tn*5*) was used to obtain TN5 insertions into the *secA* transducing phage. The bacterial strain used to examine phage-directed protein synthesis after UV irradiation was *E. coli* 159 (*uvrA gal rpsL*) and was obtained from J. Lutkenhaus.

Diploid analysis. The KLF4 episome was transferred from AB2463/KLF4 to MM59 by selecting for growth on arabinose minimal agar containing 10 μg of tetracycline per ml at 30°C. Several independent exconjugates were purified and tested for temperature sensitivity of growth at 42°C. None were temperature sensitive. These were then tested to ensure that the *secA51*(Ts) allele was still present by growing P1 *vir* on them, transducing MM60 to *leu*$^+$, and testing for the introduction of the *secA51*(Ts) allele.

Isolation of λDO2. Strain I7023 was lysogenized with λ16-25, and the lysogen was induced with UV irradia-

tion. *spi*⁻ phage were obtained by plating 0.1 ml of the lysate on I10.00. The 20 plaques which appeared were pooled, and a plate lysate was made with this same host. This plate lysate showed the presence of *secA* transducing phage as judged by spot tests on an MM52 lawn at 42°C. Therefore, a culture of MM52 was mixed with this lysate (multiplicity of infection = 0.2) and plated at 42°C for temperature-resistant transductants, which appeared at a frequency of approximately 10^{-3}. Three of these transductants were picked, purified, and induced for phage by UV irradiation. All three isolates yielded high-frequency transducing lysates for *secA*. One phage isolate, λDO2, was picked for further study.

Isolation of Tn5 insertions into *secA*. The procedure for obtaining TN5 insertions into λDO20 was essentially as described by Berg et al. (2). This Tn5 mutagenized phage stock was used to transduce MM52 to kanamycin resistance. Twelve of 600 Kan^r transductants tested were temperature sensitive for growth, indicating inactivation of the *secA* gene carried by the phage. Transducing phage from seven of these were chosen for further study and are designated λDO20(*secA1*::Tn5) through λDO20 (*secA7*::Tn5).

Preparation of phage and phage DNA. Phage were grown and purified and phage DNA was extracted essentially as described by Murray et al. (10).

Restriction and ligation conditions. The restriction and ligation conditions for phage DNA were similar to those described by Murray et al. (10).

Transfection. Strain 294 was treated with 50 mM $CaCl_2$ and transfected with 100 to 500 ng of ligated DNA according to the method of Mandel and Higa (8). The transfection mixture was plated in H top agar containing 0.02 ml of isopropylthio-β-D-galactoside (20 mg/ml) and 0.05 ml of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (20 mg/ml).

Toothpick assay for *secA* transduction. The bacterial lawn on a plaque-containing plate was killed by chloroform vapors for 20 min. Individual plaques were stabbed with a toothpick, which was subsequently touched to a plate overlaid with 0.1 ml of an overnight culture of MM52 in top agar. After overnight incubation at 42°C, *secA* transducing phage gave a region of dense bacterial growth in the weakly growing surrounding lawn, whereas nontransducing phage gave only faint plaques.

Phage-directed protein synthesis after UV irradiation. Phage-directed protein synthesis after UV irradiation of the bacterial host was performed essentially as described by Lutkenhaus and Wu (7).

RESULTS

Isolation of *secA* transducing phages. Previous genetic analysis showed that the *secA* mutation *secA51*(Ts) is located at min 2.5 on the *E. coli* map (11) in or very near a cluster of genes concerned with cell envelope biogenesis (*ftsA ftsZ envA* [6, 7]). The *secA51* allele is 98% cotransducible with an *envA* allele. We obtained two *envA* specialized transducing phage, λ16-2 and λ16-25, to determine whether they also carry the *secA* gene. Neither of these phage complemented the *secA* mutant at 42°C, although both allowed for the formation of wild-type *secA51*$^+$ recombinants at a frequency of 10^{-4} (Table 1, lines 1 and 2). Two interpretations were possible: either the two phage do carry an intact *secA* gene but the *secA51* chromosomal allele is dominant, or the phage carry only a portion of the *secA* gene. To distinguish between these possibilities, we performed dominance tests on a merodiploid strain.

A *recA* strain containing *secA51* on the chromosome and *secA51*$^+$ on the KLF4 episome, MM59/KLF4, is not temperature sensitive for growth, showing that the *secA51* allele is indeed recessive (see Materials and Methods). The *secA51* allele is also recessive for the defect in maltose binding protein secretion (data not shown). Thus, since the two λ phages do not complement the *secA51* allele, they must carry only a portion of the *secA* gene.

To isolate a specialized transducing phage carrying the *secA* gene, we made use of the orientation of the bacterial genes carried by λ16-25. When λ16-25 integrates into Δ*att*λ host by bacterial DNA homology, the phage DNA is flanked by a duplication of the bacterial DNA (Fig. 1). By selecting for Spi⁻ phage which have lost a number of phage genes by an aberrant excision, given the DNA packaging constraints of λ, one also tends to select for phage which have picked up additional bacterial DNA to the clockwise side of *envA*. Thus, we were able to enrich for λ transducing phages carrying the *secA* gene by their Spi⁻ phenotype as well as by their ability to transduce the *secA* mutant to temperature resistance. The *secA* transducing

TABLE 1. Isolation of *secA* transducing phage[a]

Phage	*murC*	*ddl*	*ftsA*	*ftsZ*	*envA*	*secA*
λ16-2	C	C	C	C	C	R
λ16-25	—	—	C	C	C	R
λDO2	NT	NT	C	C	C	C
λDO11	NT	NT	—	—	—	C
λDO20	NT	NT	—	—	—	C

[a] To test for *ts*$^+$ transduction, 0.01 ml of a phage lysate was spotted onto a tryptone-yeast extract (TYE) plate previously spread with 0.1 ml of an overnight culture of the appropriate strain. Plates were scored after overnight incubation at 42°C. To test for *envA*$^+$ transduction, 0.01 ml of a phage lysate was mixed with 0.1 ml of a log-phase culture of MM63 grown in Luria broth (LB) plus 10 mM $MgSO_4$. After 20 min of adsorption at 37°C, the mixture was diluted 10-fold into LB and grown for 3 h at 37°C. A 0.01-ml portion of this mixture was spotted onto a TYE plate containing 2 μg of rifampin per ml. Plates were scored after overnight incubation at 37°C. Confluent growth in the region of the spot was scored as complementation (C), whereas the presence of single colonies was scored as recombinant (R). —, No complementation or recombination; NT, not tested. The results obtained for *murC* and *ddl* are from Lutkenhaus et al. (6).

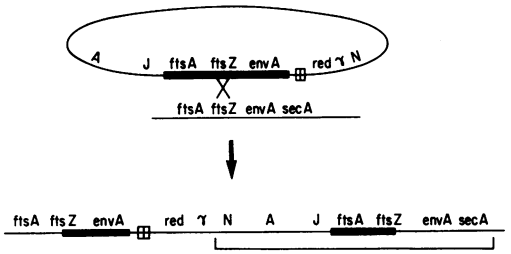


FIG. 1. Integration of λ16-25 carrying the *ftsA-envA* region into a homologous region of the chromosome of a host lacking the primary λ attachment site. The bracketed line below the integrated phage depicts a possible aberrant excision event in which the excising phage loses the *red* and γ genes while picking up additional genes to the right of *envA*, including the *secA* gene.

phage λDO2 isolated in this way (see Materials and Methods) carries not only the *ftsA*, *ftsZ*, and *envA* genes, as does its λ16-25 parent, but also the *secA* gene (Table 1, line 3). Restriction analysis of these two phage shows that λDO2 has picked up an additional 13 kilobases (kb) of bacterial DNA to the right of the bacterial insert present in λ16-25 (data not shown).

Since λDO2 carries so much bacterial DNA, we proceeded to construct a smaller *secA* transducing phage, using recombinant DNA techniques. λDO2 DNA was cleaved with *Eco*RI and mixed with the *Eco*RI-cleaved vector λ616. After ligation and transfection, individual plaques were tested for *secA* transducing ability (see Materials and Methods). One *secA* transducing phage obtained in this manner, λDO11, was chosen for further study. An *Eco*RI digest of λDO11 reveals that this phage is considerably smaller than λDO2 (Fig. 2, lanes 1 and 2). However, λDO11 carries at least three *Eco*RI fragments in addition to those comprising the phage vector (cf. lanes 2 and 6, Fig. 2). To determine the minimum number of *Eco*RI fragments needed for an intact *secA* gene, we repeated the cloning process, using λDO11. λDO11 DNA was cleaved with *Eco*RI, religated, and transfected for phage plaques. Three *secA* transducing phage, λDO20, λDO21, and λDO22, derived from λDO11 in this manner were chosen for further study. An *Eco*RI restriction analysis of these transducing phage (Fig. 2, lanes 3, 4, and 5) shows that, by the second round of restriction and selection for *secA* transducing phage, all three phages obtained gave identical restriction patterns. The *secA* gene must be coded for by portions of two *Eco*RI fragments which are 2.8 and 0.8 kb in size.

We have selected for Tn*5* insertions which inactivate the *secA* gene carried on λDO20 to further locate the *secA* gene on these two re-

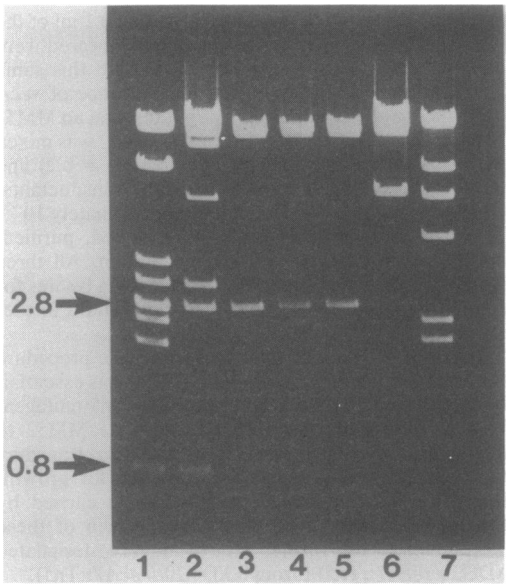


FIG. 2. A 2.5-μg portion of phage DNA in 0.05 ml of 10 mM Tris (pH 7.5)–150 mM NaCl–10 mM $MgCl_2$–1 mM dithiothreitol–100 μg of gelatin per ml was digested with 5 U of *Eco*RI at 37°C for 1 h. A 0.02-ml amount of the reaction mix was loaded onto a 0.8% horizontal agarose gel made with 40 mM Tris–20 mM sodium acetate–1 mM EDTA (pH 8)–1 μg of ethidium bromide per ml and electrophoresed at 150 mA for 2.5 h. The gel was photographed under UV light through a red filter. *Eco*RI digests of λDO2 (lane 1); λDO11 (lane 2); λDO20 (lane 3); λDO21 (lane 4); λDO22 (lane 5); λ616 (lane 6). A standard of λ cleaved with *Hin*dIII is included (lane 7).

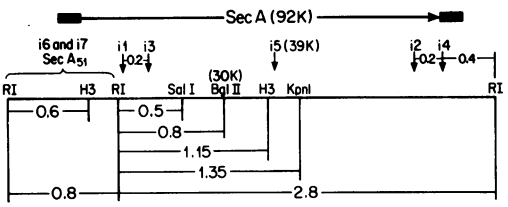


FIG. 3. Genetic organization of the *secA* gene. A restriction map of the region is given. *Eco*RI (RI), *Hin*dIII (H3), *Sal*I (SalI), *Bgl*II (BglII), and *Kpn*I (KpnI) restriction sites are shown, as well as the size of the restriction fragments (in kilobases) below the map. The positions of seven Tn*5* insertions (i1 to i7) are given above the map and were determined by using the restriction enzymes shown in the figure. The location of the *secA51*(Ts) mutation is also given. The approximate location of the *secA* gene coding sequences and the transcriptional direction of the gene are given at the top of the figure. This was possible by using the data given here as well as the translation mapping data presented in Fig. 4. The size of a truncated polypeptide (39K) synthesized by λDO20 (*secA5*::Tn*5*) and the portion of the *secA* gene (30K) included on a *secA-lacZ* fusion constructed in vitro by using the *Bgl*II site in *secA* are shown.

striction fragments. The Tn5 insertions can be easily mapped by restriction enzyme analysis (5). Of seven independent insertions analyzed, two mapped within the 0.8-kb *Eco*RI fragment and five mapped within the 2.8-kb *Eco*RI fragment. The five insertions within the larger restriction fragment map at positions covering most of this fragment (Fig. 3, insertions 1 to 5). This indicates that either the *secA* gene coding sequences occupy most of this DNA fragment (i.e., it is a large gene) or expression of the gene is abolished by insertions outside *secA* coding sequences (e.g., by polarity of Tn5 insertions within a multigene operon, or insertion within a positive regulator of *secA*). To discriminate between these two possibilities, translational mapping experiments were performed.

UV-irradiated cells were infected with λDO20 and its Tn5-containing derivatives to examine phage-directed protein synthesis. Comparison of the phage-directed proteins synthesized by λDO20 and its λ616 parent shows that only λDO20 synthesizes a rather large and prominent polypeptide of approximately 92K (Fig. 4, lanes 2 and 3). The additional two protein species synthesized by λ616, one of which is presumably β-galactosidase, are synthesized from the large *Eco*RI insert present in the λ616 displacement vector (10). Synthesis of the 92K polypeptide is abolished in UV-irradiated cells infected with phage λDO20 (*secA1*::Tn5) and λDO20 (*secA6*::Tn5) (Fig. 4, lanes 4 and 11). These phage contain Tn5 insertions within either the small *Eco*RI fragment or the proximal portion of the large *Eco*RI fragment (i.e., the insertion is immediately adjacent to the small fragment; see

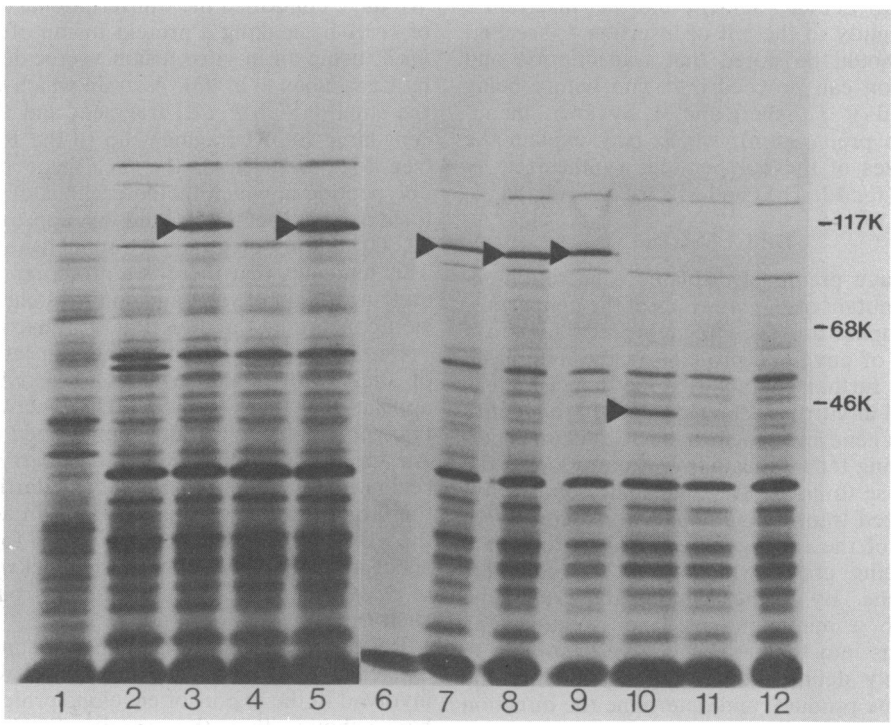


FIG. 4. Strain 159 was grown in M63 media containing 0.2% maltose until mid-logarithmic phase, when it was concentrated to approximately 2×10^9 cells per ml in 0.1 M $MgSO_4$. The cells were irradiated with a germicidal UV light at a distance of 66 cm for approximately 7 min. Phage were adsorbed to 0.1-ml samples of cells (multiplicity of infection = 10) at 37°C for 5 min, when 0.2 ml of prewarmed M63 medium containing 0.2% maltose was added. Five minutes later, 10 μCi of [^{35}S]methionine was added, and incubation was continued for an additional 5 min at 37°C. The phage-infected cells were then chilled on ice, sedimented for 5 min in an Eppendorf centrifuge, washed once, and resuspended in 0.1 ml of 125 mM Tris (pH 6.8)–2% sodium dodecyl sulfate–15% glycerol–5% β-mercaptoethanol. Samples, 0.025 ml, were incubated for 2 min at 100°C and loaded onto a 10% polyacrylamide gel. The gel was run at 35-mA constant current until the tracking dye reached the bottom. The gel was fixed and dried before autoradiography. The labeled protein profile of UV-irradiated 159 cells are given (lanes 1 and 6). Cells infected with λ616 (lane 2), λDO20 (lanes 3 and 7), λDO20(*secA1*::Tn5) (lanes 4 and 12), λDO20(*secA2*::Tn5) (lanes 5 and 8), λDO20(*secA4*::Tn5) (lane 9), λDO20(*secA5*::Tn5) (lane 10), and DO20(*secA6*::Tn5) (lane 11) are also given. The molecular weight standards are: β-galactosidase, 117,000; bovine serum albumin, 68,000; ovalbumin, 46,000.

Fig. 3). This result identifies the 92K polypeptide as the presumptive *secA* gene product (gp *secA*).

Infection with the other Tn*5*-containing phage gave a protein pattern consistent with all insertions being within the *secA* gene and the gene being transcribed left to right (Fig. 3). The λDO20 (*secA5*::Tn*5*) phage, whose insertion maps toward the middle of the large *Eco*RI fragment, synthesizes a truncated polypeptide of approximately 39K (Fig. 4, lane 10). λDO20 (*secA2*::Tn*5*) and λDO20 (*secA4*::Tn*5*) synthesize polypeptides of similar size to gp *secA* (cf. Fig. 4, lanes 7, 8, and 9). The polypeptide from λDO20 (*secA2*::Tn*5*) appears to be slightly smaller than gp *secA*, whereas the λDO20 (*secA4*::Tn*5*)-coded polypeptide is of similar molecular weight. This is compatible with the fact that both insertions map in the distal end of the 2.8-kb *Eco*RI fragment and that insertion 2 maps slightly to the left of insertion 4 (see Fig. 3). It should be noted that transcription and translation can proceed into Tn*5* before being abolished (R. R. Isberg and M. Syvanen, manuscript in preparation), which may explain the large sizes of the polypeptides synthesized by λDO20 (*secA2*::Tn*5*) and λDO20 (*secA4*::Tn*5*).

DISCUSSION

We have previously isolated a secretion-defective mutant of *E. coli* which at the nonpermissive temperature accumulates precursors to a number of envelope proteins in the cytoplasm (11). To further characterize the defect both in vivo and in vitro, we have proceeded to identify the *secA* gene and gene product. Standard genetic mapping techniques allowed us to locate the gene close to *envA*. The availability of an *envA* specialized transducing phage which contains a portion of the *secA* gene enabled us to isolate transducing phage which carry the complete *secA* gene. By further cloning the *secA* gene down to a minimum size and isolating Tn*5* insertions into the gene, we have been able to accurately define the *secA* gene, identify the *secA* gene product, and determine the direction of transcription of the gene.

secA is a new and rather large gene to the clockwise side of *envA*. The *envA* gene is contained on a 2.5-kb *Eco*RI fragment (7). *secA* is contained on two *Eco*RI fragments, 0.8 and 2.8 kb. The order of these fragments must be 2.5, 0.8, and 2.8 kb. This is the case since λ16-2 and λ16-25 both carry the 2.5-kb fragment and a portion of the 0.8-kb fragment (up to the *Hin*dIII site; see Fig. 3). This portion of the 0.8-kb fragment contains the portion of the *secA* gene corresponding to the *secA51* allele originally described (11). Furthermore, we have verified that λ16-25 specifically hydridizes only to the correct portion of the 0.8-kb fragment by Southern analysis (unpublished data).

The *secA* gene product is a polypeptide of approximately 92K. Tn*5* insertions in the very beginning of the gene result in the synthesis of no detectable gene product even in 15% polyacrylamide gels (unpublished data). A Tn*5* insertion toward the middle of the gene results in the synthesis of a truncated 39K polypeptide. Insertions at the end of the gene give rise to polypeptides of similar size to gp *secA*. These results are compatible with the observation that the former two classes of insertions fail to complement the *secA* temperature-sensitive mutant MM52 in diploid analysis, whereas the third class appears to give partial complementation (unpublished data).

The *secA* gene is transcribed in a clockwise direction, opposite to that inferred for *envA* (7). We have confirmed the transcriptional direction of *secA* by making a protein fusion of *secA* to *lacZ*, using an in vitro fusion vector developed by Casadaban et al. (3). A strain which contains the small 0.8-kb *Eco*RI fragment and the adjacent large *Eco*RI fragment up to the *Bgl*II site (see Fig. 3) fused to the *lacZ* gene makes a polypeptide in which the N terminus of β-galactosidase has been replaced by approximately 30,000 daltons of gp *secA* (unpublished data). This indicates that there is a promoter immediately upstream of *secA* and confirms our conclusion concerning the direction of transcription.

Combining our restriction and translation map of *secA* with a similar map for *envA* (7), it appears that there is enough DNA between the two genes to code for an average-sized protein. An appropriate transducing phage carrying this region intact could be constructed starting with the large transducing phage λDO2. In addition, the λDO2 transducing phage is useful for analysis of mutants which map to the clockwise side of *envA*, since this phage must carry roughly 10 such genes.

Results presented in this paper document the existence of a new gene in *E. coli* which is involved in the export of envelope proteins. We have shown that the previously described *secA*(Ts) allele (11) is recessive for both secretion and growth defects. This finding limits the number of models which can explain the nature of the secretion block in *secA* mutants. Models invoking the presence of an abnormal protein which somehow interferes with secretion by jamming up the membrane, for example, appear to be less likely. Location and quantitation of the *secA* gene product within the cell should further help to define its function.

ACKNOWLEDGMENTS

We thank Joe Lutkenhaus for advice and Ralph Isberg for suggesting the Spi⁻ selection to obtain the *secA* transducing

phage. We also thank Terry Luna for technical assistance and Ann McIntosh for assistance in the preparation of this manuscript.

This work was supported by a National Science Foundation grant (PCM-7922624) to J.B. D.B.O. is a recipient of a Public Health Service post-doctoral fellowship from the National Institute of General Medical Science.

ADDENDUM IN PROOF

In further subcloning of the *secA* gene, we found that the position of one of the restriction sites given in Fig. 3 was incorrect. The *Hin*III site in the 2.8-kb *Eco*RI fragment should be just to the left of the *Bgl*II site shown in the figure.

LITERATURE CITED

1. **Bachmann, B. J., and K. B. Low.** 1980. Linkage map of *Escherichia coli* K-12, edition 6. Microbiol. Rev. **44**:1–56.
2. **Berg, D. E., J. Davies, B. Allet, and J. D. Rochaix.** 1975. Transposition of R factor genes to bacteriophage lambda. Proc. Natl. Acad. Sci. U.S.A. **72**:3628–3632.
3. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.
4. **Gottesman, S., and J. R. Beckwith.** 1969. Directed transposition of the arabinose operon. A technique for the isolation of specialized transducing bacteriophages for any *Escherichia coli* gene. J. Mol. Biol. **44**:117–127.
5. **Jorgensen, S. J., S. J. Rothstein, and W. S. Reznikoff.** 1980. A restriction enzyme cleavage map of Tn*5* and location of a region encoding neomycin resistance. Mol. Gen. Genet. **177**:65–72.
6. **Lutkenhaus, J. F., J. Wolf-Watz, and W. D. Donachie.** 1980. Organization of genes in the *ftsA-envA* region of the *Escherichia coli* genetic map and identification of a new *fts* locus (*ftsZ*). J. Bacteriol. **142**:615–620.
7. **Lutkenhaus, J. F., and J. C. Wu.** 1980. Determination of transcriptional units and gene products from the *ftsA* region of *Escherichia coli*. J. Bacteriol. **143**:1281–1288.
8. **Mandel, M., and A. Higa.** 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159–162.
9. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Murray, N. E., W. J. Brammar, and K. Murray.** 1977. Lambdoid phages that simplify the recovery of *in vitro* recombinants. Mol. Gen. Genet. **150**:53–61.
11. **Oliver, D. B., and J. Beckwith.** 1981. *Escherichia coli* mutant pleiotropically defective in the export of secreted proteins. Cell **25**:765–772.
12. **Zissler, J., E. Signer, and F. Schaefer.** 1971. The role of recombination in growth of bacteriophage lambda. II. Inhibition of growth by prophage P2, p. 469–475. *In* A.D. Hershey (ed.), The bacteriophage lambda. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

achieve a productive and substrate-specific catalytic complex, and this induced-fit mechanism provides a direct opportunity for allosteric regulation. In analogy to an allosteric dimer, the active site of AC is formed at the interface between dyad-related homologous domains. Thus, AC provides two symmetrically related binding sites for the homologous G proteins $G_s\alpha$ and $G_i\alpha$ (*4, 5*). We suggest that $G_s\alpha$ facilitates collapse of the active-site loops from C_{1a} and C_{2a} around the substrate ATP to form an competent active site, whereas $G_i\alpha$ hinders such collapse and stabilizes an open, inactive conformation of the enzyme.

References and Notes

1. R. K. Sunahara, C. W. Dessauer, A. G. Gilman, *Annu. Rev. Pharmacol. Toxicol.* **36**, 461 (1996); D. M. F. Cooper, Ed., *Advances in Second Messenger and Phosphoprotein Research*, vol. 32 (Lippincott-Raven, Philadelphia, PA, 1998).
2. R. K. Sunahara, C. W. Dessauer, R. E. Whisnant, C. Kleuss, A. G. Gilman, *J. Biol. Chem.* **272**, 22265 (1997).
3. R. E. Whisnant, A. G. Gilman, C. W. Dessauer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6621 (1996); S.-Z. Yan, D. Hahn, Z.-H. Huang, W.-J. Tang, *J. Biol. Chem.* **271**, 10941 (1996).
4. C. W. Dessauer, J. J. Tesmer, S. R. Sprang, A. G. Gilman, *J. Biol. Chem.* **273**, 25831 (1998).
5. J. J. G. T. Tesmer, R. K. Sunahara, A. G. Gilman, S. R. Sprang, *Science* **278**, 1907 (1997).
6. D. L. Garbers and R. A. Johnson, *J. Biol. Chem.* **250**, 8449 (1975); G. Zimmermann, D. Zhou, R. Taussig, *ibid.* **273**, 19650 (1998).
7. P. J. Artymiuk, A. R. Poirrette, D. W. Rice, P. Willett, *Nature* **388**, 33 (1997).
8. S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *ibid.* **391**, 251 (1998).
9. H. Huang, R. Chopra, G. L. Verdine, S. C. Harrison, *Science* **282**, 1669 (1998).
10. Y. Li, S. Korolev, G. Waksman, *EMBO J.* **17**, 7514 (1998).
11. N. Kaushik *et al.*, *Biochemistry* **35**, 11536 (1996); A. Y. Woody, S. S. Eaton, P. A. Osumi-Davis, R. W. Woody, *ibid.*, p. 144.
12. I. Shoshani, V. Boudou, C. Pierra, G. Gosselin, R. A. Johnson, unpublished data.
13. R. K. Sunahava and J. J. G. Tesmer, data not shown.
14. F. Eckstein, P. Romaniuk, W. Heideman, D. Storm, *J. Biol. Chem.* **256**, 9118 (1981).
15. The recombinant C_{1a} domain from canine type V AC, the C_{2a} domain from rat type II AC, and bovine $G_{s\alpha}$ were expressed in *E. coli* and purified as previously described (*30*). $G_{s\alpha}$ was activated with Mg^{2+} · GTPγS and then subjected to limited proteolysis by trypsin. Crystals of the ternary complex between C_{1a}, C_{2a}, and $G_{s\alpha}$ were grown as previously described (*5*). All crystals were harvested in cryoprotectant (*5*) to which was added various combinations of inhibitors and metal ions, described as follows: for AC · βLddATP · Mg, 5 mM $MgCl_2$ and 380 μM β-L-2′,3′-dd-5′-ATP; for AC · βLddATP · Mn, 250 μM $MgCl_2$, 500 μM $MnCl_2$, and 380 μM β-L-2′,3′-dd-5′-ATP; for AC · βLddATP · Zn, 250 μM $MgCl_2$, 500 μM $ZnCl_2$, and 380 μM β-L-2′,3′-dd-5′-ATP; for AC · ATPαS-R_p · Mn, 500 μM $MgCl_2$, 500 μM $MnCl_2$, and 1 mM ATPαS-R_p. Crystals of AC · βLddATP · Mg, AC · βLddATP · Mn, and AC · βLddATP · Zn were soaked for 1 to 2 hours; the conformation of AC in these complexes was closed. AC · ATPαS-R_p · Mn was soaked for 6 hours to achieve a closed conformation; 2-hour soaks resulted in structures of the open conformation of AC. AC complexes with inhibitors such as ATP · Zn^{2+}, ATPαS-S_p · Zn^{2+}, and 2′-iodo-ATP · Mg^{2+} also adopt the open conformation, and each metal site exhibits the same preference for either Mn^{2+} or Zn^{2+} as observed in the closed conformation (*13*). Strong electron density is observed for the β and γ phosphates of these inhibitors in the presence of Mg^{2+} alone or of Mn^{2+} and Mg^{2+}, and for the entire triphosphate in the presence of Zn^{2+} and Mg^{2+}. Much weaker density corresponding to a purine ring is observed in the purine binding pocket of these complexes. However, the purine and triphosphate densities are not continuous, and a single model of ATP cannot be built to accommodate both purine and triphosphate binding sites.
16. E. E. Kim and H. W. Wyckoff, *J. Mol. Biol.* **218**, 449 (1991).
17. V. Derbyshire, J. K. Pinsonneault, C. M. Joyce, *Methods Enzymol.* **262**, 363 (1995).
18. J. A. Piccirilli, J. S. Vyle, M. H. Caruthers, T. R. Cech, *Nature* **361**, 85 (1993).
19. H. Pelletier, M. R. Sawaya, W. Wolfle, S. H. Wilson, J. Kraut, *Biochemistry* **35**, 12762 (1996); I. Slaby, B. Lind, A. Holmgren, *Biochem. Biophys. Res. Commun.* **122**, 1410 (1984); G. C. King, C. T. Martin, T. T. Pham, J. E. Coleman, *Biochemistry* **25**, 36 (1986); E. Ferrari *et al.*, *J. Virol.* **73**, 1649 (1999).
20. S.-Z. Yan, Z.-H. Huang, R. S. Shaw, W.-J. Tang, *J. Biol. Chem.* **272**, 12342 (1997).
21. Y. Liu, A. E. Ruoho, V. D. Rao, J. H. Hurley, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13414 (1997).
22. W. S. Sheldrick and E. Rieke, *Acta Crystallogr.* **B34**, 2324 (1978); K. I. Varughese, C. T. Lu, G. Kartha, *J. Am. Chem. Soc.* **104**, 3398 (1982); N. Padmaja, S. R. Ramakumar, M. A. Viswamitra, *Bull. Chem. Soc. Jpn.* **64**, 1359 (1991).
23. C. W. Dessauer and A. G. Gilman, *J. Biol. Chem.* **272**, 27787 (1997).
24. W.-J. Tang, M. Stanzel, A. G. Gilman, *Biochemistry* **34**, 14563 (1995).
25. J. J. Tesmer and S. R. Sprang, *Curr. Opin. Struct. Biol.* **8**, 713 (1998).
26. L. S. Beese and T. A. Steitz, *EMBO J.* **10**, 25 (1991); T. A. Steitz, *Curr. Opin. Struct. Biol.* **3**, 31 (1993); H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, J. Kraut, *Science* **264**, 1891 (1994).
27. S. Doublié, M. R. Sawaya, T. Ellenberger, *Structure* **7**, R31 (1999).
28. S. Doublié and T. Ellenberger, *Curr. Opin. Struct. Biol.* **8**, 704 (1998).
29. S. G. Sarafianos, V. N. Pandey, N. Kaushik, M. J. Modak, *J. Biol. Chem.* **270**, 19729 (1995); V. N. Pandey, N. Kaushik, M. J. Modak, *ibid.* **269**, 13259 (1994).
30. R. K. Sunahara *et al.*, *J. Biol. Chem.* **273**, 16332 (1998).
31. Single-letter abbreviations for the amino acid residues are as follows: D, Asp; I, Ile; K, Lys; N, Asn; R, Arg.
32. A. G. W. Leslie, MOSFLM 5.50 for image plate data (1997).
33. S. Bailey, *Acta Crystallogr.* **D50**, 760 (1994).
34. A. T. Brünger *et al.*, *ibid.* **D54**, 905 (1998).
35. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *ibid.* **A47**, 110 (1991).

36. We thank F. Eckstein at the Max-Planck Institute (Göttingen) for pure ATPαS-R_p; J. Collins for technical assistance; C. Brautigam for helpful discussion and for reading the manuscript; D. Coleman and the MacCHESS staff for their assistance with data collection at the Cornell High Energy Synchrotron Source (CHESS); and L. Esser for his assistance in preparing figures. R.K.S. was supported by a postdoctoral fellowship from the Medical Research Council of Canada. This work was supported by grants from the Agence Nationale de Recherche sur le SIDA (ANRS, France) to G.G.; by NIH grants DK38828 (R.A.J.), GM34497 (A.G.G.), and DK46371 (S.R.S.); by an Innovative Technology Grant from the Stony Brook Center for Biotechnology to R.A.J.; by Welch Foundation grants I-1271 (A.G.G.) and I-1229 (S.R.S.); and by the Raymond and Ellen Willie Distinguished Chair of Molecular Neuropharmacology (A.G.G.). Coordinates for the four models have been deposited in the Protein Data Bank with the codes 1CJT, 1CJU, 1CJV, and 1CJK.

26 April 1999; accepted 22 June 1999

Staphylococcus aureus Sortase, an Enzyme that Anchors Surface Proteins to the Cell Wall

Sarkis K. Mazmanian, Gwen Liu, Hung Ton-That, Olaf Schneewind*

Surface proteins of Gram-positive bacteria are linked to the bacterial cell wall by a mechanism that involves cleavage of a conserved Leu-Pro-X-Thr-Gly (LPXTG) motif and that occurs during assembly of the peptidoglycan cell wall. A *Staphylococcus aureus* mutant defective in the anchoring of surface proteins was isolated and shown to carry a mutation in the *srtA* gene. Overexpression of *srtA* increased the rate of surface protein anchoring, and homologs of *srtA* were found in other pathogenic Gram-positive bacteria. The protein specified by *srtA*, sortase, may be a useful target for the development of new antimicrobial drugs.

Hospital isolates of *Staphylococcus aureus*, *Staphylococcus epidermidis*, and *Enterococcus faecalis* have become resistant to most, if not all, known therapeutic regimens (*1*). Many antibiotics, including penicillin and its derivatives, target the transpeptidation reaction of bacterial cell wall synthesis, which cross-links peptidoglycan strands (*2*). To search for other cell wall synthesis reactions that may serve as targets for antimicrobial therapy, we have focused on the anchoring of surface proteins to the peptidoglycan of Gram-positive bacteria.

Department of Microbiology and Immunology, UCLA School of Medicine, University of California, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA.

*To whom correspondence should be addressed. E-mail: olafs@ucla.edu

Surface proteins not only promote interaction between the invading pathogen and animal tissues, but also provide ingenious strategies for bacterial escape from the host's immune response (*3*). In the case of *S. aureus* protein A, immunoglobulins are captured on the microbial surface and camouflage bacteria during the invasion of host tissues (*4*). Protein A is cleaved by a transpeptidase, sortase, between the threonine and the glycine of a conserved LPXTG motif (*5*). The carboxyl group of threonine is amide-linked to the amino group of the pentaglycine cross-bridge, thereby tethering the COOH-terminal end of protein A to the bacterial cell wall (*6*). This reaction, called cell wall sorting, is strikingly similar to the penicillin-sensitive transpeptidation reaction, and is likely to occur in most Gram-positive bacteria (*7*).

To identify cell wall sorting mutants, we mutagenized *S. aureus* strain OS2 with nitrosoguanidine (*8*). Temperature-sensitive (ts) mutants were identified and 1000 were transformed with pSEB-SPA$_{490\text{-}524}$, a plasmid encoding a reporter protein that allows measurement of surface protein anchoring (*9*). The SEB-SPA$_{490\text{-}524}$ precursor (P1) is exported from the cytoplasm, and its NH_2-terminal leader peptide is removed to generate the P2 intermediate (Fig. 1A) (*10*). P2 is cleaved by sortase at the LPXTG motif to generate the mature, surface-anchored protein (M). After labeling with [^{35}S]Met for 5 min, the reporter protein in strain OS2 was distributed as follows: P1 (5%), P2 (19%), and M (76%) (Fig. 1B) (*11*). We used this assay to screen 1000 ts mutants and identified two strains in which there was aberrant accumulation of P2 (47% in SM317 and 26% in SM329) (Fig. 1B). Pulse-chase analysis (*12*) revealed that in strain OS2, P2 was cleaved and anchored within 2 min, whereas in strain SM317 these events required more than 10 min (Fig. 1C). In strain SM329, P2 processing required 3 min, suggesting a mild defect in cell wall sorting (*13*).

Previous work showed that mutations in *S. aureus fem* genes slow the anchoring of surface proteins to the cell wall (*14*). These genes are thought to specify enzymes that catalyze the addition of glycines to the ε-amino of lysine within the peptidoglycan precursor lipid II (*15*). To examine whether the SM317 and SM329 mutants were defective in cell wall synthesis, we tested their sensitivity to lysostaphin, an enzyme that cuts the pentaglycine cross-bridges of the staphy-

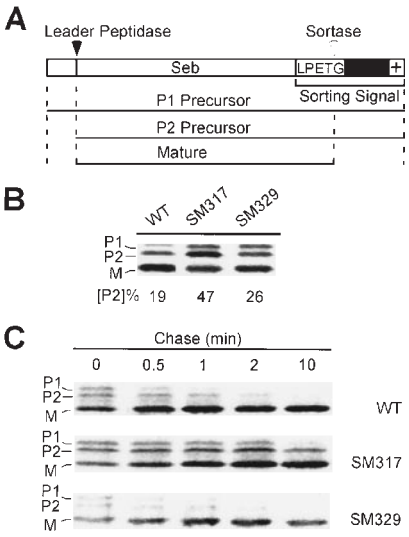


Fig. 1. Isolation of a staphylococcal mutant defective in cell wall sorting of surface proteins. (**A**) Primary structure of the surface protein precursor SEB-SPA$_{490\text{-}524}$, a fusion between enterotoxin B (SEB) and COOH-terminal protein A (SPA) sequences. P1 is directed across the cytoplasmic membrane by an NH_2-terminal leader peptide, and is then cleaved to generate P2. P2 bears a COOH-terminal sorting signal that includes an LPXTG motif, a hydrophobic domain (black bar), and a positively charged tail (boxed +). The sorting signal of P2 is cleaved at the LPXTG motif, and the mature protein (M) is linked to the cell wall. (**B**) *Staphylococcus aureus* ts mutants were screened for the accumulation of P2 by labeling with [^{35}S]Met. SM317 and SM329 accumulate more P2 than does the wild-type (WT) strain OS2. (**C**) Pulse-chase analysis of SEB-SPA$_{490\text{-}524}$ anchoring in wild-type and mutant strains.

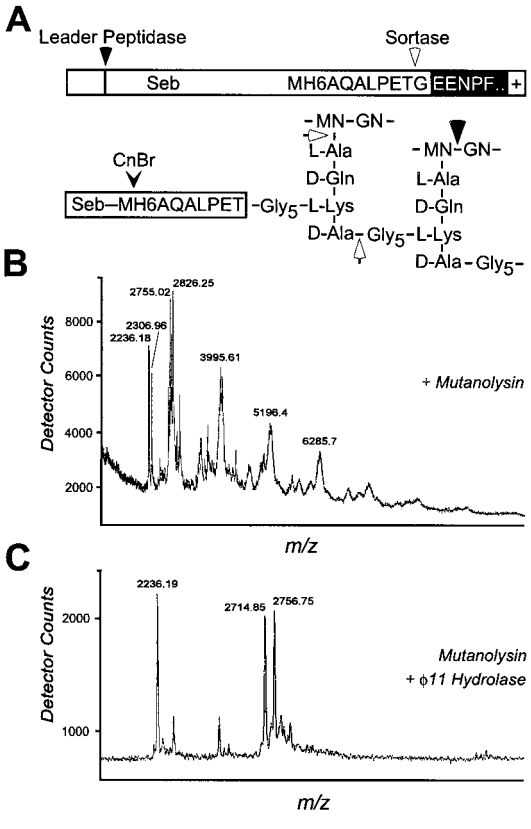


Fig. 2. Structure of surface protein anchor in strain SM317. (**A**) Primary structure of SEB-MH$_6$-CWS and its linkage to the cell wall. The glycan strands of the staphylococcal cell wall consist of a repeating disaccharide *N*-acetylmuramic acid-(β1-4)-*N*-acetylglucosamine (MN-GN) in which the lactyl of muramic acid is linked to the wall peptide (L-Ala-D-iGln-L-Lys-D-Ala). Wall peptides are linked to the pentaglycine cross-bridge, which tethers the ε-amino of L-Lys to the carboxyl of D-Ala. The cell wall can be cut at specific sites with the enzymes muramidase (solid arrow) and ϕ11 hydrolase (open arrows). (**B**) SEB-MH$_6$-CWS was solubilized by digesting the staphylococcal cell wall with muramidase (mutanolysin) and purified by affinity chromatography on Ni-NTA resin. The polypeptide was cleaved with cyanogen bromide (CnBr), and COOH-terminal anchor peptides were purified by another affinity chromatography step. The structure of the mutanolysin-released anchor peptides of strain SM317 was analyzed by MALDI-MS (*6*). (**C**) Muramidase-released anchor peptides were digested with ϕ11 hydrolase and analyzed by MALDI-MS. The observed ions represent anchor peptide linked to cell wall tetrapeptide (*m/z* 2236) and murein-disaccharide tetrapeptide (*m/z* 2715 and 2757).

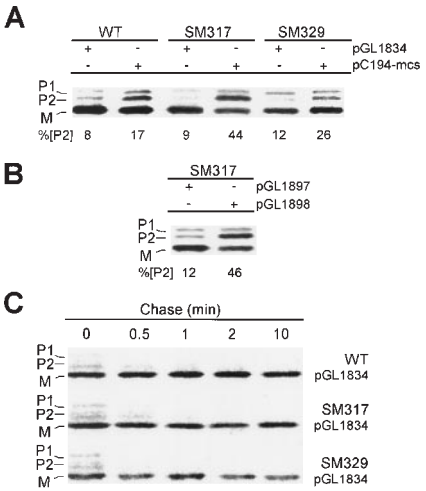


Fig. 3. Overexpression of *srtA* reduces P2 accumulation in wild-type *S. aureus* strain OS2 and in two mutant strains, SM317 and SM329. (**A**) Transformants of a multicopy plasmid library in strain SM317 were screened by labeling with [^{35}S]Met for a decrease in the accumulation of P2. Plasmid pGL1834 contains the *srtA* gene cloned into pC194-mcs. (**B**) The *srtA* gene of strain SM317 (pGL1898) or of strain OS2 (pGL1897) was transformed into SM317 and analyzed for P2 processing. (**C**) All three strains were transformed with pGL1834 and subjected to pulse-chase analysis (*12*).

lococcal cell wall (*16*). In contrast to *fem* mutants, which are resistant to lysostaphin (*17*), strains SM317 and SM329 were sensitive at concentrations that also inhibited growth of wild-type staphylococci, indicating that their sorting defects are not caused by a mutationally altered cell wall cross-bridge (*18*). To measure cell wall synthesis, we grew the wild-type and mutant strains in minimal medium containing [^{3}H]Lys or [^{3}H]Leu. Because lysine (but not leucine) is a component of the cell wall, the ratio of [^{3}H]Lys:[^{3}H]Leu incorporation into acid-precipitable and protease-resistant murein polymer is a measure of cell wall synthesis (*19*). Wild-type *S. aureus* displayed a ratio of 30, and the inhibition of cell wall synthesis by vancomycin reduced this ratio to 1.5. Strains SM317 and SM329 displayed ratios of 18 and 19, respectively, which indicates that the accumulation of P2 in strain SM317 is not caused by a defect in cell wall synthesis.

To determine the cell wall anchor structure of surface proteins in strain SM317, we introduced into cells a plasmid (pHTT4) specifying the reporter protein SEB-MH$_6$-CWS (Fig. 2A) (*6*). The cell wall was purified and digested with mutanolysin, which hydrolyzes the glycan strands (*6*). Mutanolysin-released surface protein was purified by chromatography on nickel-nitrilotriacetic acid (Ni-NTA) and cleaved at Met with cyanogen bromide. COOH-terminal peptides bearing cell wall anchor structures were purified by a second affinity chromatography step and analyzed by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) (Fig. 2). A series of ion signals with regularly spaced mass increments was revealed. These measurements are consistent with increasing numbers of peptidoglycan subunits linked to the COOH-terminal threonine of surface protein (*20*). If surface protein is tethered to cross-linked peptidoglycan of strain SM317, digestion of muramidase-solubilized anchor peptides with ϕ11 hydrolase should produce anchor peptide linked to murein tetrapeptide and disaccharide-tetrapeptide (*21*) (Fig. 2). This was tested, and the doubly digested anchor peptides generated the predicted ion signals (Fig. 2C). Thus, surface proteins of *S. aureus* SM317 are tethered to cross-linked peptidoglycan in a manner that is indistinguishable from the anchor structure of polypeptides in wild-type staphylococci. These results suggest that the accumulation of P2 in strain SM317 is caused by a defect in sortase.

We reasoned that overexpression of sortase from a multicopy plasmid should reduce the concentration of P2 in both wild-type and mutant *S. aureus*. A plasmid library of 2000 random DNA fragments from *S. aureus* OS2 was screened for sequences that reduce the accumulation of P2 in strain SM317 and two plasmids, pGL1631 and pGL1834, were identified (Fig. 3) (*22*). Transformation with pGL1834 reduced the P2 concentration in strain SM317 by 35%, in strain SM329 by 14%, and in wild-type *S. aureus* OS2 by 9%. All three strains showed a rapid increase in P2 processing (Fig. 3C). We mapped the critical *S. aureus* sequences in pGL1631 and pGL1834 to a gene that we named *srtA* (surface protein sorting A) (*18, 23*).

The *srtA* gene specifies a protein of 206 amino acids with a potential NH$_2$-terminal signal peptide/membrane anchor sequence and a presumed active-site cysteine at position 184, consistent with the observation that the cell wall sorting reaction is sensitive to reagents that modify sulfhydryl groups (Fig. 4) (*10*). Database searches revealed that *srtA* homologs are present in *Actinomyces naeslundii*, *Bacillus subtilis*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Streptococcus mutans*, *Streptococcus pneumoniae*, and *Streptococcus pyogenes*. All *srtA* homologs displayed absolute conservation of the cysteine-encoding codon 184 (*18*).

To examine whether the defect in cell wall sorting of *S. aureus* SM317 is caused by a *srtA* mutation, we amplified the genes from *S. aureus* OS2 and SM317 by PCR, and cloned them into a multicopy vector, which was then transformed into *S. aureus* SM317 (*18, 23*). The wild-type (OS2) *srtA* gene reduced the accumulation of P2 by 32%, whereas the mutant had no effect (Fig. 3B). DNA sequence analysis revealed mutations in codons 35 and 180 in the SM317 *srtA* gene. Interestingly, multicopy expression of wild-type *srtA* (pGL1894) did not fully complement the ts growth phenotype of SM317. The transformed mutant grew at 42°C but slower than the wild type, suggesting that the conditional lethal phenotype of *S. aureus* SM317 is not caused solely by the mutations in *srtA*.

Together, our data reveal that *srtA* encodes sortase, the transpeptidase that anchors surface proteins to the bacterial cell wall (*24*). In principle, purified SrtA protein can be used to screen for compounds that inhibit cell wall sorting, a strategy that may lead to new therapies for human infections caused by Gram-positive bacteria.

```
MKKWTNRLMT IAGVVLILVA AYLFAKPHID   30
NYLHDKDKDE KIEQYDKNVK EQASKDKKQQ   60
AKPQIPKDKS KVAGYIEIPD ADIKEPVYPG   90
PATPEQLNRG VSFAEENESL DDQNISIAGH  120
TFIDRPNYQF TNLKAAKKGS MVYFKVGNET  150
RKYKMTSIRD VKPTDVGVLD EQKGKDKQLT  180
LITSDDYNEK TGVWEKRKIF VATEVK*     206
```

Fig. 4. Deduced amino acid sequence of the *srtA* gene (*25*). The NH$_2$-terminal hydrophobic membrane anchor sequence is boxed. The single Cys is shaded. The *srtA* gene of strain SM317 carries two mutations, one in codon 35 replacing Asp (GAT) with Gly (GGT), and another in codon 180 replacing Thr (ACA) with Ala (AGA). The altered amino acids are indicated in bold. The DNA sequence of plasmid pGL1834 has been submitted to GenBank (accession number AF162687).

References and Notes

1. H. S. Gold and R. C. Moellering, *N. Engl. J. Med.* **335**, 1445 (1996); K. Hiramatsu *et al.*, *Lancet* **350**, 1670 (1997); K. Sieradzki, R. B. Roberts, S. W. Haber, A. Tomasz, *N. Engl. J. Med.* **340**, 517 (1999).
2. J. L. Strominger, *Johns Hopkins Med. J.* **133**, 63 (1973); C. T. Walsh, *Science* **261**, 308 (1993).
3. W. W. Navarre and O. Schneewind, *Microbiol. Mol. Biol. Rev.* **63**, 174 (1999).
4. M. Uhlén, B. Guss, B. Nilsson, F. Götz, M. Lindberg, *J. Bacteriol.* **159**, 713 (1984).
5. W. W. Navarre and O. Schneewind, *Mol. Microbiol.* **14**, 115 (1994).
6. H. Ton-That, K. F. Faull, O. Schneewind, *J. Biol. Chem.* **272**, 22285 (1997); W. W. Navarre, H. Ton-That, K. F. Faull, O. Schneewind, *ibid.* **273**, 29135 (1998).
7. O. Schneewind, A. Fowler, K. F. Faull, *Science* **268**, 103 (1995).
8. Staphylococci (10^{12} colony-forming units) were treated with *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (0.2 mg/ml) for 45 min at 30°C, and mutagenesis was quenched by the addition of 2 volumes of 100 mM sodium phosphate (pH 7.0). About 80% of the mutagenized population was killed, and the mutational frequency of rifampicin-resistant *rpoB* mutations was increased to 1.2×10^{-4}. Ts mutants were selected three times in succession by growing the mutagenized population in tryptic soy broth (TSB) at 42°C and treating with penicillin G (8 μg/ml) for 2 hours. Colonies were formed at 30°C, streaked on TSB agar, and examined for growth at 42°C.
9. O. Schneewind, D. Mihaylova-Petkov, P. Model, *EMBO J.* **12**, 4803 (1993).
10. H. Ton-That and O. Schneewind, *J. Biol. Chem.*, in press.
11. *Staphylococcus aureus* was grown in TSB supplemented with chloramphenicol (cm; 10 μg/ml) or tetracycline (tet; 2 μg/ml) at 30°C until the culture reached an optical density at 660 nm of 0.6. Cells were incubated at 42°C for 20 min, centrifuged at 15,000*g* for 3 min, and washed with 1 ml of prewarmed minimal medium (*12*). Cells were labeled with 50 μCi of [^{35}S]-Promix (Amersham) for 5 min, and surface protein processing was quenched by addition of 75 μl of 100% trichloroacetic acid (TCA). TCA precipitates were collected by centrifugation, washed in acetone, and dried. Samples were suspended in 1 ml of 0.5 M tris-HCl (pH 7.5) and incubated for 1 hour at 37°C after addition of 50 μl of lysostaphin (2 mg/ml). Proteins were again precipitated with TCA and washed with acetone; after immunoprecipitation with α-enterotoxin B, they were analyzed by 14% SDS–polyacrylamide gel electrophoresis and PhosphorImager.
12. O. Schneewind, P. Model, V. A. Fischetti, *Cell* **70**, 267 (1992).
13. In pulse-labeling experiments, strain SM329 had a more severe defect in cell wall sorting when grown in minimal medium rather than TSB. The reason for this phenomenon is not known.
14. H. Ton-That, H. Labischinski, B. Berger-Bächi, O. Schneewind, *J. Biol. Chem.* **273**, 29143 (1998).
15. B. Berger-Bächi, *Trends Microbiol.* **2**, 389 (1994).
16. C. A. Schindler and V. T. Schuhardt, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 414 (1964).
17. U. Kopp, M. Roos, J. Wecke, H. Labischinski, *Microb. Drug Resist.* **2**, 29 (1996).
18. Lysostaphin sensitivity data, mapping data, and the sequence alignment can be viewed at *Science* Online (www.sciencemag.org/feature/data/1041555.shl).
19. D. Boothby, L. Daneo-Moore, G. D. Shockman, *Anal. Biochem.* **44**, 645 (1971).
20. Ion signals of mutanolysin-solubilized anchor peptides correspond to H_6AQALPET-Gly_5 linked to cell wall tetrapeptide (predicted mass 2235; observed 2236), pentapeptide (predicted mass 2306; observed 2307), N,O6-diacetyl-MurNac-GlcNac tetrapeptide

(predicted mass 2755, observed 2755), N,O6-diacetyl-MurNac-GlcNac pentapeptide (predicted mass 2828, observed 2826), murein-tetrapeptide-murein-pentapeptide (predicted mass 3990, observed 3995), (murein-tetrapeptide)$_2$-murein-pentapeptide (predicted mass 5194; observed 5196), and (murein-tetrapeptide)$_4$ (predicted mass 6285, observed 6286).

21. W. W. Navarre, H. Ton-That, K. F. Faull, O. Schneewind, *J. Biol. Chem.* **274**, 15847 (1999).
22. Plasmid pGL4 contains the coding sequence of SEB-SPA$_{490-524}$, which was released from pSEB-SPA$_{490-524}$ by Eco RI–Bam HI digestion and inserted into the pT181 derivative pWil5. *S. aureus* SM317 (pGL4) was grown on TSB tet agar. A plasmid library of *S. aureus* OS2 chromosomal DNA was obtained by partial digestion with Sau 3A1. DNA fragments of 3 to 5 kb were purified and cloned into Bam HI–digested pC194-mcs, which contains the multiple cloning site of pUC19 inserted into the Hind III site of pC194. SM317 (pGL4) was transformed with the pC194-mcs plasmid library and transformants were selected on TSB tet-cm agar.
23. The DNA insertions of pGL1631 and 1834 were mapped and sequenced by synthesizing oligonucleotide primers. The primers for the amplification of *srtA* from the chromosomal DNA of *S. aureus* strains OS2 (pGL1897) and SM317 (pGL1898) were 5′-AAGGATCCAAAAGGAGCGGTATACATTGC-3′ and 5′-AAGGATCCTACCTTTTCCTCTAGCTGAAG-3′.
24. In another report we show that purified SrtA protein catalyzes the in vitro transpeptidation of substrates bearing an LPXTG motif (H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, in preparation).
25. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
26. We thank W. W. Navarre for help in determining the surface protein anchor structure of strain SM317, and D. Missiakas, P. Model, M. Russel, and members of our laboratory for critical reading of this manuscript. S.K.M. was supported by the Predoctoral Training Program in Genetic Mechanisms at UCLA (T32GM07104). H.T.-T. was supported by the Microbial Pathogenesis Training Grant at UCLA (AI 07323). Work in O.S.'s laboratory is supported by grant AI33987 from the National Institute of Allergy and Infectious Diseases.

6 May 1999; accepted 28 June 1999

Phosphorylation and Sequestration of Serotonin Transporters Differentially Modulated by Psychostimulants

Sammanda Ramamoorthy and Randy D. Blakely*

Many psychotropic drugs interfere with the reuptake of dopamine, norepinephrine, and serotonin. Transport capacity is regulated by kinase-linked pathways, particularly those involving protein kinase C (PKC), resulting in transporter phosphorylation and sequestration. Phosphorylation and sequestration of the serotonin transporter (SERT) were substantially impacted by ligand occupancy. Ligands that can permeate the transporter, such as serotonin or the amphetamines, prevented PKC-dependent SERT phosphorylation. Nontransported SERT antagonists such as cocaine and antidepressants were permissive for SERT phosphorylation but blocked serotonin effects. PKC-dependent SERT sequestration was also blocked by serotonin. These findings reveal activity-dependent modulation of neurotransmitter reuptake and identify previously unknown consequences of amphetamine, cocaine, and antidepressant action.

Serotonin [5-hydroxytryptamine (5-HT)] is a platelet-stored vasoconstrictor that also acts as a transmitter in the nervous system to modulate a wide spectrum of behaviors (*1*). The actions of 5-HT are terminated by active transport (*2*). Whereas 5-HT actions are mediated by >15 different types of receptors, a single 5-HT transporter (SERT) is responsible for extracellular 5-HT clearance (*3*). SERT activity is blocked by cocaine and tricyclic antidepressants. Serotonin-selective reuptake inhibitors (SSRIs) like fluoxetine (Prozac) preferentially block SERTs and enhance serotonergic signaling in affective disorders (*2, 4*). The amphetamines are substrates for SERTs, as well as for dopamine (DA) and norepinephrine (NE) transporters (DATs and NETs, respectively) (*5*) and can trigger SERT-mediated release of 5-HT (*5–8*). Repeated administration of amphetamines sensitizes monoaminergic synapses to subsequent psychostimulant challenge (*9*), which may involve modulated protein kinase cascades (*10*). Alterations in SERT activity and binding site density (*11*) and SERT gene polymorphisms (*12*) have implicated the transporter in anxiety, depression, suicide, autism, and substance abuse. Recent findings with transgenic mice (*13*) support an important role for 5-HT and SERTs in the behavioral actions of cocaine and amphetamine.

SERT expression can be rapidly modulated by receptor stimulation, second messenger production, and kinase activation (*14–16*). Suppression of SERT activity accompanying protein kinase C (PKC) activation (*17*) arises from a loss of 5-HT uptake capacity (V_{max}). The loss in 5-HT uptake capacity correlates with a loss of surface-expressed SERTs (*17*), similar to the PKC modulation of homologous γ-aminobutyric acid (GABA), DA, and NE transporters (*18*). PKC activators and phosphatase inhibitors induce SERT phosphorylation (*19*) with a similar time course and kinase antagonist sensitivity as observed for changes in 5-HT transport.

We investigated whether the regulation of SERTs was influenced by transport and whether SERT ligands differentially influenced SERT regulation. Figure 1A shows that PKC-mediated SERT phosphorylation in transfected human embryonic kidney–293 (HEK-293) cells was substantially diminished if assayed in the presence of the transported neurotransmitter, 5-HT. SDS–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of immunoprecipitates from [^{32}P]PO_4-labeled cell extracts (*20*) revealed a three- to fivefold stimulation of human SERT (hSERT) phosphorylation after phorbol 12-myristate 13-acetate (β-PMA) application. This stimulation was abolished by coapplication of PKC antagonists. In the presence of 5-HT (1 μM), SERT phosphorylation triggered by phorbol esters was also substantially blunted. At low concentrations of β-PMA (for example, 10 nM), 5-HT essentially abolished phorbol ester–induced SERT labeling, At 200 nM β-PMA, where labeling of SERTs is maximal, we consistently achieved 40 to 60% inhibition of SERT phosphorylation at maximal concentrations of 5-HT (1 μM) with a median effective concentration (EC50) of 70 nM (Fig. 1B).

If the actions of 5-HT on SERT phosphorylation arise as a consequence of transport, then an intrinsic homeostatic loop might be present to link transporter expression to extracellular amine availability, and coincubation with SERT antagonists should block this effect. Indeed, the SERT-selective tricyclic antidepressant imipramine (1 μM) or the SSRIs paroxetine (1 μM) and citalopram (1 μM) blocked the ability of 5-HT to limit PKC-dependent SERT phosphorylation (Fig. 1C). This effect was selective for SERT antagonists, as neither the DAT inhibitor GBR-12909 nor the NET antagonist nisoxetine could affect the ability of 5-HT to blunt SERT phosphorylation (Fig. 1C). There are no known 5-HT receptor subtypes on HEK-293 cells, and to our knowledge, 5-HT does not induce acute changes in cyclic adenosine 5′-monophosphate, inositol trisphosphate, or intracellular Ca^{2+} levels in these cells. Moreover, the ability of 5-HT to diminish PKC-dependent SERT phosphorylation was not

Department of Pharmacology and Center for Molecular Neuroscience, School of Medicine, Vanderbilt University, Nashville, TN 37232–6420, USA.

*To whom correspondence should be addressed. E-mail: randy.blakely@mcmail.vanderbilt.edu

Week 7

Motility and Chemotaxis

Biofilms

mice. In addition, Sk has not been reported to be present on kidney tissue. Thus, the antigen recognised by the rat anti-C1300 antiserum is a normal tissue antigen of A mice present on brain and, to a lesser extent, on kidney, but not on liver, lung, muscle, spleen or testes and distinct from the alloantigens previously determined to be present on both C1300 tumour cells and A brain. In subsequent communications, this antigen will be referred to as mouse brain antigen-1 (MBA-1).

I thank Professor N. A. Mitchison and Dr M. C. Raff in whose laboratories these experiments were carried out. S.E.M. is supported by a US Public Health Service Grant.

SUE ELLEN MARTIN*

MRC Neuroimmunology Project,
Zoology Department,
University College,
Gower Street,
London WC1E 6BT, UK

Received November 21, 1973.

* Present address: Laboratory of Biochemical Genetics, National Heart and Lung Institute, National Institutes of Health, Bethesda, Maryland 20014

1 Boyse, E. A., and Old, L. J., *A. Rev. Genet.*, **3**, 269 (1969).
2 Boyse, E. A., Miyazawa, M., Aoki, T., and Old, L. J., *Proc. R. Soc.* **B170**, 175 (1968).
3 Raff, M. C., Nase, S., and Mitchison, N. A., *Nature*, **230**, 50 (1971).
4 Amano, T., Richelson, E., and Nirenberg, M., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 258 (1972).
5 Harris, A. J., and Dennin, M. J., *Science*, **167**, 1253 (1970).
6 Harris, A. J., Heinemann, S., Schubert, D., and Tarikas, H., *Nature*, **231**, 296 (1971).
7 Blume, A. J., Gilbert, F., Wilson, S., Farber, J., Rosemberg, R., and Nirenberg, M., *Proc. natn. Acad. Sci. U.S.A.*, **67**, 786 (1970).
8 Nelson, P. G., Peacock, J. H., Amano, T., and Minna, J., *J. cell. Physiol.*, **77**, 337 (1971).
9 Schubert, D., Tarikas, H., Harris, A. J., and Heinemann, S., *Nature new Biol.*, **233**, 79 (1971).
10 Schachner, M., *Nature new Biol.*, **243**, 117 (1973).
11 Stockert, E., Old, L. J., and Boyse, E. A., *J. exp. Med.*, **133**, 1334 (1971).

Flagellar rotation and the mechanism of bacterial motility

BACTERIAL flagella are generally composed of three morphologically distinguishable regions: (*a*) the long flagellar filament which accounts for more than 95% of the flagellar protein; (*b*) the hook, which is generally 80–90 nm long and has a characteristic shape, and (*c*) the basal structure which is composed of an intricate set of disks and rods attaching the hook to the cell membrane and cell wall[1–3].

Explanations of how the flagella move[4] include the suggestion that helical waves are propagated along the filament[5,6], or that the filament behaves as a semrigid helical rotor[4,7,8]. Berg and Anderson[9] concluded that the evidence ". . . favours a model in which each filament rotates". The data presented here adds strong support to the existing evidence and provides a way to follow flagellar function that is independent of translational motility.

The methods that have been used to study motility depend on observation of the behaviour of the whole organism. This is a complex result of a series of events and is thus, at times, difficult to analyse and interpret. It would be useful to be able to follow the motion of a single flagellum or to assay for its activity in a way that does not depend on chemataxis or cell motility. To do this, a series of experiments was designed based on the observation that bacteria can be found which appear to be 'tethered' by their flagellum to a particle

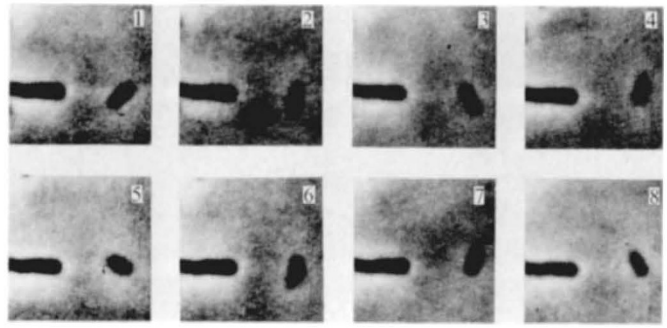


FIG. 1 The rotation of a cell bound to the microscope slide. *E. coli* strain MS1381 was grown on minimal medium with glycerol as the carbon source. Cells were put on a microscope slide and an equal volume of antipolyhook antibody diluted to 1:200 was added. After a 10 m lag, the cells began to spin. Their behaviour was recorded through a Zeiss phase contrast microscope onto video tape and then transferred to film. The pictures presented represent frames taken at intervals of 1/12 of a second. The cell at the left did not move during the sequence and it provides a reference point.

of debris or to the surface of the microscope slide[7]. These bacteria rotate rapidly. The rotation could reflect the motion of the tethered flagellum or it could result because the remaining untethered flagella cause the cell to move. To examine this motion in greater detail, we used 'polyhook' mutants of *E. coli*[10], which carry two mutations, one in the *hag* gene which eliminates the formation of the flagellar filament and another in the *flaE* gene which causes the loss of a function necessary to terminate the assembly of the hook structure. These cells make continuous polyhooks which can be 1–2 μm long and they are nonmotile. However, when dilute anti-polyhook antibody is added to a suspension of cells, they form clumps and begin to rotate rapidly. Several situations have been observed: the bacteria seem to be trapped or attached to the slide; the bacteria are bound to each other in pairs and counter-rotate or one rotates while the other is stationary; large groups of bacteria are cross linked and rotate or move with a jerking type of motion. Ten to twenty per cent of the bacteria in a microscope field can be found to be rotating. Figures 1 and 2 illustrate the first two situations. In Fig. 1, the cell seems to be attached to the surface of the microscope slide. It rotates 360° in a counterclockwise direction around an axis through one end of the cell body. The speed of rotation of an individual cell varies from two to nine revolutions per second. The cells can be observed to modulate their spinning in three ways: (*a*) by continuing to spin counterclockwise; (*b*) by stopping and then restarting in the same direction, or (*c*) by changing direction and spinning clockwise. The cells generally spin

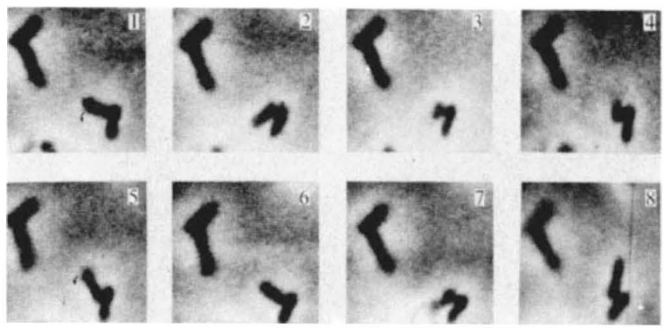


FIG. 2 The rotation of one cell bound to another. The preparation was the same as that described in the legend to Fig. 1. The cells in the upper left corner serve as reference.

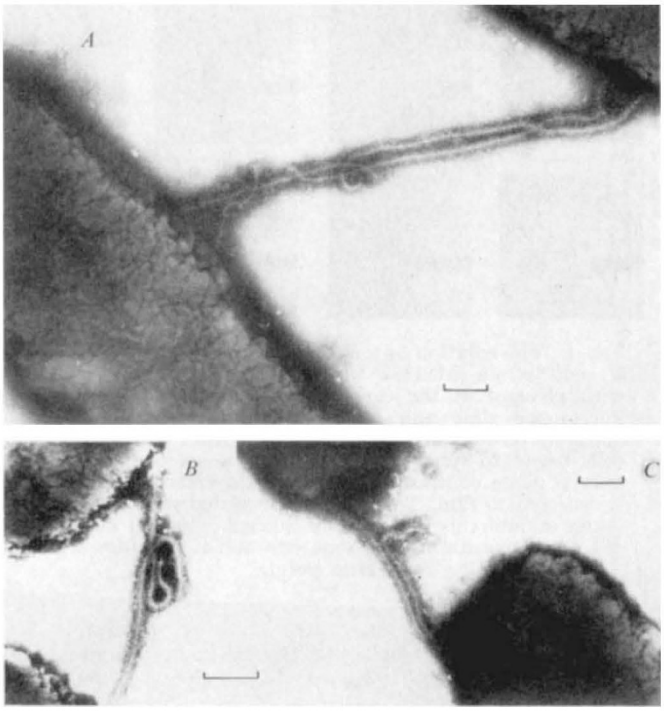


FIG. 3 Electron microscopy of cells bound together via their polyhooks. The cells were prepared as described in Fig. 1. They were placed on copper grids covered with carbon coated formover and stained with 0.5% phosphotungstic acid. The polyhook structure generally appears to be helical with a diameter of about 16 nm (ref. 10). These, however, are straightened and thickened as a result of the antibody molecules bound to them. The line in the photograph represents 0.2 nm. *A*, two cells held together by antibody bound to three polyhooks, *B* and *C*, cells held together by antibody bound to polyhooks derived from each of them.

counterclockwise and occasionally stop or reverse direction for one or two turns and then resume spinning counterclockwise. Individual cells tethered in this way have been observed to spin with intermittent short pauses for more than 30 m. Figure 2 illustrates the motion observed with a pair of tethered bacteria. Again the observations are most consistently interpreted as resulting from the complete rotation of one of the cells about a point of attachment to the other cell. The attachment appears as a narrow gap between the cells (Fig. 2, frame 4 and frame 8) which could correspond to the polyhook-antibody complex.

Thus, it is clear that the cells can rotate. However, it is possible that they rotate around the point of attachment, using the antibody as a swivel rather than rotating around the point of insertion of the basal structure into the cell body. Figure 3 shows electron micrographs of the tethered cells. It is clear that the polyhook structures are coated with antibody molecules, they bind the structures together at many points. The polyhooks have lost some of their characteristic helical appearance and are straightened, suggesting that they have some structural flexibility. The multiple attachment points along the surface of the polyhook appear to stitch them together and make it seem highly unlikely that the cells could rotate around the attachment site.

While these experiments were all done using the polyhook mutant, similar experiments were possible with bacteria that had flagellar filaments. *E. coli* strain W3110 which we used carries a mutation in the *hag* gene that results in the synthesis of straight flagellar filaments lacking the characteristic helical appearance. The cells are nonmotile. Nonetheless, indirect evidence suggests that the flagella are active[11]. To decrease the number of flagella per cell so that they would not tangle and crosslink, the bacteria were grown in minimal medium with glucose as a carbon source[12,13]. When dilute antiflagellar filament antibody was added to these cells they were tethered to the glass and to each other. They began to rotate in the same manner as the polyhook mutants. Thus, even with the straight flagellar filament the cell rotation is observed.

All these observations are concerned with the rotation of the cell, and the rotation of the filament is inferred. To observe the activity of the filaments themselves, we used small latex beads 0.7 μm in diameter (Dow), which were incubated with a 1:1,000 dilution of anti-flagellar antibody and then washed in 0.01 M Tris buffer, *p*H 7.8, and 0.1 M sodium chloride. They were added to suspensions of *E. coli* W3110 that had been grown in minimal medium with glucose. Beads were found attached 1–2 μm from the bacterium and they rotated very rapidly. In the same way, beads could be attached to flagella on wild-type bacteria and these were again observed to rotate rapidly, trailing behind motile bacteria.

All these observations, taken together, can be explained best if the hook is driven in a rotary fashion, probably by a mechanism anchored to the cell body at the base of the flagellum. This results in the rotation of the flagellar filament. Furthermore, the cell has the capacity to vary the speed and direction of rotation as well as the frequency of stopping and restarting. The ability to modulate the rotation of flagella may be the basis for the mechanism of chemotaxis.

This work was supported by a grant from the US National Science Foundation.

MICHAEL SILVERMAN*
MELVIN SIMON

Department of Biology,
University of California, San Diego,
La Jolla, California 92037

*Present address: Department of Pathology, University of Colorado School of Medicine, Denver, Colorado 80220.

Received December 27, 1973.

1 Dimmitt, K., and Simon, M., *J. Bact.*, **105**, 369–375 (1971).
2 Depamphilis, M., and Adler, J., *J. Bact.*, **105**, 376–383 (1971).
3 Abram, D., Koffler, H., and Vatter, A., *J. Bact.*, **90**, 1337–1354 (1965).
4 Doetsch, R. N., and Hageage, G. F., *Biol. Rev.*, **43**, 317–362 (1968).
5 Doetsch, R. N., *J. theor. Biol.*, **11**, 411–421 (1966).
6 Klug, A., *Symp. int. Soc. Cell Biol.*, **6**, 18–39 (1967).
7 Stocker, B. A. D., *Symp. Soc gen. Microbiol.*, **6**, 19–40 (1956).
8 Iino, T., *Behaviour of Micro-organisms*, Proc. Tenth int. Cong. Microbiol., Mexico City, 1970 (edit. by Pérez-Miravete, A.), 205–213 (Plenum, London, New York, 1973).
9 Berg, H., and Anderson, R., *Nature*, **245**, 380–383 (1973).
10 Silverman, M., and Simon, M. I., *J. Bact.*, **112**, 986–993 (1972).
11 Raimondo, L. M., Lundh, N. P., and Martinez, R. J., *J. Virol.*, **2**, 256–264 (1968).
12 Adler, J., and Tempelton, B., *J. gen. Microbiol.*, **46**, 175–184 (1967).
13 Yokota, T., and Gots, J., *J. Bact.*, **103**, 513–516 (1970).

Change in direction of flagellar rotation is the basis of the chemotactic response in *Escherichia coli*

BERG and Anderson[1] recently argued from existing evidence that bacteria swim by rotation of their helical flagella. Silverman and Simon[2] have now provided a clear demonstration of this. By means of antibodies specific for flagellar components, they tethered cells to microscope slides or to each other and observed rotation of the cell bodies. The cells were able to stop and to rotate in either direction. It seemed possible, as they proposed[2], that cessation or reversal of flagellar rotation might be involved in bacterial chemotaxis. Accordingly, we used wild-type and chemotaxis-defective mutant cells of

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 5, pp. 2390–2394, May 1979
Genetics

Interaction of the *cheC* and *cheZ* gene products is required for chemotactic behavior in *Escherichia coli*

(reversion analysis/functional suppression/flagellar rotation/sensory transduction)

JOHN S. PARKINSON AND STEPHEN R. PARKER

Department of Biology, University of Utah, Salt Lake City, Utah 84112

Communicated by A. Dale Kaiser, February 22, 1979

ABSTRACT **Previous work has shown that the *cheC* gene product of *Escherichia coli* plays a key role in regulating the direction of flagellar rotation during chemotactic responses. An attempt was made to identify other stimulus transduction elements that interact with the *cheC* component by examining *cheC* revertants for functional suppressors. Approximately two-thirds of the revertants studied appeared to be due to back mutation or to second-site mutations near or within the *cheC* structural gene. The remainder of the revertants carried suppressor mutations that mapped at the *cheZ* locus. Half of these suppressors impaired chemotaxis in a *cheC*$^+$ background and were shown by complementation analysis to be defective in *cheZ* function. These suppressors corrected *cheC* defects in an allele-specific pattern, suggesting that the *cheC* and *cheZ* proteins are in direct contact and are mutually corrective due to protein–protein interaction. Observation of swimming patterns and flagellar rotation in *cheC cheZ* mutants demonstrated that the interaction of these two gene products influences both the spontaneous frequency of flagellar reversals and the ability of the rotational machinery to respond to chemotactic stimuli. A model of this interaction and its possible role in chemotaxis are discussed.**

Stimulus detection, signaling, and behavioral response are basic features of sensory transduction systems in both prokaryotes and eukaryotes. Bacterial chemotaxis is a useful model for exploring these events at the molecular level. In *Escherichia coli* chemotactic responses are initiated by specific receptors that monitor the organism's chemical environment as it swims (1). In the absence of stimuli, wild-type cells swim in a random walk pattern (2) consisting of smooth "runs" and abrupt directional changes or "tumbles", both of which are produced by rotation of the flagellar filaments (3–5): runs by counterclockwise rotation and tumbles by clockwise rotation. Upon detecting a change in attractant or repellent concentration (6), the chemoreceptors generate signals that modulate flagellar rotation to produce an appropriate locomotor response. When headed in a favorable direction, tumble probability decreases, and when headed in an unfavorable direction, tumble probability increases (2, 6).

Studies of nonchemotactic mutants have identified a number of gene products that might be components of the signaling system in *E. coli* (7). Although the functions of most chemotaxis genes are still poorly understood, the *cheC* gene appears to play a key role in the transmission of sensory information from receptors to flagella. *CheC* mutants are motile but nonchemotactic and in the absence of stimuli exhibit very little tumbling behavior (8, 9). These mutants are typically somewhat leaky and also partially dominant (9), indicating that they probably make an altered but still functional product rather than an inactive one. This product may be a component of the flagellum, because *cheC* mutants are not complemented by nonflagellate mutants defective in *flaA* function (10). Both *cheC* and *flaA* mutants probably arise by different sorts of mutations in the same gene: null defects appear to result in a nonflagellate condition (*flaA*), whereas more subtle changes seem to permit flagellar assembly, but interfere with proper rotational behavior (*cheC*). Thus, the *cheC* (*flaA*) gene product may be an essential structural component of the flagellum that is somehow involved in determining the direction of flagellar rotation. Studies of the residual chemotactic responses in *cheC* mutants (9) and in an analogous class of *Salmonella typhimurium* mutants (11) have led to the suggestion that the *cheC* product might interact directly with the signaling system of the chemoreceptors to effect changes in rotational behavior (7, 11).

It might be possible to identify signaling functions by virtue of their ability to interact with the *cheC* product. For example, many sorts of gene interactions can result in the suppression or modification of a mutant phenotype (12). It seemed likely that *cheC* mutants, because they owe their phenotype to a seemingly minor alteration of a flagellar protein, might be suppressed by correspondingly minor changes in interacting proteins. We therefore examined a large number of chemotactic revertants of *cheC* strains to determine whether *cheC* defects could be alleviated by mutations in other chemotaxis genes. In this report, we show that many *cheC* revertants in fact carry a compensating mutation at the *cheZ* locus, and that the *cheC* and *cheZ* gene products probably interact in a direct manner. This interaction affects both the spontaneous tumbling behavior of *E. coli* and the ability of the rotational machinery to respond to chemoreceptor signals.

MATERIALS AND METHODS

E. coli K12 strains RP252 [F$^-$ *his trp* (*am*)] and RP477 (F$^-$ *thi thr leu his eda* Δ(*gal-att*λ) *strA*] and their *cheC* derivatives were used in this work. *cheC* mutations *181*, *182*, *183* (9), and *497* (13) were introduced into RP252 and RP477 by contransduction with the *his* locus. The *supD* marker employed in initial test crosses was derived from strain CR63 (14) and transferred into RP252 by selecting Trp$^+$ transductants and then testing for the ability to support the growth of *amber* mutants of phage λ. F′ strains for complementation analysis of *che* mutants have been described (9).

All other methods, including growth media, P1 transduction, and analysis of swimming behavior and flagellar rotation, have been described (9).

RESULTS

Isolation of *cheC* revertants containing external suppressors

Four different *cheC* mutations (alleles *181*, *182*, *183*, and *497*) were each introduced into strain RP252, and chemotactic revertants were selected by picking "swarms" on semisolid tryptone agar (15). For the identification of revertants in which

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

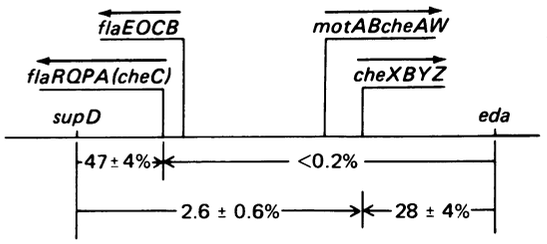


FIG. 1. Chemotaxis and flagellar genes in *cheC* region. This segment is located between minutes 41 and 43 in the *E. coli* chromosome (16) and contains genes for chemotaxis (*che*), motility (*mot*), and flagellar assembly (*fla*), not all of which are shown. Direction of transcription and extent of operons are indicated by arrows above the genes. Location of these operons relative to the outside markers *supD* and *eda* is shown approximately to scale. Below the map are P1 cotransduction frequencies (from ref. 9) for map intervals discussed in this work. Arrows point to the unselected marker in each cross.

the *cheC* defect had been corrected by mutations outside the *cheC* structural gene, we test-crossed each revertant to map the site responsible for restoring chemotaxis. In true revertants and in those with second-site mutations in the *cheC* gene, the site of reversion should map at the *cheC* locus, whereas, in revertants that carry suppressors outside the *cheC* gene, the reversion site (suppressor locus) might not map near the *cheC* gene. As shown in Fig. 1, the *cheC* gene is approximately 50% cotransducible with the *supD* locus. Thus, any reversion sites that show a significantly different linkage to *supD* should represent external suppressors of *cheC* (hereafter designated *scc* to denote suppressors of *cheC*).

Cotransduction frequencies between the site of reversion and the *supD* locus were determined for each revertant strain by means of the test cross shown in Fig. 2. Each revertant was infected with P1 phage grown on a *supD* derivative of the original *cheC* parent strain, and the proportion of *supD* transductants that were no longer chemotactic was measured. Because the donor strain also carries the original *cheC* mutation, nonchemotactic transductants should arise by inheriting both the *supD* (selected) marker and the (unselected) donor allele corresponding to the site of reversion in any particular revertant strain. The proportion of *supD* transductants that are nonchemotactic therefore provides a measure of the distance between the *supD* marker and the reversion site.

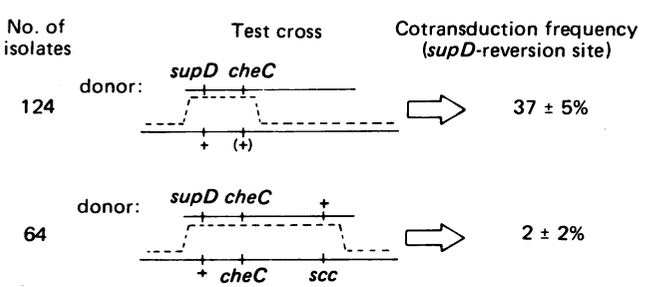


FIG. 2. Summary of test-cross results for mapping reversion sites in *cheC* revertants. Revertants of RP252 *cheC* strains were infected with P1 grown on a *supD cheC* donor, and Trp$^+$ (i.e , *supD*) transductants were selected and tested for chemotactic ability (50 transductants from each independent revertant). The revertant strains could be divided into two discrete groups based on the frequency of nonchemotactic segregants. The crossover diagrams indicate the probable genotype of each group and the exchanges required to generate a nonchemotactic recombinant. Those revertants giving a high frequency of nonchemotactic transductants appear to have reversion sites near or within the *cheC* gene (upper diagram); those yielding a low frequency of nonchemotactic transductants carry external suppressors (*scc* mutations) some distance from the *supD* and *cheC* loci (lower diagram).

Test-cross results for a sample of 188 independent revertants indicated that at least two types of revertants were obtained (Fig. 2). Approximately two-thirds of the strains had an average cotransduction frequency of 37 ± 5%. Subsequent crosses showed that in these strains the reversion site is tightly linked to the *cheC* locus, and it seems likely that many of these revertants arose by back mutation or by secondary mutations within the *cheC* gene. Because genes that specify interacting proteins are often located near one another it is conceivable that some of these revertants actually carry suppressor mutations in nearby *fla* genes (see map in Fig. 1). This might account for the fact that cotransduction frequencies between *supD* and the reversion sites in this group of strains were generally somewhat less than would be expected if the reversion events had occurred at the *cheC* locus.

The second group of revertants exhibited cotransduction values of 2 ± 2% in the test cross (Fig. 2) and clearly contain reversion sites (i.e., *scc* mutations) some distance from the *cheC* locus, but still linked to *supD*. Several clusters of *che* genes, which are loosely linked to *supD*, are cotransducible with the *eda* locus, whereas *cheC* is not (see Fig. 1). To determine whether the *scc* mutations in this group of revertants were located near these clusters, each mutation was tested for linkage to the *eda* locus. P1 lysates prepared on each *cheC scc* (*eda*$^+$) strain were used to transduce the *eda*$^+$ marker into *eda cheC* recipients derived from strain RP477, and the frequency of chemotactic transductants was measured. All 64 of the *scc* donor strains tested yielded chemotactic transductants in this cross (mean contransduction frequency of 20 ± 5%), indicating that the *scc* mutations are linked to *eda*, probably in the vicinity of the *cheX* operon (see Fig. 1).

When transferred in a similar manner into RP477 (*cheC*$^+$), half (32/64) of the *scc* mutations produced a partial or complete defect in chemotaxis, whereas the others had little or no effect on chemotactic ability. These two groups of *scc* mutations will be referred to as type I and type II, respectively. Complementation tests (performed with F′ elements carrying various *che* mutations) demonstrated that all of the type I mutations were defective in *cheZ* function. Although a definitive gene assignment for the type II mutations could not be made by complementation analysis, owing to lack of a suitable phenotype, it seems likely that these mutations are also alleles of the *cheZ* gene because type I and type II mutations have similar map positions, suppression patterns (see below), and possible modes of suppression (see below). The properties of *cheC* revertants carrying either type I or II *scc* mutations are summarized in Fig. 3 and discussed in the following sections. To simplify this discussion, we make the assumption that both groups of *scc* mutations represent alterations of *cheZ* function, and confine our attention to consideration of the CheC–CheZ interaction.

Effect of CheC–CheZ interaction on tumbling frequency

Mutants defective in *cheZ* function have very high tumbling rates (9, 17). In a *cheC*$^+$ background, type I *scc* mutations produced very high tumbling rates comparable to those of *cheZ* mutants; type II mutations also caused above normal tumbling rates, although generally not as high as in type I strains (data not shown). In combination with a *cheC* defect, which alone causes a very low tumbling rate, both types of *scc* mutations produced various tumbling frequencies (Fig. 3 *bottom*). As a general rule, revertants with type II mutations had somewhat lower tumbling frequencies than those with type I mutations, which suggests that the opposing tumbling defects caused by *cheC* and *scc* (*cheZ*) mutations may interact in a roughly ad-

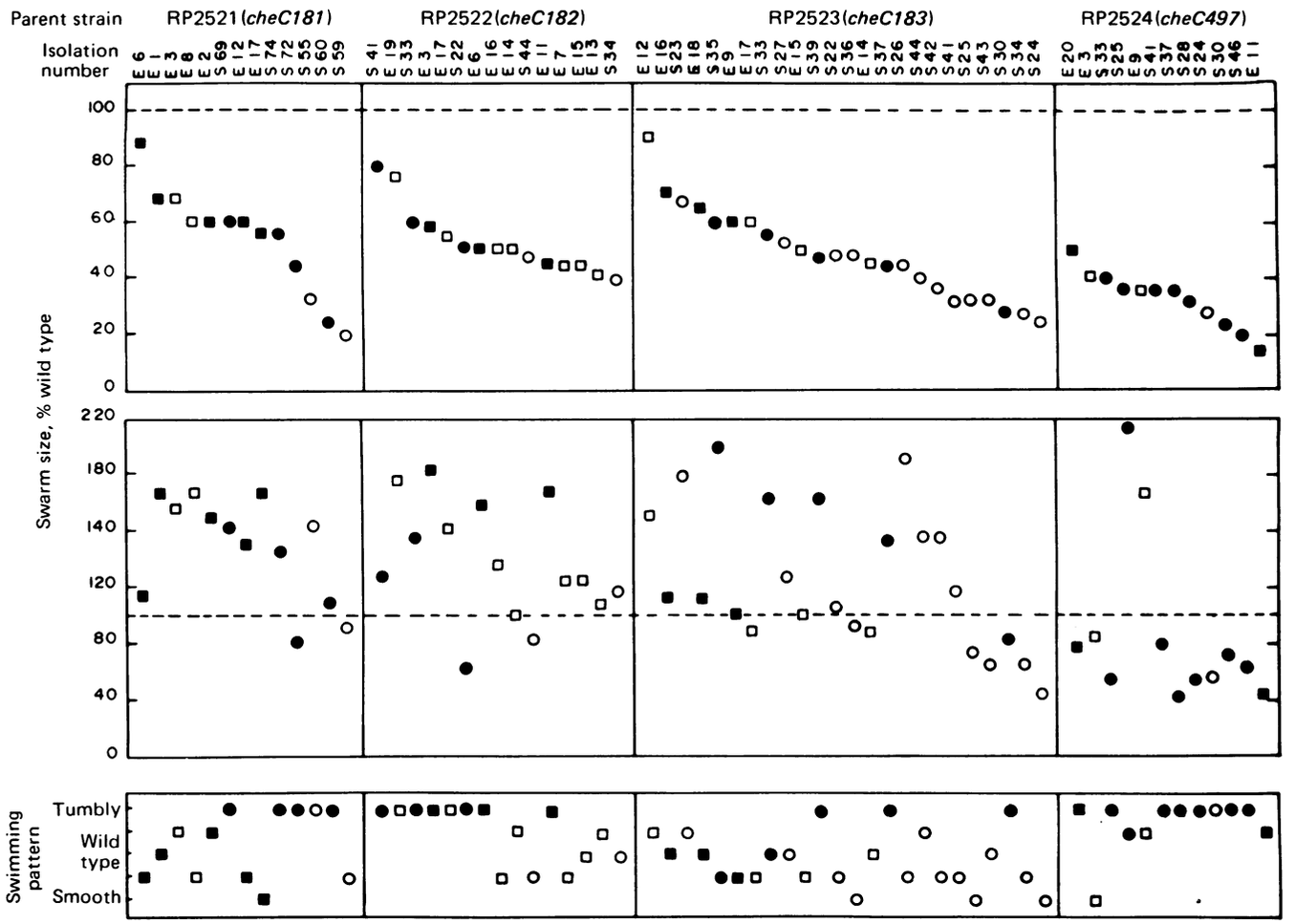


FIG. 3. Chemotactic behavior and swimming patterns of *cheC* revertants containing *scc* mutations. Chemotactic ability was assessed by measuring swarm diameters on semisolid tryptone agar after 9 hr at 35°C (*Top*) or after 17 hr at 24°C (*Middle*). Five colonies from each revertant were measured, and a wild-type control was included on each plate for normalization purposes. Variation within each strain was negligible (the height of each data point represents over two standard deviations in every case). (*Bottom*) Swimming behavior of log phase cells grown in tryptone broth at 35°C was evaluated by direct microscopic observation at room temperature (≈24°C). Each strain was assigned to one of five categories based on the frequency of tumbling: at one extreme, smooth strains (e.g., *cheC*) showed no tumbling; at the other extreme, tumbly strains (e.g., *cheZ*) showed constant tumbling. Closed symbols denote revertants carrying type I *scc* mutations; open symbols denote revertants carrying type II *scc* mutations (see text). Spontaneous revertants are indicated by circles; ethyl methanesulfonate-induced revertants are indicated by squares.

ditive fashion. Because the ability to tumble is essential for chemotaxis, this might account for the ability of *scc* mutations to suppress the chemotaxis defect of *cheC* strains. To test this notion, we examined chemotactic ability and patterns of flagellar rotation in a series of *cheC cheZ* double mutants in which the *cheZ* mutations had been derived directly from wild type rather than as suppressors of *cheC*. As shown in Fig. 4, these double mutants had rotational patterns intermediate between those of the component single mutants, which confirms that *cheC* and *cheZ* have an additive effect on tumbling behavior. However, none of these double mutants were chemotactic (data not shown), which indicates that restoration of a fairly normal tumbling rate is not a sufficient condition for chemotaxis in *cheC cheZ* strains. Comparison of chemotactic ability and swimming patterns in *cheC* revertants also demonstrates this point (Fig. 3): some revertants with wild-type tumbling rates were less chemotactic than some with very low or very high tumbling rates. That tumbling frequency and chemotactic ability are not necessarily related in the revertants implies that, in addition to setting the spontaneous tumbling rate, the CheC–CheZ interaction may influence the ability of the tumbling machinery to respond to chemoreceptor signals.

Allele-specificity of the CheC–CheZ interaction

As mentioned above, not all *cheZ* mutations are capable of suppressing the chemotaxis defect of *cheC* mutants, which shows that a specific sort of *cheZ* alteration is required. Moreover, absence of *cheZ* product evidently does not lead to suppression, because none of the *scc* alleles appear to be nonsense mutations (which would have exhibited apparent 100% linkage to *supD*, a nonsense suppressor, in the initial test crosses). These findings indicate that the *cheC* and *cheZ* products may interact directly and that only combinations that properly "fit" one another are capable of restoring chemotaxis. Examination of suppression efficiency in different *cheC scc* strains demonstrates that this is probably the case (Fig. 5). Twenty-five *scc* mutations were transferred to various *cheC* backgrounds, and chemotactic ability was determined by measuring swarm size on semisolid tryptone agar. Many of the suppressors (e.g., *scc-5*, *scc-6*) seemed to function in all four *cheC* backgrounds, which suggests that they are able to recognize and correct some aspect of the *cheC* defect common to all four mutant strains. Other suppressors, however, were able to distinguish between these *cheC* alleles. For example, *scc-12* works very well with *C182*

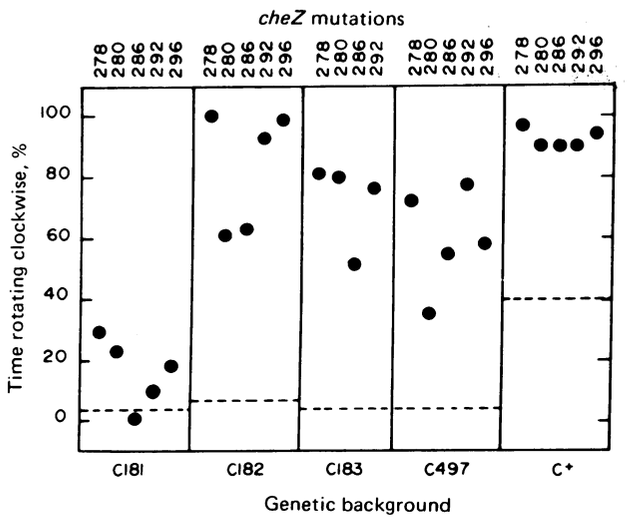


FIG. 4. Flagellar rotation patterns of *cheC cheZ* double mutants. Double mutants were constructed by introducing *cheZ* mutations into RP477 *cheC* recipients by P1 cotransduction with the *eda* locus. The *cheZ* alleles employed are described in refs. 9 and 17; alleles *286* and *292* are amber mutations. Since both the double mutants and the original recipients were nonchemotactic, doubles were identified by picking *eda*$^+$ transductants at random and backcrossing each to RP477 to test for the presence of an *eda*-linked *che* mutation. Strains were grown and tethered for rotational analysis as described in ref. 9. Each rotating cell was examined for 30 sec, and the proportions of time spent rotating clockwise and counterclockwise were measured. Each data point represents the average clockwise time for at least 20 cells. Dotted lines indicate the rotational behavior of the recipient *cheC* strains and RP477 (*cheC*$^+$).

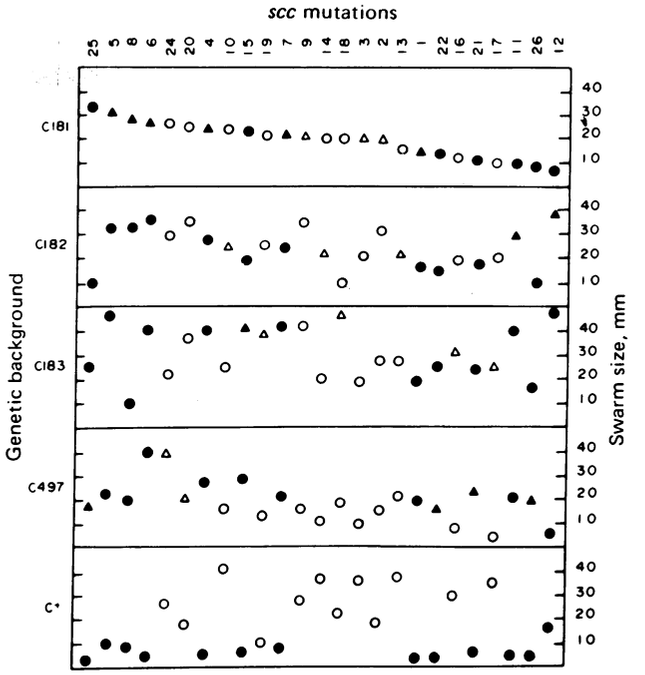


FIG. 5. Allele-specificity of *scc* mutations. Various *scc* mutations were introduced into RP477 (*cheC*$^+$) and RP477 *cheC* recipients by cotransduction with the *eda* locus. Double mutants that were nonchemotactic were confirmed by backcrosses as described in Fig. 4. Chemotactic ability was assessed by measuring swarm diameters as described in Fig. 3. Each data point represents the mean swarm size for a particular *cheC scc* combination; in all cases, SD was 2 mm or less. Closed symbols denote type I *scc* mutations; open symbols denote type II *scc* mutations (see text). The *cheC* mutation from which a particular suppressor was originally derived is indicated by a triangle. The set of suppressors has been ordered with respect to efficiency of suppression in the *cheC181* background to facilitate comparisons with the other *cheC* backgrounds.

and *C183*, but very poorly with *C181* and *C497*. Moreover, suppressors that behaved the same in one *cheC* background (e.g., *scc-12* and *scc-18* in *C183*) often behaved quite differently in another background (e.g., *C182*). In summary, the effect of an *scc* mutation on any particular *cheC* allele could not be predicted from its behavior in other *cheC* backgrounds, demonstrating that *scc* mutations act in an allele-specific fashion. The highly specific nature of this interaction implies that the *cheC* and *scc* (i.e., *cheZ*) gene products themselves are either transiently or permanently associated during the chemotaxis process.

Comparisons of chemotactic ability in *cheC scc* strains at 35°C and 24°C emphasize the specificity of the *cheC–cheZ* interaction (Fig. 3 *top* and *middle*). At 35°C, the temperature at which they were originally isolated, none of the suppressed revertants were as chemotactic as wild type, which demonstrates that the mutant products cannot function together as well as their wild-type counterparts. This implies that any protein interactions involved are probably less stable than in wild type. At 24°C, many of the revertant strains exhibited improved chemotactic ability, often even better than wild type. The lack of correlation between chemotactic ability at the two temperatures indicates that each combination of *cheC* and *scc* (*cheZ*) gene products responds to temperature changes in a unique way, which is consistent with the notion that these proteins are in direct contact.

DISCUSSION

In wild-type *E. coli*, spontaneous flagellar reversals occur about once per second (18), ensuring that, in spatial gradients of attractants or repellents, the organism's run length is sufficiently long to detect concentration differences before tumbling and yet short enough to prevent rotational diffusion from causing major course changes (19). Thus, changes in swimming direction are brought about by tumbling, and chemotaxis can be achieved by modulating the probability of flagellar reversal in response to stimuli. Like wild type, the flagella of *cheC* and *cheZ* mutants are capable of rotating in either direction (9). However, *cheC* mutants tend to remain in the counterclockwise (run) mode, whereas *cheZ* mutants rotate predominately in the clockwise (tumble) mode. Analysis of *cheC* revertants has demonstrated that certain combinations of *cheC* and *cheZ* defects can lead to restoration of chemotactic ability even though each mutation separately produces a nonchemotactic phenotype. This interaction probably involves direct contact between the *cheC* and *cheZ* proteins and generally leads to a tumbling rate that is intermediate between the very low and very high rates produced by the two component mutations separately.

A model of the CheC–CheZ interaction is shown in Fig. 6. We assume that the *cheC* product is a flagellar component, perhaps located in the basal body or the adjacent cytoplasmic membrane, that determines the direction of flagellar rotation. The *cheZ* product, which is known to be a cytoplasmic protein (20), appears to influence the pattern of flagellar rotation by binding to the *cheC* component. We suggest that transitions between clockwise and counterclockwise rotation are accompanied by the formation or dissociation of a CheC–CheZ complex: counterclockwise rotation by CheZ binding, and clockwise rotation by CheZ release.

In the absence of chemotactic stimuli, the relative affinities of the *cheC* and *cheZ* proteins probably play a major role in establishing the spontaneous tumble rate of the cell, and the

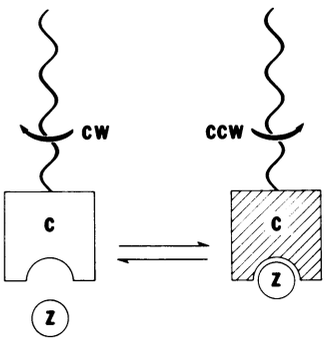


FIG. 6. Model of CheC–CheZ interaction. The *cheC* and *cheZ* proteins may interact in a reversible manner to control tumbling behavior. Additional features of the model are discussed in the text.

properties of *cheC* and *cheZ* mutants are consistent with this picture. For example, all *cheZ* mutants, including those with nonsense mutations, have high tumbling rates, which could reflect a decreased ability to bind to the *cheC* component. On the other hand, *cheC* mutants that have low tumbling rates and are partially dominant might have an increased affinity for *cheZ* protein. In such mutants tumbling should be raised to more normal levels by reducing the ability of *cheZ* protein to bind to *cheC*, which may account for the observation that *cheC* and *cheZ* mutations have a roughly additive effect on tumbling frequency. According to this model, there should exist tumbly *cheC* mutants, which have reduced affinity for *cheZ*, and nontumbling *cheZ* mutants, which have increased affinity for *cheC*. Although mutants of the latter type have not yet been observed, we have recently obtained *cheC* mutants with very high tumbling rates, and similar mutants have also been found in *S. typhimurium* (11).

What is the role of this interaction in chemotaxis? Clearly one consequence is to set the spontaneous tumbling rate of the cell; however this is not a sufficient condition for chemotaxis because *cheC scc* strains with similar tumbling rates often had very different chemotactic abilities, whereas those with similar chemotactic behavior often had very different tumbling rates (see Fig. 3). Moreover, mutations that appear to alter the relative affinities of the *cheC* and *cheZ* proteins can restore normal tumble frequencies but still preclude chemotaxis (see Fig. 4), suggesting that one or both of these proteins must participate in other processes necessary for chemotaxis. Because *cheZ* mutants still respond to chemotactic stimuli, although with high thresholds (21), it seems unlikely that *cheZ* product is responsible for *initiating* changes in flagellar rotation during chemoreceptor signaling, but it could be involved in *facilitating* or *maintaining* such changes.

Several lines of genetic evidence indicate that the *cheZ* product may also interact with another chemotaxis protein, the *cheB* product (7, 17, 22). Mutants defective in *cheB* function lack a protein methylesterase activity (23) that has been implicated in the process of sensory adaptation (24). It may be that *cheZ* protein, through its interaction with *cheB* product, somehow regulates the activity of the adaptation system and thereby controls the duration of chemotactic responses. For example, the *cheB* and *cheZ* proteins might form a tight complex so that when the *cheZ* portion is bound to the *cheC* component the methylesterase is unable to reach its target sites. It should be possible to test this notion by further studies of the behavior, particularly the sensory adaptation ability, of *cheC* and *cheZ* strains. By extending the sorts of genetic studies described in this report we may eventually be able to construct a detailed picture of the ways in which various elements of the chemotaxis machinery interact with one another to generate chemotactic behavior.

This work was supported by a U.S. Public Health Service Research Grant to J.S.P. from the National Institute of General Medical Sciences.

1. Adler, J. (1969) *Science* **166,** 1588–1597.
2. Berg, H. C. & Brown, D. A. (1972) *Nature (London)* **239,** 500–504.
3. Berg, H. C. & Anderson, R. A. (1973) *Nature (London)* **245,** 380–382.
4. Silverman, M. & Simon, M. (1974) *Nature (London)* **249,** 73–74.
5. Larsen, S. H., Reader, R. W., Kort, E. N., Tso, W.-W. & Adler, J. (1974) *Nature (London)* **249,** 74–77.
6. Macnab, R. W. & Koshland, D. R., Jr. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 2509–2512.
7. Parkinson, J. S. (1977) *Annu. Rev. Genetics* **11,** 397–414.
8. Armstrong, J. B., Adler, J. & Dahl, M. M. (1967) *J. Bacteriol.* **93,** 390–398.
9. Parkinson, J. S. (1976) *J. Bacteriol.* **126,** 758–770.
10. Silverman, M. & Simon, M. (1973) *J. Bacteriol.* **116,** 114–122.
11. Rubik, B. A. & Koshland, D. E., Jr. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2820–2824.
12. Hartman, P. E. & Roth, J. R. (1973) *Adv. Genet.* **17,** 1–105.
13. Armstrong, J. B. & Adler, J. (1969) *Genetics* **61,** 61–66.
14. Signer, E. R., Beckwith, J. R. & Brenner, S. (1965) *J. Mol. Biol.* **14,** 153–166.
15. Adler, J. (1966) *Science* **153,** 708–716.
16. Bachmann, B. L., Low, K. B. & Taylor, A. L. (1976) *Bacteriol. Rev.* **40,** 116–167.
17. Parkinson, J. S. (1978) *J. Bacteriol.* **135,** 45–53.
18. Berg, H. C. (1974) *Nature (London)* **249,** 77–79.
19. Brown, D. A. & Berg, H. C. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1388–1392.
20. Ridgway, H. F., Silverman, M. & Simon, M. (1977) *J. Bacteriol.* **132,** 657–665.
21. Parkinson, J. S. (1974) *Nature (London)* **252,** 317–319.
22. DeFranco, A. L., Parkinson, J. S. & Koshland, D. E., Jr. (1979) *J. Bacteriol.*, in press.
23. Stock, J. B. & Koshland, D. E., Jr. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3659–3663.
24. Goy, M. F., Springer, M. S. & Adler, J. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4964–4968.

Molecular Microbiology (1998) **30**(2), 295–304

Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development

George A. O'Toole and Roberto Kolter*
Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA.

Summary

The formation of complex bacterial communities known as biofilms begins with the interaction of planktonic cells with a surface in response to appropriate environmental signals. We report the isolation and characterization of mutants of *Pseudomonas aeruginosa* PA14 defective in the initiation of biofilm formation on an abiotic surface, polyvinylchloride (PVC) plastic. These mutants are designated surface attachment defective (*sad*). Two classes of *sad* mutants were analysed: (i) mutants defective in flagellar-mediated motility and (ii) mutants defective in biogenesis of the polar-localized type IV pili. We followed the development of the biofilm formed by the wild type over 8 h using phase-contrast microscopy. The wild-type strain first formed a monolayer of cells on the abiotic surface, followed by the appearance of microcolonies that were dispersed throughout the monolayer of cells. Using time-lapse microscopy, we present evidence that microcolonies form by aggregation of cells present in the monolayer. As observed with the wild type, strains with mutations in genes required for the synthesis of type IV pili formed a monolayer of cells on the PVC plastic. However, in contrast to the wild-type strain, the type IV pili mutants did not develop microcolonies over the course of the experiments, suggesting that these structures play an important role in microcolony formation. Very few cells of a non-motile strain (carrying a mutation in *flgK*) attached to PVC even after 8 h of incubation, suggesting a role for flagella and/or motility in the initial cell-to-surface interactions. The phenotype of these mutants thus allows us to initiate the dissection of the developmental pathway leading to biofilm formation.

Introduction

Biofilms are sessile bacterial communities adhered to a surface. In most environments, bacteria are thought to reside predominantly in biofilms (Costerton *et al*., 1995), in contrast to planktonic or free-swimming cells typically studied in the laboratory. *Pseudomonas aeruginosa* has been shown to form biofilms on a number of surfaces, including the tissues of the cystic fibrosis lung (Govan and Deretic, 1996) and on abiotic surfaces such as contact lenses and catheter lines (Nickel *et al*., 1985; 1989; Miller and Ahearn, 1987; Fletcher *et al*., 1993). This ubiquitous organism is also the cause of nosocomial infections in immunocompromised patients and individuals with severe burns (Bodey *et al*., 1983).

Biofilms of *P. aeruginosa* (and other microorganisms) are formed from individual planktonic cells in a complex and presumably highly regulated developmental process. Planktonic cells are thought to initiate interactions with a surface in response to various signals, including the nutritional status of the environment (Wimpenny and Colasanti, 1997; O'Toole and Kolter, 1998; Pratt and Kolter, 1998). The fully developed surface-attached community can be a highly structured with distinct architectural and physical/chemical properties (Costerton *et al*., 1995). These biofilm-grown cells are thought to be markedly different from their planktonic counterparts, based on a number of lines of evidence. For example, *P. aeruginosa* growing on a surface has increased expression of *algC*, a gene required for the synthesis of extracellular polysaccharides (Davies *et al*., 1993; Davies and Geesey, 1995). Biofilm-grown *P. aeruginosa* has also been shown to acquire increased resistance to antibiotics (Hoyle and Costerton, 1991). To date, most studies of biofilms have focused on characterizing the organisms that comprise these bacterial communities, the physical/chemical properties of biofilms and the physical forces that effect the bacterium's initial interactions with a surface (Costerton *et al*., 1995; Fletcher, 1996). However, little is known about the molecular genetic mechanisms regulating the initiation of biofilm formation and the development of these complex bacterial communities.

We present the isolation of mutants of *P. aeruginosa* PA14 defective in the initiation of biofilm formation on an abiotic surface. We report the molecular characterization of two classes of mutants defective in initiation of biofilm formation: (i) mutants defective in flagellar-mediated motility; and (ii) mutants defective in type IV pili biogenesis. The analysis of the flagellar and type IV pili mutants with time-lapse phase-contrast microscopy has allowed us to begin the dissection of the early development of a *P. aeruginosa* biofilm.

Received 24 March, 1998; revised 10 July, 1998; accepted 20 July, 1998. *For correspondence. E-mail kolter@mbcrr.harvard.edu; Tel. (617) 432 1776; Fax (617) 738 7664.

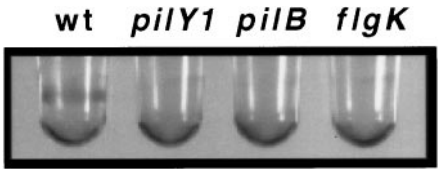


Fig. 1. Biofilm formation phenotype. Shown is the biofilm formation phenotype of the wild-type strain and three representative *sad* mutants (*pilY1*, *pilB* and *flgK*). As described previously, the biofilms form at the interface between the air and the medium (O'Toole and Kolter, 1998). Under the growth conditions used in this experiment, the only electron acceptor available is oxygen. Therefore, the biofilm forms only where oxygen levels are highest, that is at the interface between air and medium.

Results

Isolation of mutants defective in biofilm formation

We generated a collection of ≈2400 random transposon mutants of *P. aeruginosa* PA14 using the transposon Tn5-B30(Tc^r) (Simon *et al*., 1989). This collection of *P. aeruginosa* mutants was screened in microtitre dishes made of polyvinylchloride (PVC) to test for their ability to form a biofilm on an abiotic surface. The cells were allowed to grow in the wells of the microtitre dishes in a minimal M63 medium supplemented with glucose and casamino acids (CAA), using a technique described previously (O'Toole and Kolter, 1998), to assess their ability to form a biofilm. The biofilm was detected by staining with crystal violet (CV), a purple dye that stains the bacterial cells but does not stain the PVC plastic. After addition of CV and incubation at room temperature for ≈10 min, excess CV and unattached cells were removed by vigorous and repeated washing of the microtitre plates with water. An example of the phenotype of the wild-type strain is shown in Fig. 1. The biofilm is observed as a ring of CV-stained cells that forms at the interface between air and medium. Of the ≈2400 mutants screened, 15 mutants (0.5%) unable to form such a biofilm were isolated. These mutants were designated surface attachment defective or *sad*. The biofilm formation phenotypes of representative *sad* mutants are shown in Fig. 1.

Any strains exhibiting poor growth under these screening conditions might give the same phenotype as mutants unable to initiate formation of a biofilm. Therefore, all of the putative *sad* mutants were grown in liquid minimal M63 medium supplemented with glucose and CAA (the same medium used to screen for mutants). Of the 15 putative *sad* mutants tested, 13 grew as well as the wild-type strain but were unable to form a biofilm. The other two putative *sad* mutants had severe growth defects relative to the wild type and were not analysed further.

We performed Southern blot analysis of the 13 *sad* mutants that did not form a biofilm to determine the number of transposon insertions in each strain. A PCR-generated DNA fragment from the IS50 of Tn*5* was used to probe *Eco*RI-digested chromosomal DNA (there are no *Eco*RI sites in Tn*5*-B30). This analysis revealed a single hybridizing band for each strain, consistent with each *sad* mutant having only a single transposon insertion (data not shown). The further analyses of two classes of mutants (totalling 8 of 13) isolated in this screen is presented below. The analyses of the other five *sad* mutants will be presented elsewhere.

We reported previously that mutants of *P. fluorescens* originally identified as unable to initiate biofilm formation on PVC were also defective for biofilm formation on a variety of other abiotic surface (O'Toole and Kolter, 1998). We tested the *P. aeruginosa sad* mutants for their ability to form a biofilm on abiotic surfaces other than PVC, including polystyrene, polycarbonate and polypropylene. The wild-type strain can form a biofilm on all of these surfaces. All of the *sad* mutants originally isolated on PVC were also defective for biofilm formation on these other surfaces (data not shown).

Non-motile mutants are defective in biofilm formation

It has been reported that motility is required for both biofilm formation (on biotic and abiotic surfaces) and pathogenesis (Montie *et al*., 1982; de Weger *et al*., 1987; Grant *et al*., 1993; Korber *et al*., 1994; Simpson *et al*., 1995). However, these studies generally involved the analysis of molecularly uncharacterized non-motile mutants. Therefore, in addition to the phenotypic analyses described above, all *sad* mutants were assessed for their motility phenotype on 0.3% agar (minimal M63 medium supplemented with glucose and CAA). Of the 13 mutants tested, three strains (*sad-36*, *sad-39* and *sad-42*) were found to be non-motile (see Fig. 2). In a typical experiment after 24 h of growth at

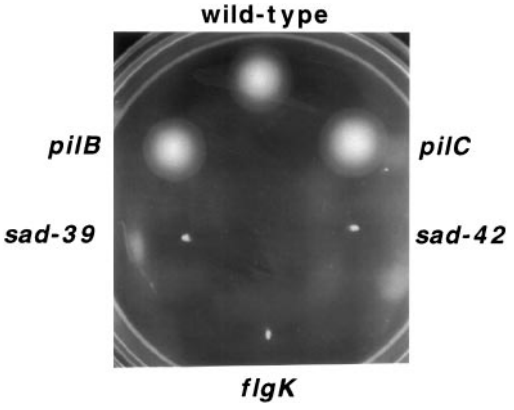


Fig. 2. Motility assays. The flagella-mediated motility of the wild-type strain, representative pili-defective mutants (*pilB* and *pilC*) and non-motile mutants (*flgK*, *sad-39*, and *sad-42*) was assessed on minimal M63 glucose/CAA medium with 0.3% agar after ≈24 h of growth at 25°C. Migration of the cells from the point of inoculation (observed as a turbid zone) indicates that the strain is proficient for flagellar-mediated motility.

room temperature, the wild type and two representative mutants defective in pili biogenesis (*pilB* and *pilC*) clearly migrated from the point of inoculation, whereas the *sad-36*, *sad-39*, and *sad-42* strains did not. One of these mutants, *sad-36*, was chosen for further analysis.

The *sad-36*::Tn*5*(Tc^r) insertion was mobilized into a wild-type genetic background by phage SN-T-mediated transduction as reported (Jensen *et al*., 1998). Eighteen out of 18 Tc^r transductants (indicating inheritance of the Tn*5* element) were non-motile and unable to make a biofilm, demonstrating that the single insertion in this strain was responsible for the observed phenotypes. The DNA sequence flanking the Tn*5* insertion in *sad-36* was determined using the arbitrary PCR method (see *Experimental procedures*) and compared with the GenBank database with BLASTX (Altschul *et al*., 1990). BLASTX translates DNA sequence in all six reading frames and compares these predicted protein sequences with GenBank. The determined DNA sequence flanking the Tn*5* element (≈375 nt), when translated, revealed a partial ORF with ≈40% identity and ≈65% similarity to HAP1 (*flgK*), the flagellar-associated hook protein 1 of *Salmonella typhimurium* and *Escherichia coli*. Mutations in the *flgK* locus in these organisms results in the synthesis of an incomplete flagellum, which renders the strains non-motile (Homma *et al*., 1990). The localization of the Tn*5* insert of the strain carrying the *sad-36* allele to a gene required for flagellar function is consistent with the non-motile phenotype of this strain.

Type IV pili are required for biofilm formation

We analysed the DNA sequence flanking the transposon inserts of the other *sad* mutants (as described in the section above and in the *Experimental procedures*). Comparison of the translated DNA sequence flanking the Tn*5* element of a number of *sad* mutants to the GenBank database revealed that five strains carried mutations in genes required for the synthesis of type IV pili.

Type IV pili have been shown to be important for the adherence to and colonization of eukaryotic cell surfaces and are thought to play a role in pathogenesis (Woods *et al*., 1980; Ramphal *et al*., 1984; Doig *et al*., 1988; Bieber *et al*., 1998). Four of the five mutants defective in type IV pili biogenesis identified in the screen had mutations in the *pilBCD* operon, which is thought to code for accessory factors required for pili assembly and function (Nunn *et al*., 1990). The strains carrying alleles *sad-31*, *sad-33* and *sad-34* have mutations in the *pilB* gene. The DNA sequence flanking the transposon insertions in *sad-33* and *sad-34* was identical, indicating that these two strains were probably siblings. The mutations carried in *sad-31* and *sad-33/sad-34* map to two different locations within *pilB* (data not shown). The strain carrying allele *sad-29* has a mutation in the *pilC* gene. Because the *pilBCD* locus may form an operon, it is possible that polarity onto *pilD* is actually causing the phenotype. However, it has been shown in *P. aeruginosa* PAO1 that mutations in any of these loci result in the loss of the synthesis of pili as indicated by resistance to the pilus-specific bacteriophage PO4 and visual inspection using electron microscopy (Nunn *et al*., 1990). The fifth mutant, *sad-25*, maps to yet a third locus, a homologue of the *pilY1* gene of *P. aeruginosa* PAO1. In *P. aeruginosa*, the *pilY1* gene is in a cluster of genes (including *pilV*, *pilW*, *pilX*, *pilY2*, and *pilE*) that is required for type IV pili biogenesis (Russell and Darzins, 1994; Alm and Mattick, 1995; Alm *et al*., 1996). Mutations in the *pilY* homologues in *Neisseria* spp. (called *pilC* in these organsims) can also result in a non-piliated phenotype (Jonsson *et al*., 1991; Rudel *et al*., 1992). Consistent with the mapping of these mutations to genes required for type IV pili biogenesis was their resistance to lysis by phage F116 (Pemberton, 1973), which uses type IV pili as its receptor (data not shown).

It has been shown that type IV pili are required for a form of surface-associated movement known as twitching motility. Twitching motility is thought to be a consequence of the extension and retraction of type IV pili, which propels the bacteria across a surface by an undescribed mechanism (Bradley, 1980; Whitchurch *et al*., 1990; Darzins, 1994). We assessed the twitching motility phenotype of the mutants carrying alleles *sad-25* (*pilY1*), *sad-29* (*pilC*), *sad-31* (*pilB*), and *sad-33* (*pilB*). The wild-type, a representative flagellar mutant (*flgK*), and four type IV pili mutants are shown in Fig. 3A. In addition to forming a colony on the surface of the agar plate (1.5% agar), Twitch$^+$ strains of *P. aeruginosa* PA14 form a haze of growth that surrounds the point of inoculation (Bradley, 1980; Whitchurch *et al*., 1990). This assay differs from the test for flagella-mediated motility, which is performed by inoculating cells onto 0.3% agar plates (see Fig. 2). Furthermore, strains capable of twitching motility have a spreading colony morphology, whereas strains defective in twitching motility produce rounded colonies (Whitchurch *et al*., 1990; Darzins, 1994). This difference in colony shape can also be observed in Fig. 3A.

Twitching motility can also be assessed by phase-contrast microscopy. At the microscopic level, the edge of the colonies of strains proficient in twitching motility are highly irregular. This is thought to be a consequence of the surface movement associated with type IV pili (Whitchurch *et al*., 1990; Darzins, 1994). Mutants lacking functional type IV pili have smooth-edged colonies. To further confirm that our strains did not have functional type IV pili, we observed the edges of wild-type and pili-deficient mutants using phase-contrast microscopy. As shown in Fig. 3B, the wild-type strain has the expected irregular colony edge and the representative pili-deficient strain (*sad-31*/*pilB*) has the expected smooth colony edge

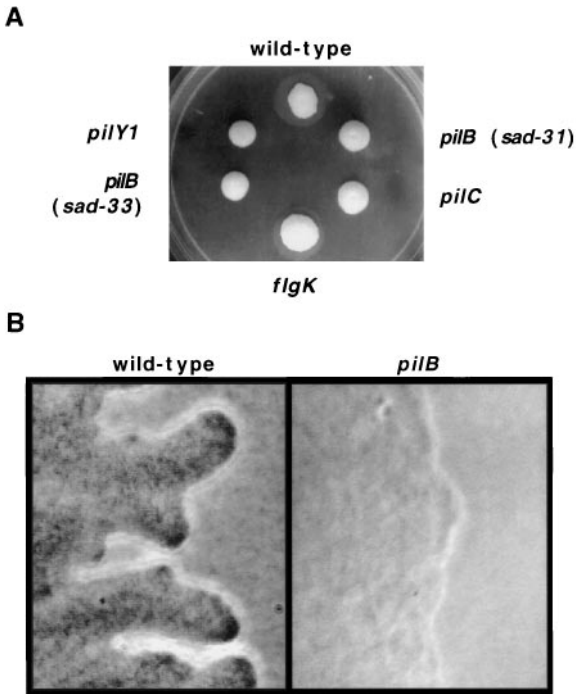


Fig. 3. Twitching motility. Twitching motility was assessed in two ways.
A. Cells were stabbed into an LB agar plate (1.5% agar) with a toothpick, incubated overnight at 37°C, then for 1–2 days at room temperature (≈25°C). Twitch$^+$ strains form a colony on the agar surface and form a hazy zone of cell growth within the agar substrate. Twitch$^-$ strains still form a colony on the agar, but lack the zone of growth within the agar. Also, the colonies of Twitch$^+$ strains are flat, spreading and irregularly shaped, whereas the colonies formed by strains defective in the synthesis of type IV pili are rounded and somewhat dome shaped (Mattick *et al.*, 1996; Darzins and Russell, 1997).
B. Direct visual inspection of the colony edges of wild type and pili mutants was performed using phase-contrast microscopy. The edge of the wild-type strain is highly irregular. In contrast, the colony edge of the representative pili mutants (*pilB*) has a smooth, regular phenotype. Micrographs were taken at 400× magnification.

phenotype. All the pili-defective mutants behaved in a fashion identical to *sad-31* (results not shown). Transmission electron-microscopic analysis of the pili mutants confirmed the lack of these structures on the surface of the mutant cells (not shown).

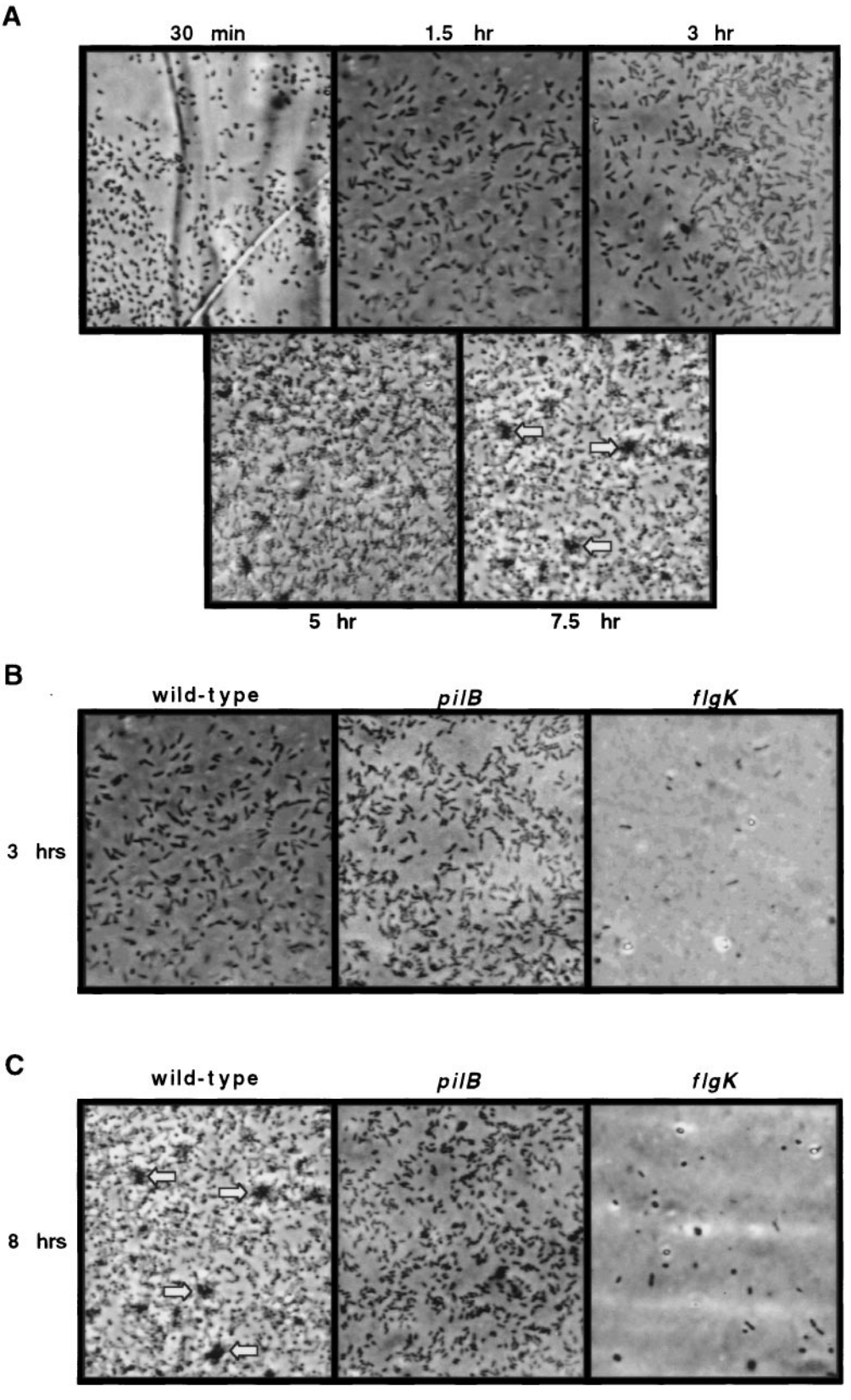
Mutants defective in flagellar-mediated motility and type IV pili biogenesis define two steps in a developmental pathway

We used the *sad* mutants isolated in this study as tools to initiate the dissection of the early steps in biofilm formation. To follow the initiation of biofilm formation by the wild-type and *sad* mutants, we directly visualized the formation of the biofilm on PVC using phase-contrast microscopy. A small tab of PVC plastic (≈3 mm × ≈6 mm) was incubated in the well of a microtitre dish that has been inoculated with $\approx 10^6$ cfu ml^{-1} of the appropriate strain in minimal M63 medium supplemented with glucose and CAA. After incubation for various times at 37°C, the plastic tab was removed from the microtitre dish with ethanol-sterilized forceps, rinsed with 1 ml of sterile minimal M63 medium and placed on a slide. The slide was examined using phase-contrast microscopy (400× magnification) as described in *Experimental procedures*.

Figure 4A shows a time-course of the development of a biofilm on PVC by the wild-type strain over 7.5 h at 37°C as observed using phase-contrast microscopy. As early as 30 min after inoculation, the wild type formed a dispersed monolayer of bacterial cells attached to the surface of the PVC plastic. A progressively more dense monolayer of cells formed on the surface over the next 3–4 h. By 5 h, and continuing until at least 7.5 h, this monolayer almost completely covered the PVC surface and became punctuated by microcolonies that are distributed across the surface of the PVC plastic and comprise multiple layers of cells. Typically, the wild-type microcolonies were ≈3–5 layers of cells thick.

We directly visualized the ability of the type IV pili-deficient and non-motile strains to form a biofilm on PVC using phase-contrast microscopy and compared their phenotypes with the wild-type strain. For the representative non-motile strain (carrying a mutation in *flgK*), few to no cells were observed attached the PVC plastic even after 8 h of incubation in the presence of the PVC surface. All other non-motile strains analysed had a phenotype identical to the *flgK* mutant. This observation is consistent with previous results showing the importance of motility in biofilm formation (Montie *et al.*, 1982; de Weger *et al.*,

Fig. 4. Direct observation of biofilm formation on PVC using phase-contrast microscopy.
A. Direct visualization of the initiation of biofilm formation over 7.5 h in our model system. Incubation of *P. aeruginosa* for up to 3 h in the presence of PVC results in the attachment of cells in a monolayer on this surface. At later time points (5 and 7.5 h), microcolonies can be observed dispersed throughout the monolayer (microcolonies are indicated by arrows).
B. Shown are phase-contrast micrographs of the wild-type strain, a representative pili-defective mutant (*pilB*) and a representative non-motile mutant (*flgK*) after incubation for 3 h at 37°C in the presence of PVC plastic. The wild-type and *pilB* strains attach to the PVC surface in a monolayer. Few or no cells are seen on the PVC plastic when the biofilm formation phenotype of the non-motile mutants is assessed.
C. The biofilm formation phenotype of the wild type, *pilB* (*sad-31*) and *flgK* at 8 h is shown. The wild-type strain forms a biofilm on the surface of the PVC plastic, which comprises a monolayer of cells punctuated by dispersed microcolonies. The *pilB* mutant has a slightly more dense monolayer than observed at 3 h but does not form the microcolonies evident with the wild-type strain. The *flgK* mutant has very few cells attached to the PVC plastic. Micrographs were taken at 400× magnification and ≈50 fields were searched for each strain tested and representative fields are shown.


A
30 min
1.5 hr
3 hr
5 hr
7.5 hr
B
wild-type
pilB
flgK
3 hrs
C
wild-type
pilB
flgK
8 hrs

1987; Grant *et al*., 1993; Korber *et al*., 1994; Simpson *et al*., 1995).

We also directly visualized the biofilm formation phenotype of a representative mutant defective in pili biogenesis (*pilB*). At the early time points (≤ 3 h), there was little difference in the biofilm formation phenotype of the wild type and the type IV pili mutants; both the wild-type and the pili-defective strain form a dispersed monolayer of cells on the surface of the PVC plastic. By 8 h, in contrast to the aggregates of cells formed by the wild-type strain, the pili-defective mutants did not develop these characteristic microcolonies. Furthermore, the wild-type strain almost completely covered the PVC surface with a dense, tightly packed layer of cells. The phenotype of the type IV pili mutants at this 8 h time point was unchanged from that observed at 3 h, that is a dispersed monolayer of cells. The other mutants defective in pili biogenesis (*pilC* and *pilY1*) had similar phenotypes (data not shown).

A role for twitching motility in biofilm formation

To define better the events that lead to microcolony formation by the wild type and to determine whether surface-based twitching motility plays a role in biofilm formation, we used phase-contrast time-lapse microscopy to follow a developing biofilm. Using time-lapse microscopy, we watched individual microcolonies formed by the wild-type strain over a period of 56 min (with images acquired at 15 s intervals). Shown in Fig. 5 is a montage of nine phase-contrast micrographs taken during biofilm formation by the wild-type strain every 7 min between 360 and 416 min after inoculation. Several microcolonies were followed through the course of this experiment to illustrate the movement of cells across the PVC plastic surface.

In Fig. 5, the white arrow indicates the position of a microcolony that is first clearly visible in B, becomes larger (C) but has dispersed by D. This microcolony does not reform during the course of this experiment (D–I). A series of time-lapse micrographs taken at 15 s intervals between 374 min (C) and 381 min (D) show that this microcolony disperses because the cells comprising the colony move apart while still remaining associated with the plastic surface (data not shown but can be viewed as a time-lapse movie at http://gasp.med.harvard.edu/GO5.html).

The black arrow points to a large microcolony evident in Fig. 5A. This large microcolony becomes progressively smaller (B–F) and eventually splits into two small, adjacent microcolonies (G). In H, these two adjacent microcolonies form a larger single colony that has grown slightly in size when visualized 7 min later (I).

The formation of microcolonies in this system is due to, in a large part, the aggregation of cells found dispersed in the monolayer of cells on the surface and not solely to the growth of the bacterial cells. This point is illustrated further by data presented in H and I. The dark circle in I indicates a dense, well-formed microcolony. However, this colony is

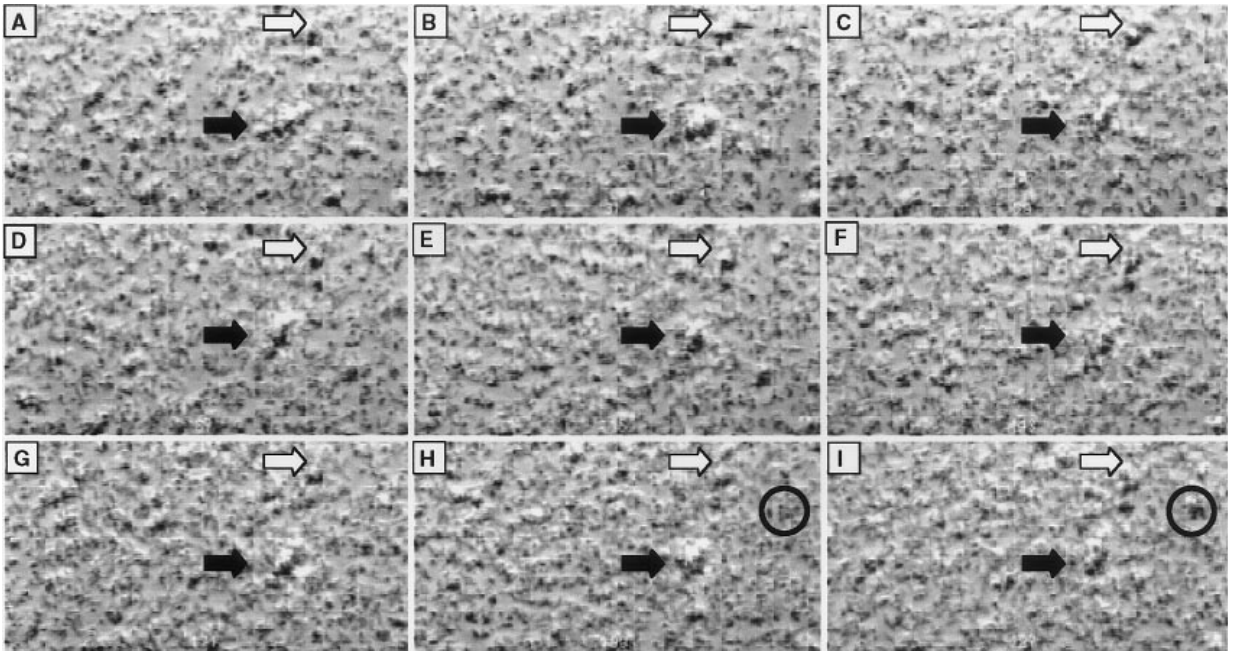


Fig. 5. A role for twitching motility in biofilm formation. Shown are a series of phase-contrast micrographs of a biofilm prepared as described in *Experimental procedures*. The first image (A) was taken at 6 h after inoculation and the subsequent images (B–I) are at 7 min intervals. Arrows indicate microcolonies that form and/or disperse over the course of the experiment. The black circles indicate the identical spot on the field in H and I. See the text for additional details.

not evident 7 min previously in H. The elapsed 7 min between the micrograph shown in H and the micrograph shown in I represents less than the time needed for a single population doubling under these growth conditions. Furthermore, analysis of the time-lapse film shows that this microcolony forms by recruiting adjacent cells from the monolayer (data not shown here, but can be viewed as a time-lapse movie at http://gasp.med.harvard.edu/GO5.html). The data described above and shown in Fig. 5 demonstrate the dynamic nature of microcolony formation and dispersal during the course of biofilm development.

As discussed above, type IV pili are required for surface-based twitching motility and mutants defective in type IV pili biogenesis do not make the microcolonies characteristic of the wild-type strain. It is important to note that none of the behaviours described above for the wild-type were observed in the representative type IV pili mutant, *pilB*. As shown above in Fig. 4, this strain does not form microcolonies when observed either after 8 h of growth or when monitored by time-lapse microscopy (data not shown).

Discussion

Flagellar-mediated motility, type IV pili and the initiation of biofilm formation

Among the non-biofilm forming mutants isolated in this screen were those defective in flagellar-mediated motility (Fig. 2). Motility has also been suggested to be involved in biofilm formation in other model systems (Montie *et al*., 1982; de Weger *et al*., 1987; Smit *et al*., 1989; Korber *et al*., 1994; O'Toole and Kolter, 1998; Pratt and Kolter, 1998). Therefore, the isolation of non-motile strains helps to validate our experimental approach. Furthermore, for one of the mutants (*sad-36*), we have shown that the insertion element in this strain is in a structural gene required for the synthesis of a functional flagellum. For a more complete discussion of the role of flagella-mediated motility in biofilm development in *Pseudomas* and *E. coli* see Pratt and Kolter (1998).

Another class of mutants isolated in this screen are those defective in the synthesis of type IV pili, including strains with mutations in *pilB*, *pilC* and *pilY1*. These polar-localized pili are virulence factors and have been implicated in both the adherence of bacteria to eukaryotic cell surfaces (Woods *et al*., 1980; Ramphal *et al*., 1984; Bieber *et al*., 1998; Doig *et al*., 1988; Tang *et al*., 1995) and in a form of surface-associated movement termed twitching motility (Bradley, 1980; Mattick *et al*., 1996; Darzins and Russell, 1997). The *pilBCD* operon is thought to encode accessory factors required for type IV pili biogenesis (Nunn *et al*., 1990). The *pilY1* gene has sequence similarity to the *pilC* genes of *Neisseria meningitidis* and *Neisseria gonorrhoeae*, which have been shown to be required for type IV pili biogenesis and attachment to eukaryotic cells in these organisms (Jonsson *et al*., 1991; Rudel *et al*., 1995; Rahman *et al*., 1997).

Using the sad *mutants to initiate the dissection of the biofilm developmental pathway*

We have used the *sad* mutants, in conjunction with phase-contrast microscopy, to begin to elucidate the early steps in biofilm formation on an abiotic surface. The goal of these studies is to correlate specific bacterial structures with defined steps in the biofilm developmental pathway. Our current model for biofilm formation is shown in Fig. 6. The direct visual inspection of the biofilm formation phenotype of non-motile strains on PVC plastic revealed that, compared with the wild-type strain, only very few cells could make a stable interaction with this abiotic surface (Fig. 4B). Furthermore, our studies were performed with a molecularly characterized non-motile strain with a defined defect in flagellar synthesis. Earlier studies used uncharacterized strains that may have had pleiotrophic defects (Lawrence *et al*., 1987; Mills and Powelson, 1996). Our observations, which concur with previous studies (Lawrence *et al*., 1987), suggest that motility is important for the cells to make initial contacts with an abiotic surface. From these experiments it is not clear whether the flagellum plays a direct role as an adhesin as previously suggested for *P. fluorescens* (Lawrence *et al*., 1987), or, as also proposed, that flagellar-mediated motility is required to bring the cell within close proximity of the surface to overcome repulsive forces between the bacterium and the surface to which it will eventually attach (Mills and Powelson, 1996). It is possible that flagellar-mediated motility is required for both of these processes.

The direct visual analysis of mutants defective in type IV pili biogenesis revealed that early biofilm formation by these mutants ($\leq$ 3 h) was very similar to the wild-type strain, in

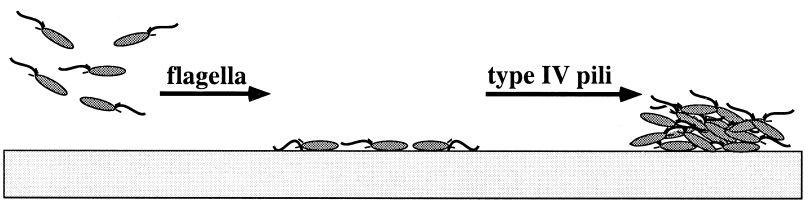


Fig. 6. A model for biofilm formation. Shown is our current model for the role of flagella and type IV pili in biofilm formation. Flagella or flagella-mediated motility appears to be important for the formation of a bacterial monolayer of the abiotic surface. Type IV pili appear to play a role in downstream events such as microcolony formation. See the text for additional details.

that all of these strains form a dispersed monolayer of cells on the PVC plastic. However, the wild-type strain eventually forms a dense, tightly packed monolayer of cells punctuated by microcolonies on the plastic surface; the pili-defective mutants remained as a dispersed monolayer of cells (Fig. 4B and C). Based on our observations, it appears that type IV pili are required downstream of flagella but still early in this developmental pathway.

The data above suggest possible roles for type IV pili and type IV pili-mediated twitching motility in *P. aeruginosa* biofilm development. It is possible that type IV pili play a direct role in stabilizing interactions with the abiotic surface (that may have been initiated via flagella or flagella-mediated motility) and/or in the cell-to-cell interactions required to form a microcolony. Type IV pili-mediated twitching motility may also be necessary for cells to migrate along the surface to form the multicell aggregates characteristic of the wild-type strain. In support of such a role for twitching motility in biofilm formation, we present evidence that the wild-type strain does move across the surface and form cell aggregates by recruiting cells from the adjacent monolayer (Fig. 5). It is important to note that strains defective in pili biogenesis (like the *pilB* mutant) express neither twitching motility nor microcolony formation phenotypes.

The microscopic colonies formed by *P. aeruginosa* PA14 on PVC plastic during biofilm development are reminiscent of the macroscopic aggregates formed by *Myxococcus* during the development of fruiting bodies (Kaiser, 1984). Furthermore, the development of these fruiting bodies by *Myxococcus* has also been shown to require type IV pili (Wu and Kaiser, 1995). Interestingly, recent studies suggest a requirement for homoserine lactones (HSLs) to express type IV pili-mediated twitching motility (Glessner *et al*., 1998). These data suggest a role for cell-to-cell signalling early in microcolony formation in addition to the established role of HSLs in the later stages of biofilm development (Davies *et al*., 1998). However, it is also important to note that type IV pili mutants can still interact with the abiotic surface, suggesting the existence of additional, unidentified adhesions that promote cell-to-surface interactions.

The requirement for type IV pili in biofilm formation on an abiotic surface has an additional important implication. As mentioned above, type IV pili have been shown to be important for bacterial adhesion to eukaryotic cell surfaces and pathogenesis (Woods *et al*., 1980; Ramphal *et al*., 1984; Sato *et al*., 1988; Ramphal *et al*., 1991; Tang *et al*., 1995; Bieber *et al*., 1998). These data suggest that there may be an overlap in factors required for the initiation of biofilm formation on an abiotic surface and the factors necessary for bacterial attachment and pathogenesis *in vivo*. If this is the case, the biofilm formation assay presented here may serve as a simple primary screen for identifying novel virulence factors. The analysis of additional mutants isolated in this screen and their testing in models of pathogenesis is in progress.

Experimental procedures

Bacterial strains, media and chemicals

P. aeruginosa PA14 was grown on rich medium (Luria Bertani; LB) or minimal medium (as indicated in each experiment) at 37°C, unless otherwise noted. The minimal medium used was minimal M63 salts (Pardee *et al*., 1959) supplemented with glucose (0.2%), $MgSO_4$ (1 mM) and, where indicated, casamino acids (CAA, 0.5%). Antibiotics were added at the following concentrations: for *E. coli*, ampicillin (Ap), 150 μg ml^{-1}; naladixic acid (Nal), 20 μg ml^{-1}; for *P. aeruginosa*, Tc, 150 μg ml^{-1}. All enzymes for DNA manipulation were purchased from New England Biolabs. All plasmids were constructed in *E. coli* JM109 using standard protocols (Ausubel *et al*., 1990). Plasmids were transferred to *P. aeruginosa* by electroporation (Bloemberg *et al*., 1997). Transductions were performed as reported (Jensen *et al*., 1998). Assays for assessing flagellar-mediated motility were performed as reported (O'Toole and Kolter, 1998). Twitching motility was assessed as described (Whitchurch *et al*., 1990).

Genetic techniques

Transposon mutants were generated with Tn*5*-B30(Tc^r) using a modification of published protocols (Simon *et al*., 1989) as described (O'Toole and Kolter, 1998). The resulting transposon mutants were screened as described below.

Biofilm formation assay

Screen for mutants defective in biofilm formation

This assay is based on the ability of bacteria to initiate biofilm formation on polyvinylchloride plastic (PVC). The initiation of biofilm formation was assayed as described (O'Toole and Kolter, 1998).

Quantification of biofilm formation

Biofilm formation was quantified as described (O'Toole and Kolter, 1998). Briefly, the crystal violet was solubilized in 95% ethanol and the absorbance was determined at 600 nm.

Microscopy

Visualization of bacterial cells attached to PVC was performed by phase-contrast microscopy (400× magnification) using a Nikon Diaphot 200 inverted microscope (Nikon). The images were captured with a black and white CCD72 camera integrated with a Power Macintosh 8600/300 computer with video capability (Cupertino). The images were processed using SCION IMAGE software, a modification of NIH IMAGE (NIH) by the Scion Corporation.

Molecular techniques

DNA sequence flanking transposon mutants were determined

using arbitrary PCR (Caetano-Annoles, 1993) as described (O'Toole and Kolter, 1998). Southern blots were performed as follows: chromosomal DNA of the *sad* mutants was prepared (Pitcher *et al*., 1989), digested with *Eco*RI (Tn*5*-B30 does not have a *Eco*RI site) and transferred to GeneScreen Plus (NEN Research Products) as reported (Ausubel *et al*., 1990). The hybridization was performed with the ECL direct nucleic acid labelling and detection system (Amersham Life Science) according to the manufacturer's instructions without modification. The DNA probe used was derived from the insertion sequence element (IS50) of Tn*5* and generated using PCR with the Tn*5* element as a template. The PCR primers used to generate the probe were IS50R.1 (5′-GCTTCCTTTAGCAGCCCTTGCGC-3′) and IS50R.2 (5′-CTTCCATGTGACCTCCTAACATGG-3′).

Acknowledgement

We thank L. G. Rahme for providing us with strain *P. aeruginosa* PA14. This work was supported by grant number GM58213 from the NIH to R.K. and Fellowship DRG of the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation to G.A.O. We also thank the Micro Core Facility at the Harvard Medical School for DNA sequencing and L. A. Pratt and S. E. Finkel for a critical review of the manuscript.

References

Alm, R.A., and Mattick, J.S. (1995) Identification of a gene, *pilV*, required for type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*, whose product possesses a pre-pilin-like leader sequence. *Mol Microbiol* **16:** 485–496.

Alm, R.A., Hallinan, J.P., Watson, A.A., and Mattick, J.S. (1996) Fimbrial biogenesis genes of *Pseudomonas aeruginosa: pilW* and *pilX* increase the similarity of type 4 fimbriae to the GSP protein-secretion systems and *pilY1* encodes a gonococcal PilC homologue. *Mol Microbiol* **22:** 161–173.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Ausubel, F.A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., *et al*. (1990) *Current Protocols in Molecular Biology*. New York: Wiley Interscience.

Bieber, D., Ramer, S.W., Wu, C.-Y., Murray, W.J., Tobe, T., Fernandez, R., and Schoolnik, G.K. (1998) Type IV pili, transient bacterial aggregates, and virulance of enteropathogenic *Escherichia coli*. *Science* **280:** 2114–2118.

Bloemberg, G.V., O'Toole, G.A., Lugtenberg, B.J.J., and Kolter, R. (1997) Green fluorescent protein as a marker for *Pseudomonas* spp. *Appl Environ Microbiol* **63:** 4543–4551.

Bodey, G.P., Bolivar, R., Fainstein, V., and Jadega, L. (1983) Infections caused by *Pseudomonas aeruginosa*. *Rev Infect Dis* **5:** 279–313.

Bradley, D.E. (1980) A function of *Pseudomonas aeruginosa* PAO polar pili: twitching motility. *Can J Microbiol* **26:** 146–154.

Caetano-Annoles, G. (1993) Amplifying DNA with arbitrary oligonucleotide primers. *PCR Methods Appl* **3:** 85–92.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., and Lappin-Scott, H.M. (1995) Microbial biofilms. In *Annual Review of Microbiology*, Vol. 49. Ornston, L.N., Barrows, A., and Greenberg, E.P. (eds). Palo Alto, CA: Annu Rev Inc, pp. 711–745.

Darzins, A. (1994) Characterization of a *Pseudomonas aeruginosa* gene cluster involved in pilus biosynthesis and twitching motility: sequence similarity to the chemotaxis proteins of enterics and the gliding bacterium *Myxocoocus xanthus*. *Mol Microbiol* **11:** 137–153.

Darzins, A.L., and Russell, M.A. (1997) Molecular genetic analysis of type-4 pilus biogenesis and twitching motility using *Pseudomonas aeruginosa* as a model system-a review. *Gene* **192:** 109–115.

Davies, D.G., and Geesey, G.G. (1995) Regulation of the alginate biosynthesis gene *algC* in *Pseudomonas aeruginosa* during biofilm development in continuous culture. *Appl Environ Microbiol* **61:** 860–867.

Davies, D.G., Chakabarty, A.M., and Geesey, G.G. (1993) Exopolysaccharide production in biofilms: substratum activation of alginate gene expression by *Pseudomonas aeruginosa*. *Appl Environ Microbiol* **59:** 1181–1186.

Davies, D.G., Parsek, M.R., Pearson, J.P., Iglewski, B.H., Costerton, J.W., and Greenberg, E.P. (1998) The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* **280:** 295–298.

Doig, P., Todd, T., Sastry, P.A., Lee, K.K., Hodges, R.S., Paranchych, W., and Irvin, R.T. (1988) Role of pili in adhesion of *Pseudomonas aeruginosa* to human respiratory epithelial cells. *Infect Immun* **56:** 1641–1646.

Fletcher, M. (1996) Bacterial attachment in aquatic environments: a diversity of surfaces and adhesion strategies. In *Bacterial Adhesion: Molecular and Ecological Diversity*. Fletcher, M. (ed.). New York: John Wiley & Sons, pp. 1–24.

Fletcher, E.L., Weissman, B.A., Efron, N., Fleiszig, S.M.J., Curcio, A.J., and Brennan, N.A. (1993) The role of pili in the attachment of *Pseudomonas aeruginosa* to unworn hydrogel contact lenses. *Curr Eye Res* **12:** 1067–1071.

Glessner, A., Iglewsski, B.H., and Robinson, J.B. (1998) Role of *Pseudomonas aeruginosa lasR* and *rhl* quorum-sensing systems in control of twitching motility. *General Meeting of the American Society for Microbiology*, Poster D-88.

Govan, J.R.W., and Deretic, V. (1996) Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol Rev* **60:** 539–574.

Grant, C.C.R., Konkel, M.E., Cieplak, J., W., and Tompkins, L.S. (1993) Role of flagella in adherence, internalization, and translocation of *Campylobacter jejuni* in nonpolarized and polarized epithelial cell cultures. *Infect Immun* **61:** 1764–1771.

Homma, M., DeRosier, D.J., and Macnab, R.M. (1990) Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their relationship to other axial components of the flagellum. *J Mol Biol* **213:** 819–832.

Hoyle, B.D., and Costerton, W.J. (1991) Bacterial resistance to antibiotics: the role of biofilms. *Prog Drug Res* **37:** 91–105.

Jensen, E.C., Schrader, H.S., Rieland, B., Thompson, T.L., Lee, K.W., Nickerson, K.W., and Kokjohn, T.A. (1998) Prevalence of broad-host-range lytic bacteriophages of

Sphaerotilus natans, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Appl Environ Microbiol* **64:** 575–580.

Jonsson, A.B., Nyberg, G., and Normark, S. (1991) Phase variation of gonococcal pili by frameshift mutation in *pilC*, a novel gene for pilus assembly. *EMBO J* **10:** 477–488.

Kaiser, D. (1984) Regulation of multicellular development in *Myxobacteria*. In *Microbial Development*. Losick, R. & Shapiro, L. (eds). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 197–218.

Korber, D.R., Lawrence, J.R., and Caldwell, D.E. (1994) Effect of motility on surface colonization and reproductive success of *Pseudomonas fluorescens* in dual-dilution continuous culture and batch culture systems. *Appl Environ Microbiol* **60:** 1421–1429.

Lawrence, J.R., Delaquis, P.J., Korber, D.R., and Caldwell, D.E. (1987) Behavior of *Pseudomonas fluorescens* within the hydrodynamic boundary layers of surface microenvironments. *Microb Ecol* **14:** 1–14.

Mattick, J.S., Whitchurch, C.B., and Alm, R.A. (1996) The molecular genetics of type-4 fimbriae in *Pseudomonas aeruginosa* – a review. *Gene* **179:** 147–155.

Miller, M.J., and Ahearn, D.G. (1987) Adherence of *Pseudomonas aeruginosa* to hydrophilic contact lenses and other substrata. *J Clin Microbiol* **25:** 1392–1397.

Mills, A.L., and Powelson, D.K. (1996) Bacterial interactions with surfaces in soil. In *Bacterial Adhesion: Molecular and Ecological Diversity*. Fletcher, M. (ed.). New York: John Wiley & Sons, pp. 25–57.

Montie, T.C., Doyle-Huntzinger, D., Craven, R.C., and Holder, I.A. (1982) Loss of virulence associated with absence of flagellum in an isogenic mutant of *Pseudomonas aeruginosa* in the burned-mouse model. *Infect Immun* **38:** 1296–1298.

Nickel, J.C., Ruseska, I., Wright, J.B., and Costerton, J.W. (1985) Tobramycin resistance of *Pseudomonas aeruginosa* cells growing as a biofilm on urinary tract catheter. *Antimicrob Agents Chemother* **27:** 619–624.

Nickel, J.C., Downey, J.A., and Costerton, J.W. (1989) Ultrastructural study of microbiologic colonization of urinary catheters. *Urology* **34:** 284–291.

Nunn, D., Bergman, S., and Lory, S. (1990) Products of three accessory genes, *pilB*, *pilC*, and *pilD* are required for biogenesis of *Pseudomonas aeruginosa* pili. *J Bacteriol* **172:** 2911–2919.

O'Toole, G.A., and Kolter, R. (1998) The initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signaling pathways: a genetic analysis. *Mol Microbiol* **28:** 449–461.

Pardee, A.B., Jacob, F., and Monod, J. (1959) The genetic control and cytoplasmic expression of 'inducibility' in the synthesis of β-galactosidase in *E. coli*. *J Mol Biol* **1:** 165–178.

Pemberton, J.M. (1973) F116. a DNA bacteriophage specific for the pili of *Pseudomonas aeruginosa* strain PAO. *Virology* **55:** 558–560.

Pitcher, D.G., Saunders, N.A., and Owen, R.J. (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* **8:** 151–156.

Pratt, L.A., and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: defining the roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* **30:** 285–293.

Rahman, M., Kallstrom, H., Normark, S., and Jonsson, A.-B. (1997) PilC of pathogenic *Neisseria* is associated with the bacterial cell surface. *Mol Microbiol* **25:** 11–25.

Ramphal, R., Sadoff, J.C., Pyle, M., and Silipigni, J.D. (1984) Role of pili in the adherence of *Pseudomonas aeruginosa* to injured tracheal epithelium. *Infect Immun* **44:** 38–40.

Ramphal, R., Koo, L., Isimoto, K.S., Totten, P., Lara, J.C., and Lory, S. (1991) Adhesion of *Pseudomonas aeruginosa* pilin-deficient mutants to mucin. *Infect Immun* **59:** 1307–1311.

Rudel, T., van Putten, J.P.M., Gibbs, C.P., Haas, R., and Meyer, T.F. (1992) Interaction of two variable proteins (PilE and PilC) required for pilus-mediated adherence of *Neisseria gonorrhoeae* to human epithelial cells. *Mol Microbiol* **6:** 3439–3450.

Rudel, T., Schleuerflug, I., and Meyer, T. (1995) *Neisseria* PilC protein identified as type-4 pilus tip-located adhesin. *Nature* **373:** 357–359.

Russell, M.A., and Darzins, A. (1994) The *pilE* gene product of *Pseudomonas aeruginosa*, required for pilus biogenesis, shares amino acid sequence identity with the N-termini of type 4 prepilin proteins. *Mol Microbiol* **13:** 973–985.

Sato, H., Okinaga, K., and Saito, H. (1988) Role of pili in the pathogenesis of *Pseudomonas aeruginosa* burn infections. *Microbiol Immunol* **32:** 131–139.

Simon, R., Quandt, J., and Klipp, W. (1989) New derivatives of transposon Tn*5* suitable for mobilization of replicons, generation of operon fusions and induction of genes in Gram-negative bacteria. *Gene* **80:** 160–169.

Simpson, D.A., Ramphal, R., and Lory, S. (1995) Characterization of *Pseudomonas aeruginosa fliO*, a gene involved in flagellar biosynthesis and adherence. *Infect Immun* **63:** 2950–2957.

Smit, G., Kijne, J.W., and Lugtenberg, B.J.J. (1989) Roles of flagella, lipopolysaccharide, and a Ca^{+2}-dependent cell surface protein in attachment of *Rhizobium leguminosarum* biovar *viciae* to pea root hair tips. *J Bacteriol* **171:** 569–572.

Tang, H., Kays, M., and Prince, A. (1995) Role of *Pseudomonas aeruginosa* pili in acute pulmonary infection. *Infect Immun* **63:** 1278–1285.

de Weger, L.A., van der Vlught, C.I.M., Wijfjes, A.H.M., Bakker, P.A.H.M., Schippers, B., and Lugtenberg, B. (1987) Flagella of a plant-growth-stimulating *Pseudomonas fluorescens* are required for colonization of potato roots. *J Bacteriol* **169:** 2769–2773.

Whitchurch, C.B., Hobbs, M., Livingston, S.P., Krishnapillai, V., and Mattick, J.S. (1990) Characterization of a *Pseudomonas aeruginosa* twitching motility gene and evidence for a specialized protein export system widespread in eubacteria. *Gene* **101:** 33–44.

Wimpenny, J.W.T., and Colasanti, R. (1997) A unifying hypothesis for the structure of microbial biofilms based on cellular automaton models. *FEMS Microbiol Ecol* **22:** 1–16.

Woods, D.E., Strauss, D.C., Johanson, J.W.G., Berry, V.K., and Bass, J.A. (1980) Role of pili in adherence to buccal epithelial cells. *Infect Immun* **29:** 1146–1151.

Wu, S.S., and Kaiser, D. (1995) Genetic and functional evidence that Type IV pili are required for social gliding motility in *Myxococcus xanthus*. *Mol Microbiol* **18:** 547–558.

Molecular Microbiology (2014) 93(3), 494–504 ■
doi:10.1111/mmi.12673
First published online 9 July 2014

A conserved type IV pilin signal peptide H-domain is critical for the post-translational regulation of flagella-dependent motility

Rianne N. Esquivel and Mechthild Pohlschroder*
Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA.

Summary

In many bacteria and archaea, type IV pili facilitate surface adhesion, the initial step in biofilm formation. *Haloferax volcanii* has a specific set of adhesion pilins (PilA1-A6) that, although diverse, contain an absolutely conserved signal peptide hydrophobic (H) domain. Data presented here demonstrate that these pilins (PilA1-A6) also play an important role in regulating flagella-dependent motility, which allows cells to rapidly transition between planktonic and sessile states. Cells lacking adhesion pilins exhibit a severe motility defect, however, expression of any one of the adhesion pilins in trans can rescue the motility and adhesion. Conversely, while deleting *pilB3-C3*, genes required for PilA pilus biosynthesis, results in cells lacking pili and having an adhesion defect, it does not affect motility, indicating that motility regulation requires the presence of pilins, but not assembled pili. Mutagenesis studies revealed that the pilin-dependent motility regulatory mechanism does not require the diverse C-terminal region of the PilA pilins but specifically involves the conserved H-domain. This novel post-translational regulatory mechanism, which employs components that promote biofilm formation to inhibit motility, can provide a rapid response to changing environmental conditions. A model for this regulatory mechanism, which may also be present in other prokaryotes, is discussed.

Introduction

Biofilms are complex microbial communities, bound by a matrix of extracellular polymeric substance (EPS) that allow cells to tolerate stress conditions such as high UV exposure (Hansen *et al.*, 2007; Monds and O'Toole, 2009; Haussler and Fuqua, 2013; Orell *et al.*, 2013a). Once environmental conditions become unfavourable, cells must rapidly transition from a planktonic to a sessile state (McDougald *et al.*, 2011). For instance, flagella are required for the swimming motility of planktonic cells and enhance initial type IV pilus-dependent attachment of many bacteria and archaea to surfaces (O'Toole and Kolter, 1998; Ghosh and Albers, 2011). However, flagella are not required, and indeed hinder, subsequent stages of biofilm development whereas adhesive pili are crucial to the formation of cell aggregates in many prokaryotic biofilms (Fröls *et al.*, 2008; Karatan and Watnick, 2009; Pohlschroder *et al.*, 2011; Esquivel *et al.*, 2013).

Several strategies to inactivate already existing flagella and to inhibit the biosynthesis of new flagella have evolved (Guttenplan and Kearns, 2013). For example, in *Bacillus subtilis* the glycosyltransferase, EpsE, which is required for matrix biosynthesis, binds to the flagella rotor and disengages motor force-generating elements, providing a rapid mechanism for inhibiting flagella rotation (Blair *et al.*, 2008). High levels of c-di-GMP, produced by many bacteria during biofilm formation, also inhibit transcription of genes that encode flagella biosynthesis components, and upregulate the expression of genes involved in adhesive pili formation. Conversely, once a biofilm begins to disperse, a condition where pili expression becomes disadvantageous, c-di-GMP levels decrease, resulting in repression of *pil* genes and increased *fla* gene expression (Kuchma *et al.*, 2007; Boyd and O'Toole, 2012; Guttenplan and Kearns, 2013).

Interestingly, unlike the analogous surface filaments of bacteria, which are produced by distinct biosynthesis machineries, archaeal pili and flagella (the latter also known as archaella; Jarrell and Albers, 2012) are assembled by machineries that use homologous, or even share, core components (Ghosh and Albers, 2011; Pohlschroder *et al.*, 2011). These components are also homologous to proteins involved in bacterial pilus biosynthesis. As such, flagellins and pilins in many archaeal species are processed by the same PilD homologue, PibD (FlaK) (Albers *et al.*, 2003; Bardy and Jarrell, 2003; Giltner *et al.*, 2012). Unlike signal peptidase I and II which have recognition sites that follow the signal peptide hydrophobic (H) domain,

Accepted 13 June, 2014. *For correspondence. E-mail pohlschr@sas.upenn.edu; Tel. (+1) 215 573 2283; Fax (+1) 215 898 8780.

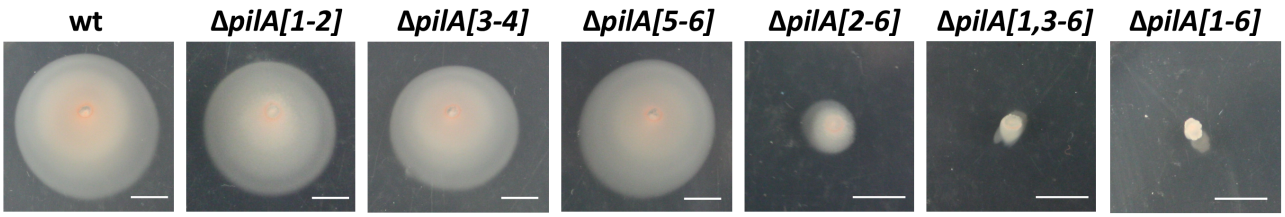


Fig. 1. Deleting at least five pilin genes impairs flagella-dependent motility. Motility assays of *H. volcanii* wild-type (wt) and *pilA* deletion strains after three days of incubation in MGM. Bars, 5 mm.

the PibD recognition sites in type IV pilins precede the H-domain and the processed pilins and flagellins are thought to be retained in the membrane prior to assembly. In the mature protein, this H-domain forms an alpha helix that is predicted to serve as the core of the flagella or pili (Craig *et al.*, 2006; Pohlschroder *et al.*, 2011; Giltner *et al.*, 2012). Archaeal pili and flagella assembly also requires the presence of homologues of the pilus biosynthesis ATPase PilB and of the membrane protein PilC (Takhar *et al.*, 2013). However, distinct PilB and PilC paralogues are required for archaeal type IV pilus and flagella biosynthesis (Albers and Pohlschroder, 2009; Ghosh and Albers, 2011; Lassak *et al.*, 2012). Consistent with having unique biosynthesis machineries, it has recently been shown that the transcriptional regulation of archaeal flagella and pili expression within biofilms is, as in bacteria, inversely related (Reimann *et al.*, 2012; Lassak *et al.*, 2013; Orell *et al.*, 2013b).

While recent advancements have been made in understanding the relationship between the transcriptional regulation of flagella and pili, little is known about the post-transcriptional regulation of these archaeal filaments. In the euryarchaeon *Haloferax volcanii*, which expresses two flagellin genes, *flgA1* and *flgA2*, the deletion of the gene encoding the flagellar subunit, *flgA2*, results in hypermotility (Tripepi *et al.*, 2013). The effect that deletion of a flagellar subunit leads to hypermotility is unprecedented, but the regulatory mechanism underlying this effect has not yet been determined.

H. volcanii is also the first motile organism to be reported where the initial adhesion to a surface does not require flagella (Tripepi *et al.*, 2010) and where a subset of six identified adhesion pilins (PilA1-A6) are required for microcolony formation, while others appear to inhibit this early step in biofilm formation (Esquivel *et al.*, 2013). Finally, each of these pilins, even though they are rather diverse, has a completely conserved H-domain that is required for the assembly of a pilus (Szabó *et al.*, 2007; Esquivel *et al.*, 2013).

In this study, we have shown that the PilA adhesion pilins play essential roles in the post-translational regulation of flagella-dependent motility and that the absolutely conserved H-domain specific to these adhesion pilins is essential to this regulatory function. Although this additional pilin function, the third pilin function identified that is involved in processes required for biofilm formation, has not been reported previously, it may not be unique to halophiles.

Results and discussion

ΔpilA[1-6] has a severe motility defect

Considering the hypermotility phenotype observed in *Sulfolobus acidocaldarius* lacking pilus biosynthesis genes *aapE*, *aapF* or the major pilin gene, *aapB* (Henche *et al.*, 2012) we stab-inoculated deletion mutants, lacking either single or multiple pilins, on motility plates and determined their motility phenotypes. Our results show that the deletion of a single *H. volcanii pilA* gene or two *pilA* genes does not significantly affect motility compared to that of the H53 parent strain (from hereon referred to as the wild-type) (Fig. 1). Unexpectedly, while these mutants (Δ*pilA1*, Δ*pilA2*, Δ*pilA*[*1-2*], Δ*pilA*[*3-4*], or Δ*pilA*[*5-6*]) and the wild-type strain form a visible halo after two days, a deletion strain expressing only chromosomally encoded PilA1 exhibited decreased motility, only forming a visible halo after three days (Fig. 1 and data not shown). This mutant, Δ*pilA*[*2-6*], also displays a severe adhesion defect (Esquivel *et al.*, 2013). Moreover, the non-adhering mutant strain expressing only chromosomal PilA2 and the mutant strain lacking all six *pilA* genes do not show any significant motility even after five days of incubation (Fig. 1 and Fig. S1). This is consistent with both of these mutants, Δ*pilA*[*1,3-6*] and Δ*pilA*[*1-6*], being unable to adhere (Esquivel *et al.*, 2013). These results strongly suggest that *H. volcanii* swimming motility is regulated in a pilin-dependent manner through a novel regulatory mechanism not previously identified.

The majority of ΔpilA[1-6] cells are non-flagellated

Motility, adhesion and filament assembly phenotypes of *H. volcanii* wild-type and mutant strains discussed in this manuscript are summarized in Table 1. The inhibited and delayed motility observed for the Δ*pilA*[*1-6*] strain might be due to the fact that: (1) only a few cells have functional

Table 1. Motility, filament assembly and adhesion phenotypes exhibited by *H. volcanii* wt and mutant strains.

Strain	Plasmid	Motility[a]		Filaments	Adhesion	Reference
			EM	α-His(cell/CsCl)		
wt	–	+++	+++	NA	+++	Tripepi *et al.* (2012)
Δ*pilA[1-6]*	–	–	–	NA	–	This study/Esquivel *et al.* (2013)
Δ*pilA[1-6]*	PilA1His	+++	++[b]	++/–	++	This study/Esquivel *et al.* (2013)
Δ*pilA[1-6]*	PilA1HybHis	–	+	–/–	–	This study/Esquivel *et al.* (2013)
Δ*pilA[1-6]*	FlgA1HybHis	++	++	ND	–	This study
Δ*pilA[1-6]*	FlgA1His	–	ND	ND	–	This study
Δ*flgA[1-2]*	–	–	+++	NA	+++	Tripepi *et al.* (2012)
Δ*flgA[1-2]*	FlgA[1-2]His	+++	+++	+++/ND	+++	Tripepi *et al.* (2012)
Δ*pilA[1-6]*Δ*flgA[1-2]*	–	–	–	NA	–	This study
Δ*pilA[1-6]*Δ*flgA[1-2]*	FlgA[1-2]His	–	+	+/–	–	This study
Δ*pilB3-C3*	–	+++	+++	NA	+	This study/Tripepi *et al.* (2013)
Δ*pilB3-C3*	PilB3-C3His	+++	+++	NA	++++	This study/Tripepi *et al.* (2013)
Δ*pilB3-C3*Δ*flgA[1-2]*	–	–	–	NA	+	This study
Δ*flgA1*	–	–	++	NA	+++	Tripepi *et al.* (2013)
Δ*flgA1*	FlgA1His	+++	+++	++/++	+++	Tripepi *et al.* (2013)
Δ*flgA1*	FlgA1HybHis	–	++	–/ND	+++	This study
Δ*pilB3-C3*Δ*flgA1*[c]	–	++	+	NA	+	This study
Δ*pilB3-C3*Δ*flgA1*	FlgA1His	++	ND	++/ND	+	This study
Δ*pilB3-C3*Δ*flgA1*	FlgA1HybHis	++	ND	++/–	+	This study
Δ*flgA2*	–	++++	++++	NA	+++	Tripepi *et al.* (2013)
Δ*pilA[1-6]*Δ*flgA2*	–	–	ND	NA	–	This study

a. Motility after 3 days of incubation.
b. Pili were only detected associated with the cells.
c. Only motile on semi-defined media.

flagella; or (2) the flagella are not fully functional in this mutant background. To distinguish between these possibilities, we compared *H. volcanii* Δ*pilA*[*1-6*] with the wild-type strain using transmission electron microscopy (TEM). While about 40% of the wild-type cells have filaments associated with them, filamentous surface structures were observed on only 2 of approximately 100 Δ*pilA*[*1-6*] cells analysed (Fig. 2A).

We have previously demonstrated that flagella and pili are released into the culture supernatants upon centrifugation (Tripepi *et al.*, 2012). To determine whether the Δ*pilA*[*1-6*] cells produce flagella that are unstable and shed, even under the mild conditions used for TEM preparations, we analysed culture supernatant fractions of the wild-type, Δ*flgA*[*1-2*] and Δ*pilA*[*1-6*] strains using cesium chloride (CsCl) gradient centrifugation. While TEM cannot

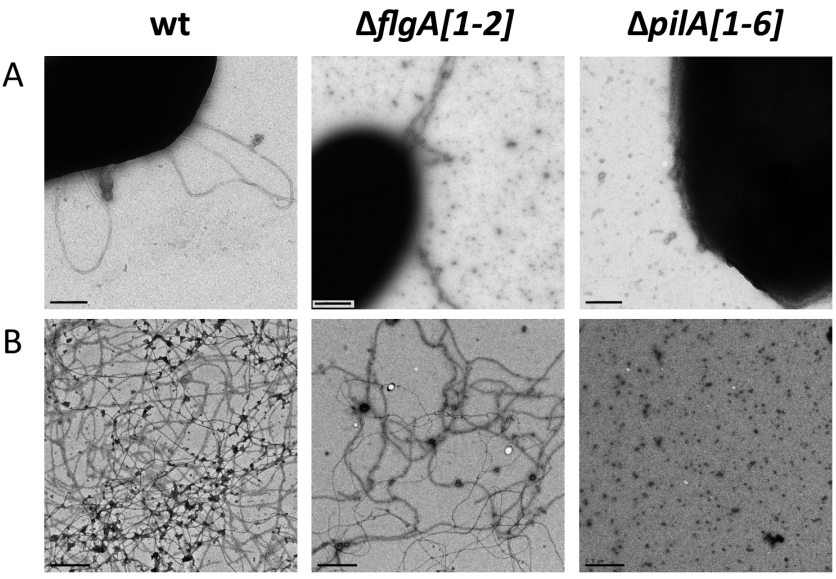


Fig. 2. Only a small subset of cells lacking PilA1-A6 contain flagella.
A. TEM of whole cells of wt and deletion strains. Bars, 200 nm. Images represent approximately 40% of wt and Δ*flgA*[*1-2*] cells and 98% of Δ*pilA*[*1–6*] cells analysed.
B. TEM of surface filaments purified by CsCl gradient centrifugation. Bars, 0.5 μm.

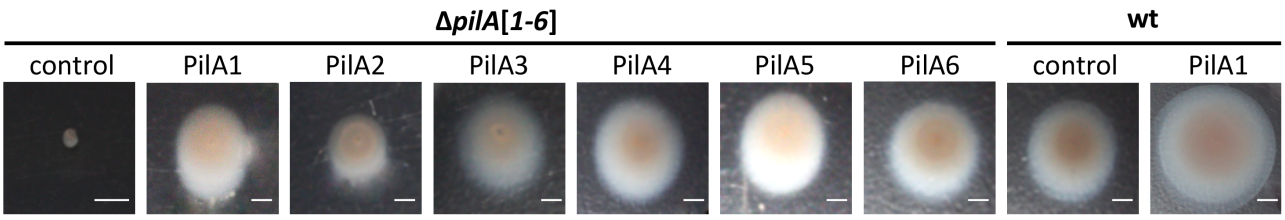


Fig. 3. One PilA pilin is sufficient to complement the *H. volcanii* Δ*pilA[1–6]* motility defect. Motility assays of *H. volcanii* Δ*pilA[1-6]* strain transformed with pTA963 (control) or pTA963 encoding a His-tagged version of one of the six pilin genes under the regulation of a *trp*-inducible promoter or wt transformed with pTA963 or pTA963 encoding *pilA1His*. Bars, 5 mm.

distinguish between flagella and pili, examination of the CsCl gradient purified filaments revealed that the Δ*flgA*[*1-2*] culture supernatant contains a significantly reduced number of filaments compared to the wild-type, indicating that an ample portion of the filaments produced by the wild-type are flagella (Fig. 2B). Conversely, in the Δ*pilA*[*1-6*] fractions, no filaments were detected. These observations indicate a significantly decreased synthesis or stability of flagella when pilin genes are absent (Fig. 2B).

Each of the six PilA pilins can complement the motility defect of the ΔpilA[1-6] strain

The motility phenotypes of the Δ*pilA* strains suggest that effective flagella-dependent motility requires the presence of at least some PilA pilins (Fig. 1). Consistent with the delayed motility phenotype observed for a Δ*pilA*[*1-6*] strain being due to the absence of all PilA adhesion pilins, this defect can be rescued by the expression of any one of the six PilA pilins in trans (Fig. 3). While the Δ*pilA*[*1-6*] deletion strain expressing PilA2 is somewhat less motile than the wild-type, the expression of PilA1, 3, 4, 5 or 6 in trans fully rescues the Δ*pilA*[*1-6*] motility defect. Similarly, PilA2 only partially rescues the adhesion defect of the Δ*pilA*[*1-6*] strain (Esquivel *et al*., 2013). Additionally, expression of each of the pilins individually in trans in a wild-type strain results in a hypermotility phenotype, supporting our hypothesis that a novel mechanism, involving PilA pilins, regulates *H. volcanii* flagella-dependent motility (Fig. 3 and Fig. S2). Since the expression of a single adhesion pilin gene can successfully complement this motility defect, this regulatory mechanism might operate in other archaea or bacteria but may yet have gone unnoticed.

PilA dependent regulation of motility is post-translational

We previously determined that FlgA[1-2]His can complement the motility defect of a Δ*flgA*[*1-2*] strain when expressed in trans from a *trp* inducible promoter (Large *et al*., 2007; Tripepi *et al*., 2010). However, expressing this construct in the Δ*pilA*[*1-6*]Δ*flgA*[*1-2*] strain does not complement the motility defect, indicating that the lack of pili may have a post-translational effect (Fig. 4A). While raising antibodies against *H. volcanii* flagellins and pilins has proven difficult, FlgA2His can be readily detected in cell extracts by using an anti-His antibody in Western blot analyses (Fig. 4B). However, only a faint band was identified in Western blots of protein from the supernatant fractions. To confirm these results we used CsCl gradient centrifugation to isolate flagella in a large-scale preparation from 1 l of culture supernatant of the Δ*pilA*[*1-6*]Δ*flgA*[*1-2*] strain expressing FlgA[1-2]His from the *trp* inducible promoter. FlgA2His was not detected in this preparation (Fig. 4B). In fact, consistent with flagella being unstable in these cells, none of the approximately 100 cells analysed by TEM had surface filaments (Fig. 4C) and only two small,

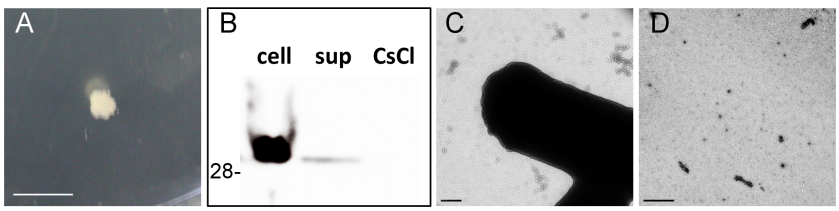


Fig. 4. Lack of pilins results in post-translational regulation of flagella-dependent motility. Δ*pilA[1-6]*Δ*flgA[1-2]* strain expressing *flgA[1-2]His* under the regulation of a *trp*-inducible promoter were analysed by (A) motility assay, (B) anti-His Western blot, and TEM of (C) whole cells or (D) filaments purified using CsCl gradient centrifugation. The Westerns were performed on protein extracts from cell lysates, TCA-precipitated proteins from culture supernatants (sup), or CsCl gradient purifications of surface filaments. Comparable culture volumes were used in cell and sup protein preparations. Molecular mass standard indicated on the left (in kDa). Bars, 5 mm (motility assay), 200 nm (cell) or 0.5 μm (CsCl).

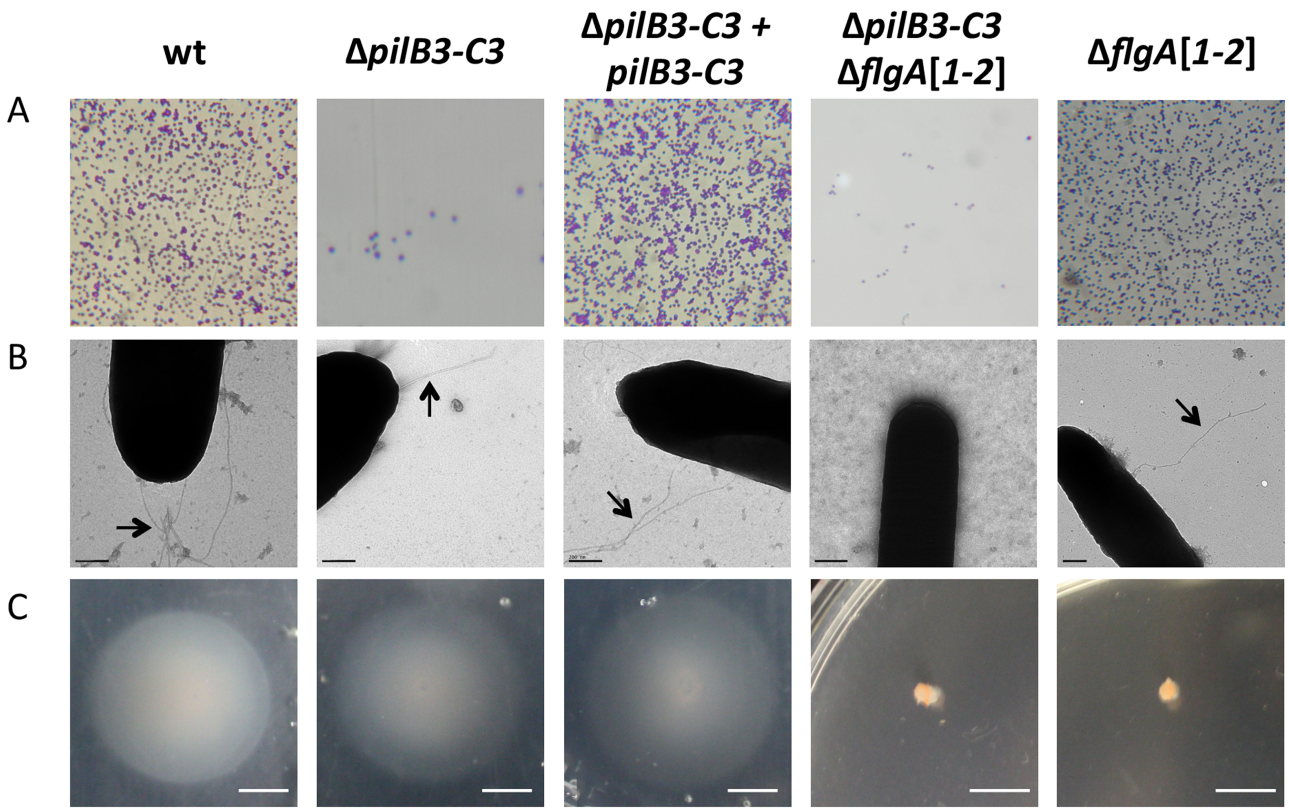


Fig. 5. Pilus assembly is not required for motility. (A) Surface adhesion assays (× 35 magnification), (B) TEM of whole cells and (C) motility assays of wt, Δ*pilB3-C3*, Δ*pilB3-C3*Δ*flgA[1-2]* and Δ*flgA[1-2]* strains transformed with pTA963 and Δ*pilB3-C3* transformed with pTA963 encoding *pilB3-C3His*. Surface filaments are observed on wt, Δ*pilB3-C3*, Δ*pilB3-C3* expressing PilB3-C3 and Δ*flgA[1-2]* strains (arrows). Bars, 5 mm (motility assay) or 200 nm (cell).

unattached filaments were observed in the surrounding supernatant. Moreover, no filaments were detected in CsCl gradient fractions (Fig. 4D). Taken together, these results strongly suggest that this pilin-dependent regulation of flagella biosynthesis is, at least in part, post-translational.

H. volcanii *PilB3 and PilC3 are required for PilA pilus biosynthesis*

We previously showed that the expression of any one of the six PilA pilins in a Δ*pilA*[*1-6*] strain results in the synthesis of pili (Esquivel *et al.*, 2013). To address whether flagella-dependent motility requires the presence of pili or whether pilins alone are sufficient, it was critical to identify pilus biosynthesis components of these PilA adhesion pili. Since the *H. volcanii* genome has five *pilB* and *pilC* containing operons in addition to its paralogues *flaI* and *flaJ*, respectively, in the flagella biosynthesis operon (Hartman *et al.*, 2010), our first objective was to identify the PilB/C paralogues required for the PilA adhesion pilus biosynthesis.

Consistent with the observed colocalization of *pilB* and *pilC* with pilin genes in many characterized *pil* operons, all *H. volcanii pilB* and *pilC* containing operons, except for the *pilB3-C3* containing operon, colocalize with at least one gene that encodes a predicted pilin-like protein (Fig. S3), a protein containing a pilin cleavage site as predicted by the program FlaFind (Szabó *et al.*, 2007). Since none of the conserved *pilA1-A6* pilin genes are associated with a *pilB* or *pilC* gene, we hypothesized that PilB3-C3 might be involved in PilA pilus biosynthesis.

We generated an *H. volcanii* deletion strain lacking *pilB3* and *pilC3* (Δ*pilB3-C3*) using homologous recombination (Allers and Ngo, 2003). The deletion was confirmed by PCR using primers homologous to sequences lying just outside these genes (Fig. S4A, Table S2) as was done previously for all other knockout strains (Esquivel *et al.*, 2013). The Δ*pilB3-C3* strain was tested for adhesion to plastic coverslips and it was found that it has a severe adhesion defect, indicating that the pilus-biosynthesis components PilB3-C3 are required for PilA adhesion pili biosynthesis (Fig. 5A). Consistent with a defect in pilus biosynthesis in this strain, when examined by TEM, only pili, not flagella, can be detected on a Δ*flgA[1-2]* strain, as previously determined (Tripepi *et al.*, 2012), while introducing a *pilB3-C3* deletion into this

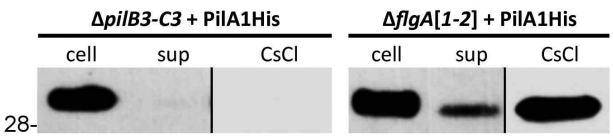


Fig. 6. PilA pilins are not incorporated into flagella. Western blot analysis using anti-His antibodies was performed on protein extracts from cell lysates (cell), TCA-precipitated proteins from the supernatant (sup), or CsCl gradient purifications of the surface filaments from the Δ*pilB3-C3* or Δ*flgA[1-2]* strains expressing PilA1His. Molecular mass standard indicated on the left (in kDa).

strain results in bald cells (Fig. 5B). Expression of PilB3-C3 in trans restored adhesion and pilus biosynthesis (Fig. 5A and B). These results also suggest that the limited level of adhesion observed in the Δ*pilB3-C3* strain is not due to residual assembled pili. Perhaps membrane-associated pilins are capable of mediating some adhesion.

Pili are not required for H. volcanii *motility*

Since PilB3-C3 are required for PilA pilus biosynthesis, we could use a Δ*pilB3-C3* strain to determine whether an assembled pilus is required to affect flagella-dependent motility or whether pilins alone are sufficient. When stab-inoculated into motility plates, the Δ*pilB3-C3* mutant displays motility similar to that of the wild-type strain, suggesting that assembled pili are not required for motility (Fig. 5C). Consistent with these results, while a Δ*pilA*[*1-6*] strain is bald, lacking surface-associated flagella as well as pili (Fig. 2A), a Δ*pilB3-C3* strain does have surface filaments, as determined by TEM (Fig. 5B). As noted above TEM does not distinguish between flagella and pili. However, considering that a Δ*pilB3-C3*Δ*flgA*[*1-2*] strain is bald (Fig. 5B), these data indicate that the filaments observed on this strain are flagella. To confirm that filaments observed in a Δ*pilB3-C3* background are flagella and that this strain is unable to assemble pili, we overexpressed PilA1His in the Δ*pilB3-C3* and the Δ*flgA[1-2]* strains. Using anti-His antibodies, Western blot analyses readily detected PilA1His in protein extracts from cell preparations of both strains, while this tagged pilin subunit can only be detected in supernatant fractions of cells that have functional PilB3-C3 (Fig. 6). Immunogold labelling of pili including His-tagged pilins was attempted but not successful, possibly due to the high salt concentration used or the His-tag being buried within the structure. However, lack of PilA1His in the supernatant of a Δ*pilB3-C3* strain was confirmed by Western blot analysis of a CsCl gradient purified protein preparation derived from 1 l of culture (Fig. 6). The absence of any His signal in the concentrated pili/flagella preparation of the Δ*pilB3-C3* supernatant fraction not only shows that pili are not formed but also strongly suggests that PilA1His is not incorporated into flagella and that the regulation of motility is mediated by membrane associated pilins.

The conserved H-domain, but not the pilin, is critical for motility regulation

Having established that pilin subunits, but not the pili, are required for regulation of motility, we set out to determine which parts of the pilin are important in regulating motility. We have shown that expression of a hybrid pilin, PilA1Hyb, containing the FlgA1 signal peptide hydrophobic stretch, instead of the conserved PilA H-domain, does not rescue the adhesion defect of the Δ*pilA*[*1-6*] strain nor does it result in the biosynthesis of pili in this deletion strain (Fig. 7A; Esquivel *et al.*, 2013). To assess the importance of the conserved H-domain for motility regulation, we stab-inoculated the Δ*pilA*[*1-6*] strain expressing PilA1Hyb in

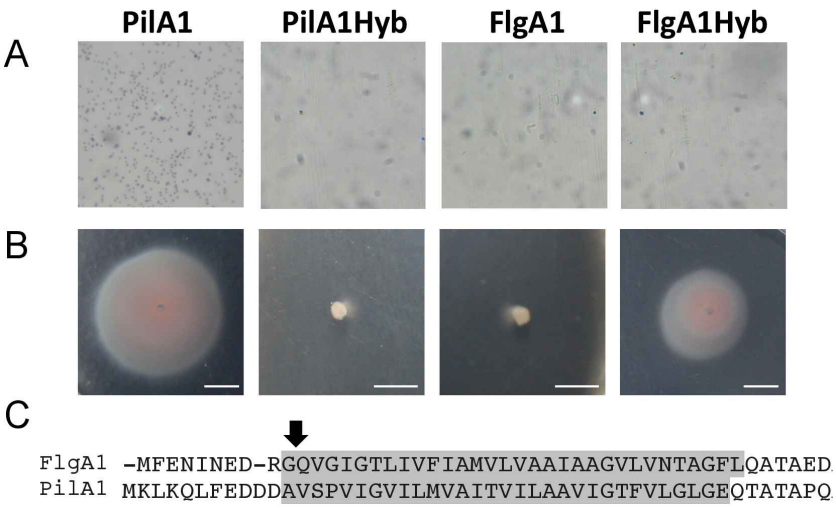


Fig. 7. The PilA1-A6 conserved H-domain is critical for pilin-dependent motility regulation.
A and B. (A) Adhesion assays (×35 magnification) and (B) motility assays of *H. volcanii* Δ*pilA*[*1-6*] strain transformed with pTA963 encoding *pilA1His*, *pilA1HybHis*, *flgA1His* or *flgA1HybHis*. A spot on the microscope lens is seen in FlgA1 and FlgA1Hyb.
C. N-terminal amino acid sequence of *H. volcanii* FlgA1 and PilA1. PilA1 H-domain used to construct the FlgA1Hybrid and FlgA1 H-domain replaced in that flagellin hybrid, are highlighted in gray. Arrow indicates predicted PibD processing site. Bars, 5 mm.

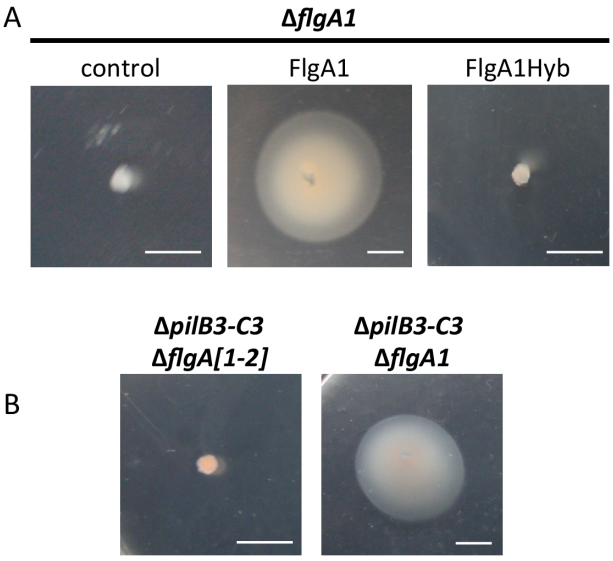


Fig. 8. FlgA1Hyb cannot make functional flagella.
A. Motility assays of Δ*flgA1* strain transformed with pTA963 (control) or pTA963 encoding *flgA1His* or *flgA1HybHis*.
B. Motility assays of Δ*pilB3-C3*Δ*flgA[1-2]* and Δ*pilB3-C3*Δ*flgA1* transformed with pTA963, demonstrating that FlgA2 can support motility in an FlgA1-independent manner, in the absence of *pilB3-C3*. Bars, 5 mm.

trans into motility plates. Unlike PilA1 expressed in trans this hybrid construct cannot complement the Δ*pilA*[*1-6*] motility defect, suggesting that multiple pilin functions critically depend on the H-domain (Fig. 7B).

To determine whether the hydrophobic stretch can only regulate motility in the context of a mature pilin, we created a hybrid flagellin containing the conserved pilin H-domain (Fig. 7C). When expressed in a Δ*pilA*[*1-6*] background, this hybrid protein, FlgA1Hyb, as expected, does not rescue the adhesion defect (Fig. 7A). However, it does rescue motility in the Δ*pilA*[*1-6*] strain, albeit not to wild-type levels, indicating the conserved H-domain can regulate swimming motility independent of the C-terminal pilin sequence (Fig. 7B).

To confirm that the His-tagged FlgA1Hyb construct could not make functional flagella, we expressed this hybrid flagellin in the Δ*flgA1* strain, which is non-motile. We have previously shown that this deletion strain can be complemented by expressing FlgA1His in trans (Tripepi *et al*., 2010). Unlike FlgA1His, expression of FlgA1HybHis does not complement the motility defect in the Δ*flgA1* strain (Fig. 8A), probably because the PilA H-domain is not compatible with the flagella assembly machinery.

Surprisingly, a Δ*flgA1* strain became motile upon the deletion of *pilB3-C3* when grown on semi-defined media rather than complex media (Fig. 8B and Tripepi *et al*., 2013). Possibly, similar to the overexpression of pilins, deletion of pilus-biosynthesis genes results in an increased number of pilins in the membrane, hence promoting motility, and allowing FlgA2, in the absence of FlgA1 to confer motility. Multiple mechanisms seem to be involved in regulating *H. volcanii* motility. This could explain why an increase in motility is not observed in a Δ*pilB3-C3* strain but is observed upon the deletion of *pilB3-C3* and *flgA1* (Fig. 8B) or overexpression of pilins in the wild-type strain (Fig. 3 and Fig. S2). While a Δ*flgA2* mutant is hypermotile, suggesting a regulatory role for this flagellin (Tripepi *et al*., 2013), these results demonstrate that FlgA2 is sufficient to make functional flagella, which has not been observed before.

FlgA2 and pilin motility regulation are independent

Considering the Δ*flgA2* hypermotility phenotype and the pilin-like signal peptide structure of archaeal flagellins, we had previously hypothesized that the PilA H-domain might interact with the FlgA2 H-domain in the cell membrane, sequestering inhibitory concentrations of FlgA2. While this FlgA1-independent motility indicates that the FlgA2 and pilin-dependent motility regulation are independent, these results do not exclude the possibility that pilin-FlgA2 interactions are occurring in the presence of FlgA1. If the Δ*pilA*[*1-6*] strain had a severe motility defect because FlgA2 was not sequestered, the Δ*pilA*[*1-6*] in which *flgA2* was also deleted should be motile. However, our results show that the deletion of the six pilins in the absence of FlgA2 (Fig. S4B), still results in a severe motility defect, supporting two novel, independent regulatory mechanisms (Fig. 9).

Conclusions

Most prokaryotic cells can reversibly exist as either motile planktonic cells or as sessile cells in biofilms. These cells require flagella to swim in liquid media, and many use type

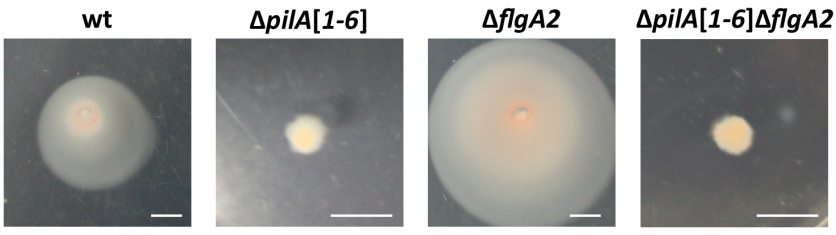


Fig. 9. PilA pilin motility regulation is FlgA2-independent. Motility assays of wt and deletion strains in MGM. Bars, 5 mm.

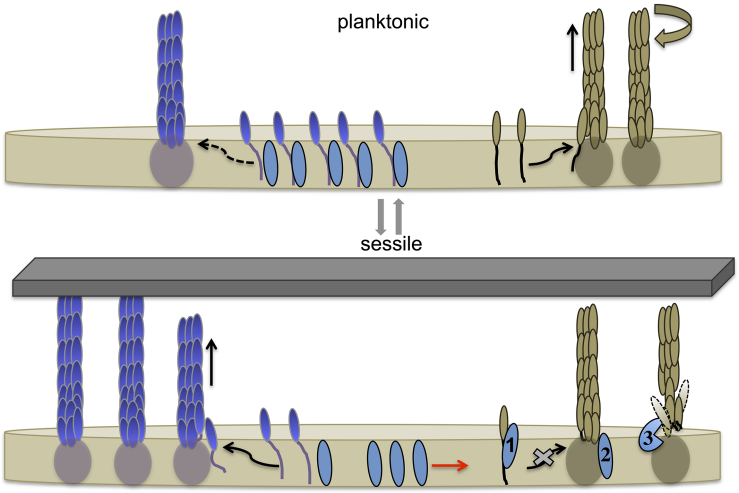


Fig. 10. Model for pilin-mediated inhibition of swimming motility. During planktonic growth *H. volcanii* cells synthesize flagellins that are readily incorporated into flagella, supporting swimming motility. The cells also express pilins, which are incorporated into pili at a slow rate. The H-domain of membrane-associated pilins interacts with, and hence sequesters, a protein that directly or indirectly inhibits flagella motility. Upon adhesion, pilus-assembly kinetics shifts and the affinity of pilins for pili increases, depleting the membrane of pilins and releasing the inhibitor proteins. These inhibitors interfere with flagella biosynthesis and/or stability. Taken together, this allows cells to rapidly respond to environmental conditions that favour biofilm formation rather than motility. Three possible mechanisms through which an inhibitor might obstruct swimming motility are: (1) direct interaction with flagellins, thus preventing the incorporation of subunits into the flagellum; (2) inactivation of a flagella-biosynthesis component(s) or (3) degradation/destabilization of the flagella.

IV pili to mediate attachment to surfaces. Therefore, it is not surprising that cells have evolved transcriptional and post-transcriptional mechanisms to tightly regulate the biosynthesis and function of pili and flagella during the transition between a motile and sessile state. We have demonstrated here, for the first time, that type IV pilins themselves can be involved in the post-translational regulation of flagella-dependent motility. Although other modes of regulating flagella-dependent motility likely play a role in regulating *H. volcanii* motility, the novel post-translational pilin-dependent regulatory mechanism reported in this study is unique in allowing a rapid transition of cells from a sessile to a planktonic state when local environmental conditions change. Our data show that the highly conserved pilin H-domain, which is required for membrane targeting and pilus biosynthesis, is also critical for this motility regulation. It may be that this functional role has resulted in the 100% conservation of its amino acid sequence across the H-domain of all six, otherwise diverse, adhesion pilins. While a construct only expressing the H-domain was unstable, we were able to complement motility using a FlgA1Hyb protein lacking the C-terminal pilin domain. Localization studies of this construct also suggest that the H-domain does not promote motility by incorporating into the flagella structure. Perhaps the adhesion pilins sequester another non-flagellin protein that directly or indirectly inhibits motility. We have also observed that *H. volcanii* synthesizes flagella as well as pili during planktonic growth and that subunits of these surface filaments can be detected in the membrane of planktonic cells (Tripepi *et al.*, 2013). Given these data we propose the model described in Fig. 10.

We were able to identify this regulatory mechanism because the surface adhesion of *H. volcanii* is not affected unless at least five of the six genes encoding adhesion pilins are deleted (Esquivel *et al.*, 2013), which results in a strain having a cell membrane that is depleted of adhesion pilins. To the best of our knowledge, in studies of pili in other prokaryotic systems no attempts have been made to obtain strains depleted of all major and minor adhesion pilins involved in the biosynthesis of pili that are required for surface adhesion. Hence, our results, which suggest a novel system for the regulation of expression of surface filaments during biofilm formation, might also be identified in other archaea and bacteria when the corresponding strains, lacking all pilins involved in adhesion, are generated. A closer examination of domains conserved between pilins might ultimately lead to the determination that this regulatory mechanism is used across a wide variety of prokaryotic species.

Future studies will focus on the identification of the predicted pilin-interacting motility inhibitor using co-immunoprecipitation studies as well as genetic selections. Moreover, using the Δ*pilB3-C3* knockout strains as well as hybrid pilins and flagellins constructed for this study, will shed light on the details of the initial interactions between pilins and PilB/C, a step in bacterial pilus biosynthesis that has long remained elusive.

Experimental procedures

Strains and growth conditions

The plasmids and strains used in this study are listed in Table S1. *H. volcanii* H53 and its derivatives were grown at 45°C in liquid or on solid agar (1.5% w/v) semi-defined Casamino Acid (CA) medium, supplemented with tryptophan and uracil (50 μg ml^{-1} final concentration) or complex Modified Growth Medium (MGM). Strains transformed with pTA963 are grown on CA medium supplemented with tryptophan (50 μg ml^{-1} final concentration) (Dyall-Smith, 2004). For selection of the deletion mutants (see below), 5-FOA was added at a final concentration of 150 μg ml^{-1} in CA medium, and uracil was added to a final concentration of 10 μg ml^{-1}. Strain H53 and the deletion mutants transformed with pTA963 or its derivatives were grown in CA medium supplemented with tryptophan. *Escherichia coli* strains were grown at 37°C in NZCYM medium supplemented with ampicillin (200 μg ml^{-1}) (Blattner *et al*., 1977).

Generation of chromosomal deletions

Chromosomal deletions were generated by using a homologous recombination (pop-in pop-out) method previously described by Allers and Ngo (2003). Plasmid constructs were generated using overlap PCR as described previously by Hammelmann and Soppa (2008) and modified as described in Tripepi *et al*. (2010). To confirm the chromosomal replacement event occurred at the proper location on the chromosome, the genomic DNA isolated from colonies derived using these techniques was screened by PCR. The identities of the PCR products were verified by sequencing using the primers lying outside the gene of interest (primers used are listed in Table S2).

Surface adhesion assay

H. volcanii surface adhesion was assayed using a modified air-liquid interface (ALI) assay (O'Toole *et al*., 1999) as described in Esquivel *et al*. (2013). Briefly, 3 ml of culture in CA medium supplemented with tryptophan and/or uracil as necessary, at an optical density of 600 nm (OD_{600}) of ~ 0.3, was incubated in each well of a 12-well plate. Plastic coverslips (22 by 22 mm; 0.19 to 0.25 mm thick) were inserted into each well and incubated overnight at 45°C without shaking. Upon acetic acid fixing, coverslips were stained in 0.1% w/v crystal violet solution for 10 min. The coverslips were then washed with distilled water, air-dried and examined using light microscopy.

Motility assay

Motility assays were performed by stab-inoculating *H. volcanii* into motility plates containing 0.3% w/v agar as described in Tripepi *et al*. (2010) and incubated for 3 days at 45°C, unless otherwise noted.

Isolation and purification of surface filaments

The isolation of *H. volcanii* flagella or type IV pili supernatant fractions was performed by CsCl gradient purification as described previously by Fedorov (1994), with modifications as described in Tripepi *et al*. (2013). Briefly, to select for motile cells, colonies from a solid-agar plate were stab-inoculated into motility plates and cells from the outer motility ring, formed after 3 days, were inoculated into 5 ml CA liquid medium. Two litre of CA medium were inoculated with this 5-ml culture, and the cultures were harvested at an OD_{600} of approximately 0.3 by centrifugation at 8700 rpm (JA-10 rotor; Beckman) for 30 min. The supernatant was centrifuged again (8700 rpm for 30 min) and incubated at room temperature with 4% w/v polyethylene glycol (PEG) 6000 for 1 h. The PEG-precipitated proteins were then centrifuged at 16 000 rpm (JLA-16.250 rotor; Beckman) for 50 min at 4°C, and the surface filaments were purified by cesium chloride (CsCl) density gradient centrifugation (overnight centrifugation at 50 000 rpm) (VTI-65.1 rotor; Beckman). CsCl was dissolved in a 3 M NaCl saline solution to a final density of 1.37 g cm^{-1}.

Protein extraction, LDS-PAGE and Western blotting

Protein from cell pellets, TCA-precipitated supernatants, or surface filament containing CsCl fractions of *H. volcanii* strains were separated by LDS gel chromatography and stained by Coomassie or, in strains expressing His-tagged constructs, analysed by western blot using anti-His antibodies as described in Tripepi *et al*. (2012). Liquid cultures were grown until the early-log phase (OD_{600} ~ 0.3). Cells were collected by centrifugation at 4300 *g* for 10 min at 4°C. Cell pellets were resuspended and lysed in 1% v/v NuPAGE lithium dodecyl sulphate (LDS) supplemented with 50 mM dithiothreitol (DTT). The supernatants of relevant strains containing secreted proteins were precipitated with cold trichloroacetic acid (TCA) (10%, v/v), and pellets were washed twice with cold acetone (80%, v/v) and then resuspended in 1% LDS buffer supplemented with 50 mM DTT. The electrophoresis of protein samples was performed with 12% v/v Bis-Tris NuPAGE gels under denaturing conditions using morpholinepropanesulphonic acid (MOPS) buffer at pH 7.7. Proteins were transferred from the gel onto a polyvinylidene difluoride membrane using a Bio-Rad Transblot-SD semidry transfer cell at 15 V for 30 min. Western blots of whole-cell lysates of strains expressing C-terminally His-tagged constructs with a three amino acid linker sequence were probed with an anti-His antibody at a dilution of 1:1000, followed by a secondary anti-mouse antibody at a dilution of 1:10 000. Antibody-labelled protein bands were identified using the Amersham ECL Plus Western blotting detection system.

Electron microscopy

H. volcanii whole cells and CsCl gradient fractions were prepared as described in Tripepi *et al*. (2013). Cell cultures were fixed in CA medium with 2% v/v glutaraldehyde and 1% v/v paraformaldehyde for 1 h. Ten microlitres of the fixed culture was put onto glow-discharged copper grids coated with carbon-Formvar for 10 min. The grids were rinsed two times in ddH_2O and negatively stained using 1% w/v uranyl formate. Grids were then analysed using a Philips Tecnai 12 operating at 120 kV, and a 0Gatan US1000 2K × 2K camera. CsCl gradient density fractions were applied onto the glow

discharged grids and were left for 5 min at room temperature, washed with water, blotted with a filter paper, and stained with 2% w/v uranyl acetate for 10 s. Grids were then analysed using a Philips Tecnai 12 instrument operating at 120 kV and a Gatan US1000 2K-by-2 K (2024- by 2024-pixel) camera.

Acknowledgements

M.P. and R.E. were supported by National Aeronautics and Space Administration grant NNX10AR84G. R.E. was also supported by the National Institutes of Health Cell and Molecular Biology Training grant TM32 GM-07229.We thank Dewight Williams for advice on microscopy and Dieter Schifferli, Jay Zhu, Friedhelm Pfeiffer, Michael Donnenberg, Ben Garcia, Fevzi Daldal and the Pohlschroder lab for helpful discussions.

References

Albers, S.V., and Pohlschroder, M. (2009) Diversity of archaeal type IV pilin-like structures. *Extremophiles* **13:** 403–410.

Albers, S.V., Szabo, Z., and Driessen, A.J. (2003) Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J Bacteriol* **185:** 3918–3925.

Allers, T., and Ngo, H.P. (2003) Genetic analysis of homologous recombination in Archaea: *Haloferax volcanii* as a model organism. *Biochem Soc Trans* **31:** 706–710.

Bardy, S.L., and Jarrell, K.F. (2003) Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus voltae*. *Mol Microbiol* **50:** 1339–1347.

Blair, K.M., Turner, L., Winkelman, J.T., Berg, H.C., and Kearns, D.B. (2008) A molecular clutch disables flagella in the *Bacillus subtilis* biofilm. *Science* **320:** 1636–1638.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L., *et al.* (1977) Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. *Science* **196:** 161–169.

Boyd, C.D., and O'Toole, G.A. (2012) Second messenger regulation of biofilm formation: breakthroughs in understanding c-di-GMP effector systems. *Annu Rev Cell Dev Biol* **28:** 439–462.

Craig, L., Volkmann, N., Arvai, A.S., Pique, M.E., Yeager, M., Egelman, E.H., and Tainer, J.A. (2006) Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. *Mol Cell* **23:** 651–662.

Dyall-Smith, M. (2004) The halohandbook – protocols for haloarchaeal genetics. [WWW document]. URL http://www.haloarchaea.com

Esquivel, R.N., Xu, R., and Pohlschroder, M. (2013) Novel archaeal adhesion pilins with a conserved N terminus. *J Bacteriol* **195:** 3808–3818.

Fedorov, O.V. (1994) Protofilament as a structural element of flagella of haloalkalophilic archaebacteria. *Can J Microbiol* **40:** 45–53.

Fröls, S., Ajon, M., Wagner, M., Teichmann, D., Zolghadr, B., Folea, M., *et al.* (2008) UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation. *Mol Microbiol* **70:** 938–952.

Ghosh, A., and Albers, S.V. (2011) Assembly and function of the archaeal flagellum. *Biochem Soc Trans* **39:** 64–69.

Giltner, C.L., Nguyen, Y., and Burrows, L.L. (2012) Type IV pilin proteins: versatile molecular modules. *Microbiol Mol Biol Rev* **76:** 740–772.

Guttenplan, S.B., and Kearns, D.B. (2013) Regulation of flagellar motility during biofilm formation. *FEMS Microbiol Rev* **37:** 849–871.

Hammelmann, M., and Soppa, J. (2008) Optimized generation of vectors for the construction of *Haloferax volcanii* deletion mutants. *J Microbiol Methods* **75:** 201–204.

Hansen, S.K., Rainey, P.B., Haagensen, J.A., and Molin, S. (2007) Evolution of species interactions in a biofilm community. *Nature* **445:** 533–536.

Hartman, A.L., Norais, C., Badger, J.H., Delmas, S., Haldenby, S., Madupu, R., *et al.* (2010) The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE* **5:** e9605.

Haussler, S., and Fuqua, C. (2013) Biofilms 2012: new discoveries and significant wrinkles in a dynamic field. *J Bacteriol* **195:** 2947–2958.

Henche, A.L., Ghosh, A., Yu, X., Jeske, T., Egelman, E., and Albers, S.V. (2012) Structure and function of the adhesive type IV pilus of *Sulfolobus acidocaldarius*. *Environ Microbiol* **14:** 3188–3202.

Jarrell, K.F., and Albers, S.V. (2012) The archaellum: an old motility structure with a new name. *Trends Microbiol* **20:** 307–312.

Karatan, E., and Watnick, P. (2009) Signals, regulatory networks, and materials that build and break bacterial biofilms. *Microbiol Mol Biol Rev* **73:** 310–347.

Kuchma, S., Kimberly, M.B., Judith, H.M., Nicole, T.L., Frederick, M.A., and O'Toole, G.A. (2007) BifA, a cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *J Bacteriol* **189:** 8165–8178.

Large, A., Stamme, C., Lange, C., Duan, Z., Allers, T., Soppa, J., and Lund, P.A. (2007) Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *cct1* gene. *Mol Microbiol* **66:** 1092–1106.

Lassak, K., Ghosh, A., and Albers, S.V. (2012) Diversity, assembly and regulation of archaeal type IV pili-like and non-type-IV pili-like surface structures. *Res Microbiol* **163:** 630–644.

Lassak, K., Peeters, E., Wrobel, S., and Albers, S.V. (2013) The one-component system ArnR: a membrane-bound activator of the crenarchaeal archaellum. *Mol Microbiol* **88:** 125–139.

McDougald, D., Rice, S.A., Barraud, N., Steinberg, P.D., and Kjelleberg, S. (2011) Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. *Nat Rev Microbiol* **10:** 39–50.

Monds, R.D., and O'Toole, G.A. (2009) The developmental model of microbial biofilms: ten years of a paradigm up for review. *Trends Microbiol* **17:** 73–87.

Orell, A., Frols, S., and Albers, S.V. (2013a) Archaeal biofilms: the great unexplored. *Annu Rev Microbiol* **67:** 337–354.

Orell, A., Peeters, E., Vassen, V., Jachlewski, S., Schalles,

S., Siebers, B., and Albers, S.V. (2013b) Lrs14 transcriptional regulators influence biofilm formation and cell motility of Crenarchaea. *ISME J* **7:** 1886–1898.
O'Toole, G.A., and Kolter, R. (1998) Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Mol Microbiol* **30:** 295–304.
O'Toole, G.A., Pratt, L.A., Watnick, P.I., Newman, D.K., Weaver, V.B., and Kolter, R. (1999) Genetic approaches to study of biofilms. *Methods Enzymol* **310:** 91–109.
Pohlschroder, M., Ghosh, A., Tripepi, M., and Albers, S.V. (2011) Archaeal type IV pilus-like structures – evolutionarily conserved prokaryotic surface organelles. *Curr Opin Microbiol* **14:** 357–363.
Reimann, J., Lassak, K., Khadouma, S., Ettema, T.J., Yang, N., Driessen, A.J., *et al*. (2012) Regulation of archaella expression by the FHA and von Willebrand domain-containing proteins ArnA and ArnB in *Sulfolobus acidocaldarius*. *Mol Microbiol* **86:** 24–36.
Szabó, Z., Stahl, A.O., Albers, S.V., Kissinger, J.C., Driessen, A.J., and Pohlschröder, M. (2007) Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases. *J Bacteriol* **189:** 772–778.
Takhar, H.K., Kemp, K., Kim, M., Howell, P.L., and Burrows, L.L. (2013) The platform protein is essential for type IV pilus biogenesis. *J Biol Chem* **288:** 9721–9728.
Tripepi, M., Imam, S., and Pohlschroder, M. (2010) *Haloferax volcanii* flagella are required for motility but are not involved in PibD-dependent surface adhesion. *J Bacteriol* **192:** 3093–3102.
Tripepi, M., You, J., Temel, S., Onder, O., Brisson, D., and Pohlschroder, M. (2012) N-glycosylation of *Haloferax volcanii* flagellins requires known Agl proteins and is essential for biosynthesis of stable flagella. *J Bacteriol* **194:** 4876–4887.
Tripepi, M., Esquivel, R.N., Wirth, R., and Pohlschroder, M. (2013) *Haloferax volcanii* cells lacking the flagellin FlgA2 are hypermotile. *Microbiology* **159:** 2249–2258.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Week 8

Yeast

Archaea

Cell, Vol. 21, 205–215, August 1980, Copyright © 1980 by MIT

Identification of 23 Complementation Groups Required for Post-translational Events in the Yeast Secretory Pathway

Peter Novick, Charles Field and Randy Schekman*
Department of Biochemistry
University of California, Berkeley
Berkeley, California 94720

Summary

Cells of a Saccharomyces cerevisiae mutant that is temperature-sensitive for secretion and cell surface growth become dense during incubation at the nonpermissive temperature (37°C). This property allows the selection of additional secretory mutants by sedimentation of mutagenized cells on a Ludox density gradient. Colonies derived from dense cells are screened for conditional growth and secretion of invertase and acid phosphatase. The *sec* mutant strains that accumulate an abnormally large intracellular pool of invertase at 37°C (188 mutant clones) fall into 23 complementation groups, and the distribution of mutant alleles suggests that more complementation groups could be found. Bud emergence and incorporation of a plasma membrane sulfate permease activity stop quickly after a shift to 37°C. Many of the mutants are thermoreversible; upon return to the permissive temperature (25°C) the accumulated invertase is secreted. Electron microscopy of *sec* mutant cells reveals, with one exception, the temperature-dependent accumulation of membrane-enclosed secretory organelles. We suggest that these structures represent intermediates in a pathway in which secretion and plasma membrane assembly are colinear.

Introduction

Studies of the secretory process in eucaryotic cells have focused on the molecular events associated with synthesis and processing of specific secretory and membrane proteins, and on the organelles which mediate passage from the endoplasmic reticulum to the cell surface. While much is known about the maturation of certain secretory and membrane polypeptides (such as insulin, Chan, Keim and Steiner, 1976; VSV glycoprotein, Katz et al., 1977), the molecular events associated with sorting, packaging, transport and exocytosis of the exported proteins remain obscure.

We have undertaken a study of the secretory apparatus in the yeast Saccharomyces cerevisiae. While yeast cells are not specialized for secretion as are, for example, the acinar cells of the pancreas (Palade, 1975), the ease of a combined genetic and biochemical approach allows the use of techniques which have been less feasible with traditional secretory systems. The use of conditional mutants has been crucial for the analysis of complex bacteriophage morphogenesis pathways, both in identifying intermediate structures and in providing biochemical assays for assembly steps (Wood and King, 1979). We believe that a similar approach may be useful in unraveling a eucaryotic morphogenesis pathway.

Yeast cell surface growth is restricted primarily to enlargement of the bud followed by cell division. Incorporation of new cell wall material, including secretion of the wall-bound enzymes invertase and acid phosphatase, is also restricted to the bud (Tkacz and Lampen, 1972, 1973; Field and Schekman, 1980). Membrane-enclosed vesicles have been implicated in secretion and bud growth (Moor, 1967; Matile et al., 1971). Our recent report of a conditional mutant blocked in secretion and cell surface growth, which accumulates membrane-enclosed vesicles containing a secretory enzyme (Novick and Schekman, 1979), supports such a role for vesicles.

In this report we describe a technique for the enrichment of conditional secretory and cell surface growth mutants. We have identified a large number of complementation groups that are required for the movement of at least two secretory enzymes and one plasma membrane permease through a series of distinct membrane-enclosed organelles in a pathway that leads to the cell surface.

Results

Secretory mutants are defined as those strains which fail to export active invertase and acid phosphatase, but continue to synthesize protein under restrictive growth conditions. In a previous report (Novick and Schekman, 1979) we described a screening procedure that allowed the identification of two nonallelic secretory mutants (*sec*1-1, *sec*2-1) among a group of randomly selected temperature-sensitive yeast mutants. The first mutant (HMSF 1) stopped dividing and enlarging at the nonpermissive temperature (37°C), yet protein and phospholipid synthesis continued for at least 3 hr. This situation produced dense cells. Henry et al. (1977) showed that during inositol starvation of an auxotrophic strain, net cell surface growth stopped while cell mass increased. Starved cells could be resolved from normal cells on a Ludox density gradient. We have used the Ludox density gradient technique to select additional secretory mutants.

Density Enrichment

In the experiment shown in Figure 1, about 5×10^6 *sec*1-1 cells were mixed with 5×10^8 X2180 cells, and after 3 hr at 37°C the mixture was sedimented in a solution of Ludox. The resulting gradient was fractionated and the genotype of cells in diluted aliquots was determined. The 5% increase in density allowed *sec*1 cells to be separated completely from a 100 fold larger population of wild-type cells.

The density separation made feasible the isolation

* To whom correspondence should be addressed.

of a large number of additional secretory mutants. Mutagenized cultures were allowed to grow for several generations at 24°C and then transferred to 37°C for 3 hr. The cells were then sedimented in a Ludox gradient, and the densest 1–2% of the cells were pooled. Temperature-sensitive growth mutants were identified among the dense cells, and the secretion of acid phosphatase and invertase was measured by a modification of the previous procedure (Novick and Schekman, 1979). The density gradient procedure enriched for a variety of temperature-sensitive mutants and among them about 15% were secretion-defective (Table 1).

The procedure was used on three strains; NF1R and SF182-3B were mutagenized with ethyl methanesulfonate (EMS) and X2180-1A was treated with nitrous acid. A total of 485 secretion-defective mutants were isolated, among which three classes were found. Class A *sec* mutants (188 total) showed accumulation of invertase at the nonpermissive temperature. Class B *sec* mutants showed no accumulation of active invertase, although protein synthesis continued at a high rate at 37°C. Class C mutants did not secrete because protein synthesis was temperature-sensitive. This report will deal with the *sec*A mutants; analysis of the *sec*B mutants is in progress.

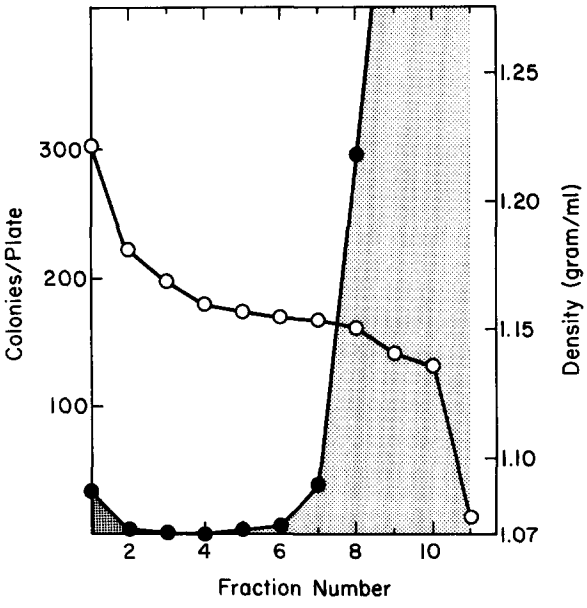


Figure 1. Density Gradient Separation of *sec*1-1 and X2180 Cells

SF150-5C and X2180-1A cells were grown in YPD medium at 25°C. After 3 hr at 37°C, the cells were sedimented, washed and resuspended in 1 ml of water. Cell aliquots were mixed (49.5 A_{600} units of X2180 and 0.5 A_{600} unit of SF150-5C) and sedimented on a Ludox gradient. Fractions were collected and diluted 10^4 fold, and 0.1 ml portions were spread on minimal medium (7 mM phosphate) agar plates. Colonies formed after 2 days at room temperature were stained for acid phosphatase activity (Hansche et al., 1978). Heavy stippling represents the phosphatase-constitutive (*pho*80) *sec*1 colonies (SF150-5C); light stippling represents phosphatase-repressed colonies (X2180).

The enrichment procedure was first performed on an *a* strain (NF1R) and then on an α strain (SF 182-3B). The *sec* mutants were arranged into complementation groups by standard genetic techniques. The enrichment was repeated with strain X2180-1A, and complementation analysis of the temperature-sensitive clones was performed with tester strains derived from the previously isolated *sec* mutants. New *sec* mutants that complemented all of the tester strains were crossed with X2180-1B, diploids were selected and sporulated, and the *sec* mutants were obtained in *a* and α mating type strains. Complementation analysis was performed again. By this procedure, 23 complementation groups were identified. EMS and nitrous acid produced a similar spectrum of mutant alleles (Table 2); *sec*2 and *sec*5 were the most common groups for each mutagen.

All the *sec* mutants were recessive in heterozygous diploids. Analysis of each complementation group was conducted with a representative allele chosen for optimum growth at 25°C, maximum inhibition of secretion at 37°C and maximum secretion of accumulated invertase upon return to 25°C. Each group showed 2:2 segregation of the temperature-sensitive phenotype which always coincided with the secretory defect.

Thermoreversible Accumulation of Invertase

External invertase synthesis is derepressed by a decreased supply of glucose, and the secreted enzyme remains in the yeast cell wall and can be assayed in a whole-cell suspension (Dodyk and Rothstein, 1964). Intracellular forms of the enzyme are measured in a spheroplast lysate, and this level does not rise significantly during derepression (Novick and Schekman, 1979).

The level of secreted invertase increased 13 fold when X2180 cells were transferred from YP + 5% glucose medium to YP + 0.1% glucose medium (Table 3). This increase was blocked to varying degrees in the *sec* mutants at 37°C. Derepression for 1 hr at 37°C produced a 4–18 fold increase in the intracellular level of invertase. Upon return to 25°C, in the

Table 1. Comparison of Screening Procedure with and without Density Enrichment

	Without Enrichment		With Enrichment	
Screening Stage	Colonies	%	Colonies	%
(1) Colonies tested	5,600	100	18,500	100
(2) TS mutants	291	5.2	2,830	15
(3) TS phosphatase secretion	63	1.1	980	5
(4) TS invertase secretion	16	.29	485	2.6
(5) TS invertase accumulation	2	.04	188	1.0

Table 2. Distribution of Mutants in the *sec*A Complementation Groups: EMS versus Nitrous Acid

	EMS		Nitrous Acid	
sec	Isolates	%	Isolates	%
1	8	11	4	3
2	28	39	41	35
3	3	4	0	0
4	7	10	2	2
5	10	14	16	14
6	3	4	3	3
7	1	1	3	3
8	6	8	4	3
9	3	4	4	3
10	1	1	2	2
11	1	1	11	9
12	1	1	3	3
13			4	3
14			4	3
15			2	2
16			2	2
17			1	1
18			2	2
19			1	1
20			1	1
21			1	1
22			4	3
23			1	1

presence of cycloheximide, all the mutants showed an increased level of secreted invertase. In most cases this represented the secretion of a large fraction of the accumulated enzyme. The mutants produced nearly normal levels of secreted and intracellular invertase when derepression was conducted at 25°C.

All 12 alleles of *sec*11 produced nearly normal levels of secreted invertase at 37°C. However, like the other mutants, *sec*11 accumulated about 7 fold more internal invertase than normal. Furthermore, 35% of the accumulated enzyme was secreted upon return to 25°C.

Other Defects

Secreted acid phosphatase first appeared in wild-type cells 1.5 hr after a transfer from YPD + 7 mM phosphate medium into phosphate-depleted YPD medium, and the rate of secretion was maximal from 2.5 to 5 hr after the shift. During this 2.5 hr period, the *sec* mutants secreted normally at 25°C, but produced at least 5 fold less external phosphatase at 37°C (Table 4).

Incorporation of a sulfate permease activity was used to assess the role of the *sec* gene products in plasma membrane assembly. In X2180, sulfate permease activity first appeared 30 min after cells were transferred from a minimal medium containing 1.5 mM methionine to a sulfate-free minimal medium. During a 2.5 hr period of derepression, most of the *sec* mutants produced normal levels of permease activity at 25°C, but showed significantly lower incorporation at 37°C (Table 4).

Bud emergence stopped quickly, as indicated by the nearly constant number of cells and buds, when the *sec* mutants were transferred from 25° to 37°C (Table 4). Cells arrested at all stages of the cell cycle, and no increase in cell size was noted. As expected from the enrichment procedure, all the *sec* mutants became denser than X2180 during a 3 hr incubation at 37°C (Table 4), although only a few of the strains became as dense as *sec*1. Certain other conditions, such as inhibition of protein synthesis or growth to stationary phase, caused X2180 cells to become dense at 37°C.

Mutants Accumulate Secretory Organelles

The reversible accumulation of invertase and the reduced incorporation of a membrane permease suggested that the *sec* mutants might accumulate an organelle of the secretory apparatus. This was confirmed for all but one of the mutants when thin sections were examined by electron microscopy. Wild-type cells grown at 37°C (Novick and Schekman, 1979, Figure 6A) or *sec* mutant cells grown at 25°C (Figure 2A) showed occasional enrichment of vesicles in the bud; short, thin tubules of endoplasmic reticulum (ER) were also seen apposed to the inner surface of the plasma membrane or in continuity with the nuclear membrane. Mutant cells incubated for 2 hr at 37°C showed several cytological aberrations. The groups were classified according to the organelle accumulated. The most common class, with 10 members (Table 5, Figure 2B), accumulated membrane-enclosed vesicles of 80–100 nm in diameter. These vesicles were not enriched in the bud.

A second class, with nine representatives, developed a more extensive network of ER than was seen in wild-type cells (Figure 3). The ER often lined the inner surface of the plasma membrane and wound through the cytoplasm where multiple connections with the nuclear membrane were visible (Figures 3A and 3B). The lumen of both the ER and the nuclear membrane was wider than the corresponding wild-type structure. Eucaryotic rough and smooth ER can be distinguished by the presence or absence of attached ribosomes. However, due to the high concentration of free ribosomes, it was not possible to identify specific associations with the yeast ER. While all nine members of this class showed extensive ER at 37°C, three complementation groups also produced small vesicles (~ 40 nm) which were often arranged in patches in the cytoplasm (Table 5, Figure 3C).

Table 3. Invertase Secretion and Accumulation by the *sec* Mutants

Strain	*sec* Group	Units/mg Dry Weight: External[a] (1 Hr 37°C)	Internal (1 Hr 37°C)	External (1 Hr 37°C → 3 Hr 25°C)	% Release[b]	External (1 Hr 25°C)	Internal (1 Hr 25°C)
X2180-1A		.38	.08	.33	0	.34	.14
HMSF 1	1-1	.02	.61	.28	43	.29	.15
HMSF 106	2-56	.03	.87	.36	38	.24	.18
HMSF 68	3-2	.02	.31	.05	9	.31	.22
HMSF 13	4-2	.05	.63	.13	11	.32	.30
HMSF 134	5-24	.03	.84	.08	6	.39	.17
HMSF 136	6-4	.03	.84	.46	52	.36	.14
HMSF 6	7-1	.04	.39	.10	16	.42	.29
HMSF 95	8-6	.03	.57	.07	7	.37	.22
HMSF 143	9-4	.09	1.05	.53	42	.20	.28
HMSF 147	10-2	.03	.68	.15	18	.31	.17
HMSF 154	11-7	.40	.53	.59	35	.56	.26
HMSF 162	12-4	.04	1.3	.90	64	.22	.11
HMSF 163	13-1	.19	.77	.64	58	.28	.14
HMSF 169	14-3	.07	.54	.32	46	.30	.12
HMSF 171	15-1	.17	.47	.33	34	.36	.19
HMSF 174	16-2	.06	1.50	.69	42	.45	.18
HMSF 175	17-1	.17	.70	.58	59	.29	.14
HMSF 176	18-1	.02	.97	.64	63	.36	.15
HMSF 178	19-1	.02	1.05	.49	45	.43	.21
HMSF 179	20-1	.07	.98	.68	63	.49	.24
HMSF 180	21-1	.29	.43	.48	44	.39	.18
HMSF 183	22-3	.04	.84	.58	64	.41	.18
HMSF 190	23-1	.03	.83	.57	65	.36	.15

[a] Cultures were grown overnight in YP + 5% glucose medium at 25°C to an A_{600} of 0.5–5.5. Cells (3 A_{600} units) were sedimented in a clinical centrifuge, resuspended in 3 ml of YP + 0.1% glucose medium and incubated at 37°C for 1 hr. An aliquot (1 ml) was then removed and added to a tube containing 1 mg of glucose and 0.1 mg of cycloheximide, and the mixture was incubated at 25°C for 3 hr. In a parallel experiment, 2 A_{600} units of the overnight culture were sedimented, and the cell pellet was resuspended in 2 ml of YP + 0.1% glucose medium and incubated at 25°C for 1 hr. At the end of each experiment samples were chilled on ice, centrifuged and resuspended in one half volume of 10 mM azide at 0°C.

[b] $\%\ \text{release} = \left(\frac{\text{Ext}_{37°C \rightarrow 25°C} - \text{Ext}_{37°C}}{\text{Int}_{37°C}}\right) 100.$

A third class of mutant produced an organelle with no obvious counterpart in other eucaryotic cells. Because of its unique morphology, we call this structure a Berkeley body (Bb). The Bb, although varied in form, appeared to consist of two curved membranes with an enclosed electron-transparent lumen (Figure 4). In some sections the Bb was closed to form a toroid; in other sections it was open at one end to form a cup (Figure 4B). The toroid structure, with enclosed ribosomes and cytoplasm, may be an alternate view of the cup form; perpendicular planes of sectioning would give the image of one or the other. Two complementation groups (*sec*7, 14) made predominantly Bbs; two alleles of *sec*7, derived from different parent strains, were examined and each produced only Bbs. In addition to Bbs, *sec*14 also produced 80–100 nm vesicles (Figure 5A). Bbs were occasionally seen in sections of two complementation groups (*sec*2, 9) where 80–100 nm vesicles were the dominant structure.

Two exceptions to the major classes were observed: *sec*19 produced a mixture of the major organelles (Figure 5B), and *sec*11 did not build up any of the organelles.

Discussion

The results presented here show that in yeast, at least 23 gene products are required for the transport of secretory proteins from the site of synthesis to the cell

Table 4. Acid Phosphatase Secretion, Sulfate Permease Incorporation, Cell Division and Cell Density of the *sec* Mutants

Strain	*sec*	Acid Phosphatase[a] (Units/ml)			Sulfate Permease[b] $\left(\frac{\text{Units}}{\text{mg Dry Weight}}\right)$		Cell Number[c] $\left(\frac{\text{2 Hr 37°C}}{\text{0 Hr}}\right)$	Cell Density[d] (g/ml)	
		2.5 Hr	5 Hr 37°C	5 Hr 25°C	25°C	37°C		25°C	37°C
X2180 1A		27	193	174	6.3	5.0	2.03	1.110	1.122
HMSF 1	1-1	27	28	147	7.4	.2	1.10	1.113	1.161
HMSF 106	2-56	50	48	170	5.8	.1	.92	1.109	1.141
HMSF 68	3-2	55	82	411	6.8	.3	1.10	1.109	1.142
HMSF 13	4-4	25	29	202	4.8	.2	1.07	1.116	1.146
HMSF 134	5-24	27	31	177	5.8	.4	1.04	1.110	1.161
HMSF 136	6-4	25	27	178	5.1	.1	1.20	1.111	1.159
HMSF 6	7-1	86	95	320	3.6	.03	.91	1.103	1.142
HMSF 16	8-1	44	81	223	4.5	.9	1.11	1.103	1.135
HMSF 143	9-4	30	30	177	6.1	.1	1.04	1.111	1.146
HMSF 147	10-2	25	47	70	7.2[e]	.03	.92	1.117	1.152
HMSF 154	11-7	16	15	107	6.1	2.0	1.48	1.117	1.142
HMSF 162	12-4	25	25	89	5.3[e]	.65	1.05	1.107	1.143
HMSF 163	13-1	22	18	154	4.2	.1	1.01	1.113	1.141
HMSF 169	14-3	25	26	121	5.3	1.5	.93	1.117	1.144
HMSF 171	15-1	23	41	190	6.7	1.7	1.15	1.117	1.159
HMSF 174	16-2	26	25	93	6.2	.02	.95	1.114	1.139
HMSF 175	17-1	32	30	200	6.1	.4	1.07	1.118	1.146
HMSF 176	18-1	27	25	202	5.6	.04	.93	1.117	1.155
HMSF 178	19-1	32	32	237	6.3	.3	.92	1.119	1.156
HMSF 179	20-1	23	21	136	1.5	.1	.94	1.116	1.141
HMSF 180	21-1	24	42	171	6.3	.2	1.08	1.113	1.142
HMSF 183	22-3	25	26	204	1.7	.1	1.09	1.115	1.151
HMSF 190	23-1	30	27	82	4.4	.1	1.02	1.114	1.145

[a] Cultures were grown overnight in YPD + 7 mm phosphate medium. Cells (4.5 A_{600} units) were centrifuged, resuspended in 3 ml of phosphate-depleted YPD medium and incubated at 25°C. After 2.5 hr, 1 ml aliquots were transferred to 37° and 0°C, and the rest were left at 25°C. Incubation was continued for 2.5 hr and the samples were chilled, centrifuged and resuspended in 1 ml of 10 mM azide at 0°C.

[b] Cultures were grown overnight in minimal medium + 1.5 mM methionine and 50 μM $(NH_4)_2SO_4$. Cells (1.2 A_{600} units) were centrifuged, washed once and resuspended in 1.2 ml of sulfate-free minimal medium. After 2.5 hr at 25° or 37°C, the tubes were chilled and 1 ml aliquots were removed for the permease assays; the rest were used for A_{600} determination.

[c] Cultures were grown overnight in YPD medium at 25°C. Cells (2 A_{600} units) were centrifuged and resuspended in 2 ml YPD medium; 0.5 ml was diluted with 10 mM azide at 0°C and the rest were incubated for 2 hr at 37°C. The ratio of the 2 hr/0 hr cell number is listed.

[d] Cultures were grown overnight in YPD medium. Cells (8 A_{600} units) were centrifuged and resuspended in 4 ml of YPD medium; 2 ml of each were incubated at 25° and 37°C. After 3 hr the cells were sedimented and resuspended in 0.5 ml of 10 mM azide at 0°C. X2180 cells grown to stationary phase at 37°C had a density of 1.131 g/ml; cells treated with 0.1 mg/ml of cycloheximide for 3 hr at 37°C had a density of 1.134 g/ml.

[e] HMSF 147 and 162 were auxotrophic and the sulfate permease experiment was carried out with prototrophic strains derived from crosses with X2180 (SF 226-1C, *sec*10; SF 292-2C, *sec*12).

surface. Thermosensitive defects in these gene products also block incorporation of a plasma membrane permease and stop bud growth. Taken together, these observations suggest that membrane growth and secretion are accomplished by parallel if not identical pathways. Furthermore, membrane-enclosed organelles accumulate in 22/23 of the mutants at 37° but not at 25°C. We propose that these structures are intermediates in the secretory pathway; their soluble contents are destined for secretion by exocytosis and their membranes will be incorporated into the plasma membrane by fusion.

In a previous report (Novick and Schekman, 1979), we described the detection of *sec*1-1 and *sec*2-1 in a collection of randomly selected temperature-sensitive mutants. No new *sec* mutants were found in a larger collection of temperature-sensitive strains, and therefore the density enrichment procedure was adopted. Although both of the original *sec* mutants become dense at 37°C and survive the enrichment, density

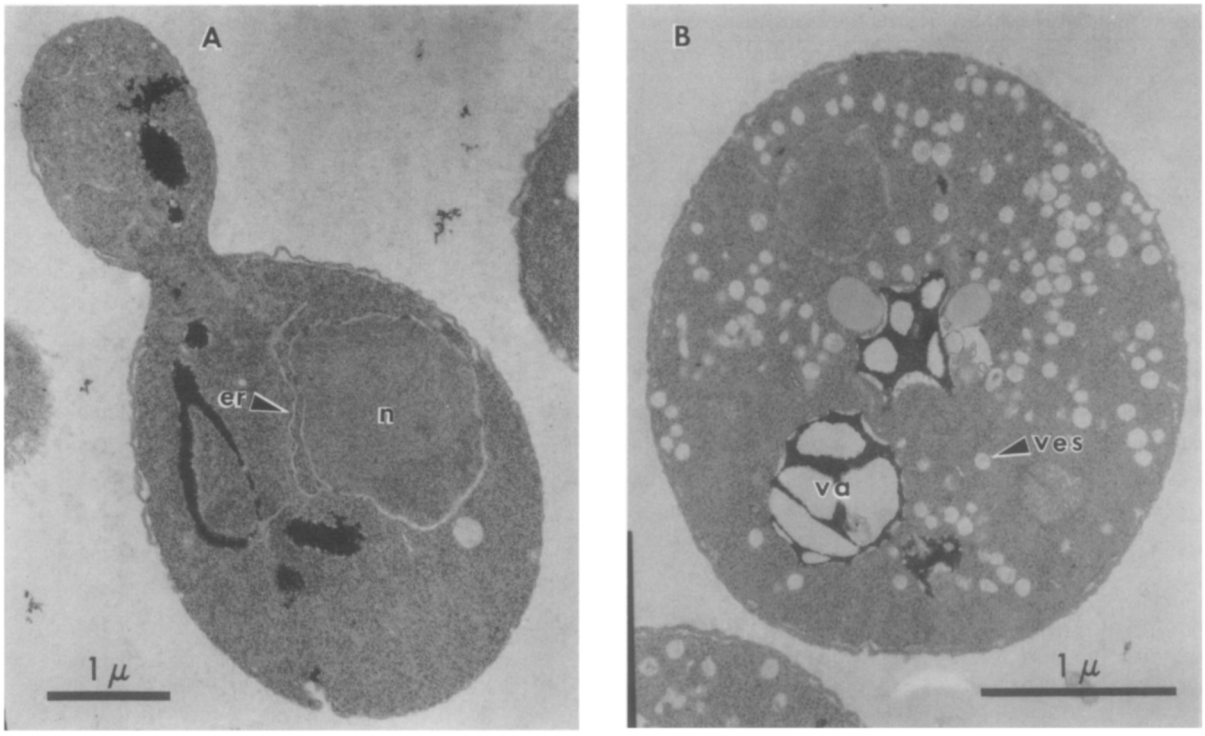


Figure 2. Thin Section Electron Micrographs of Cells Grown in YPD Medium

(A) HMSF 13 (*sec*4-2) grown at 25°C; (B) HMSF 171 (*sec*15-1) incubated at 37°C for 2 hr. Symbols: (n) nucleus; (va) vacuole; (er) endoplasmic reticulum; (ves) vesicles.

selection has the disadvantage that it eliminates mutants that die rapidly at 37°C. We assume that the density enrichment selects mutants that accumulate mass without a corresponding cell surface increase. The screening procedure, on the other hand, assumes only that secretion of cell wall mannoproteins is necessary for cell viability. For this reason, the density enrichment may eliminate mutants which fail to secrete but continue to expand, while the screening procedure will remove mutants which fail to enlarge but continue to secrete. Nevertheless, a large number of gene products are implicated in both aspects of cell surface growth.

It is likely that there are more than 23 *sec*A complementation groups required for the secretory process. Five of the groups reported here contain only one mutant allele, suggesting that groups exist for which no mutant has been found. Furthermore, the 23 *sec*A groups may represent fewer than 23 separate gene products; gene clusters coding for multifunctional proteins have been found in fungi and yeast. However, in such circumstances a single mutant frequently appears to fall into two otherwise distinct complementation groups (Fincham, 1977). No example of overlapping *sec* groups was found.

A trivial explanation for the large number of *sec* complementation groups is that mutant forms of various secreted proteins can act as inhibitors and block the passage of other cell surface molecules. Bassford and Beckwith (1979) and Bassford, Silhavy and Beck-

Table 5. Organelles Accumulated in the *sec* Strains

Strain (HMSF)	*sec*	Structure(s)
1	1-1	vesicles, Berkeley bodies
47	2-7	vesicles
3	3-1	vesicles
13	4-2	vesicles
81	5-8	vesicles
12	6-1	vesicles
6	7-1, -2	Berkeley bodies
93	8-4	vesicles
89	9-3	vesicles, Berkeley bodies
147	10-2	vesicles
154	11-7	
162	12-4	ER
163	13-1	ER
169	14-3	Berkeley bodies, vesicles
171	15-1	vesicles
174	16-2	ER
175	17-1	ER, small vesicles
176	18-1	ER, small vesicles
178	19-1	vesicles, Berkeley bodies, ER
179	20-1	ER
180	21-1	ER
183	22-3	ER, small vesicles
190	23-1	ER

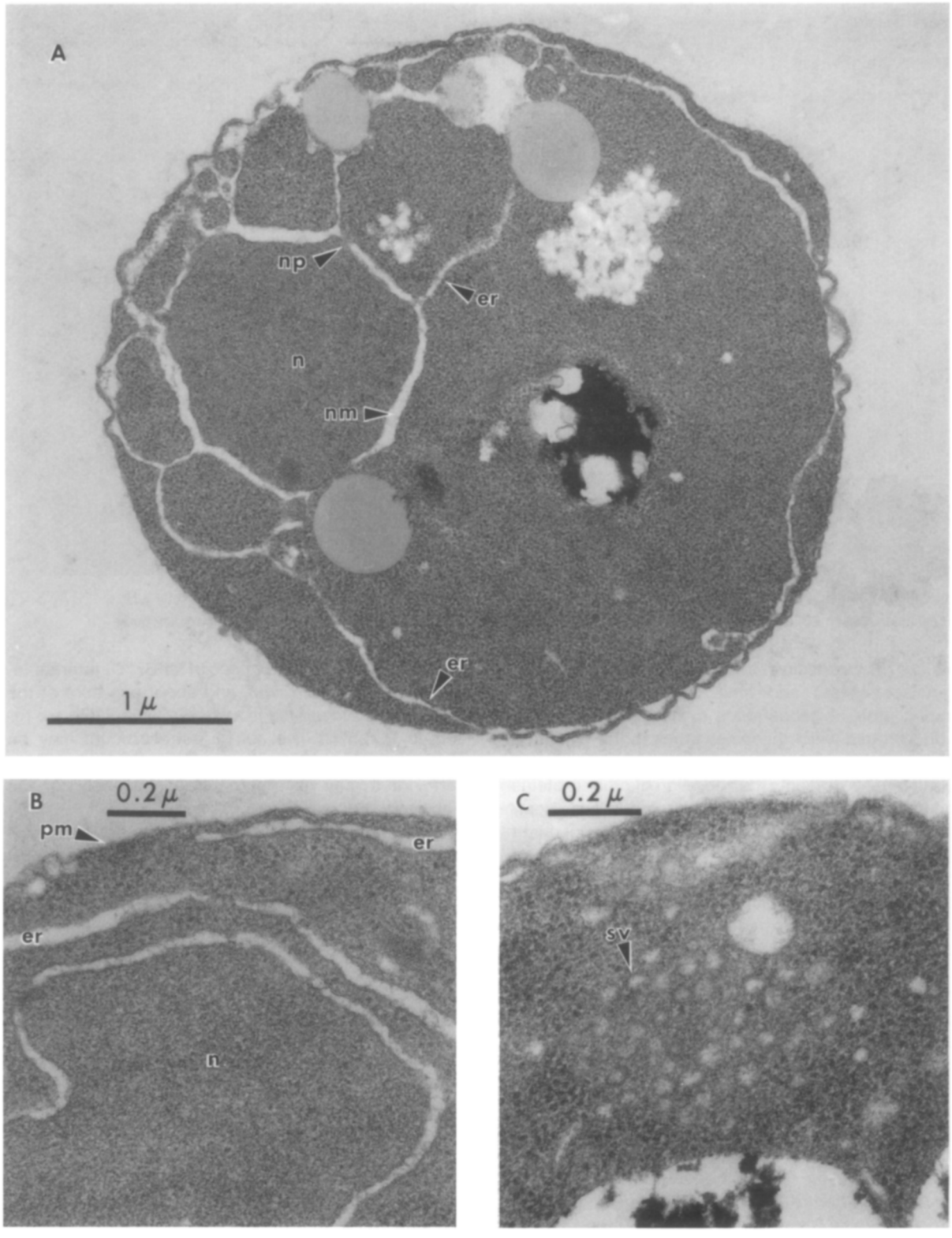


Figure 3. Thin Section Electron Micrographs of Cells Grown in YPD Medium at 25°C, Then Shifted to 37°C for 2 Hr

(A) HMSF 174 (*sec*16-2); (B) HMSF 190 (*sec*23-1); (C) HMSF 175 (*sec*17-1). Symbols are as in Figure 2 and (np) nuclear pore; (sv) small vesicle; (pm) plasma membrane; (nm) nuclear membrane.

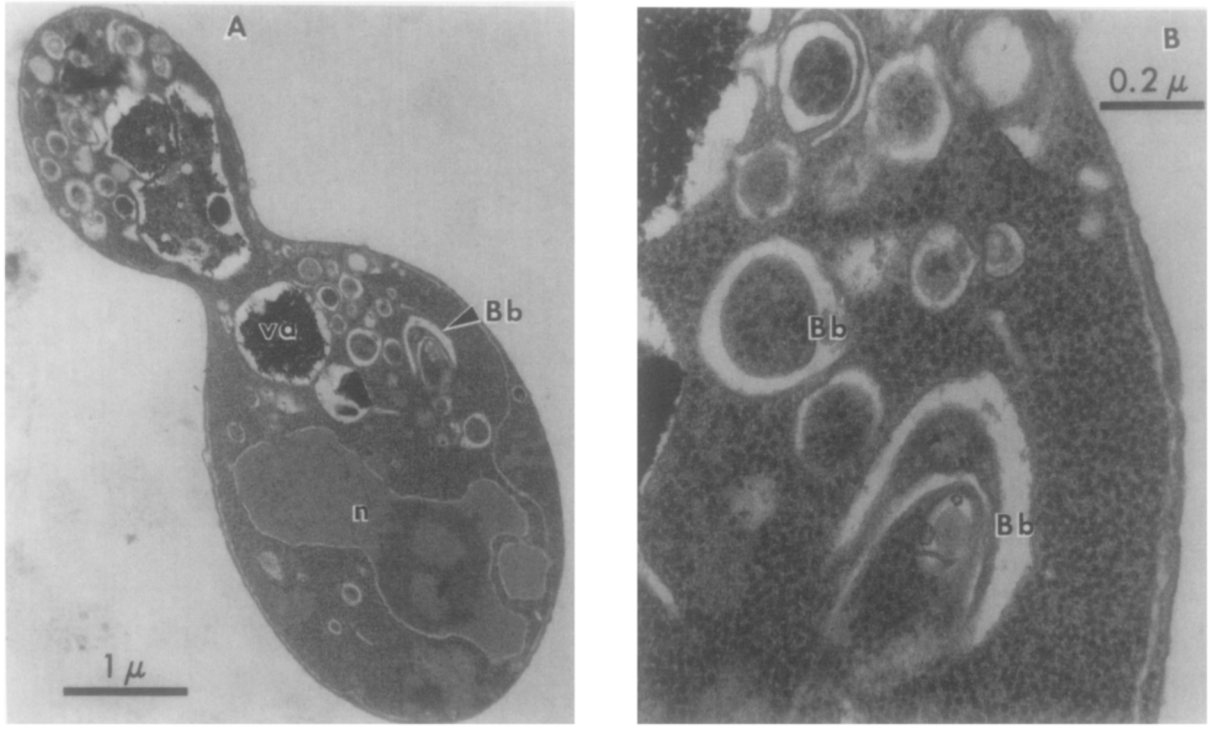


Figure 4. Thin Section Electron Micrograph of HMSF 6 (*sec*7-1) Grown in YPD Medium at 25°C, Then Shifted to 37°C for 2 Hr
(A) Low magnification; (B) a portion of the same cell at higher magnification. Symbols are as in Figure 2 and (Bb) Berkeley body.

with (1979) have shown that fusion of the E. coli *lac*Z (β-galactosidase) and *mal*E (periplasmic maltose binding protein) genes leads to the production of a hybrid protein which becomes stuck in the cytoplasmic membrane. The aberrant incorporation of this hybrid protein prevents the secretion of other proteins and the cells die. Such mutations are genetically dominant (T. Silhavy, personal communication); cell death resulting from the production of hybrid protein is not prevented by synthesis of a normal *mal*E gene product. In contrast, all the *sec* mutants are recessive and therefore the defective gene products are not likely to be secretion inhibitors.

Although 16 of the 23 representative mutant alleles are at least 5 fold reduced at 37°C in each of the four parameters of cell surface growth (Tables 3 and 4), several of the mutants are less restrictive (leaky). Some of the leaky strains (*sec*15, 17, 21) are in complementation groups for which only 1–2 mutant alleles are available. Among the groups with many members, some alleles are very restrictive at 37°C, while others are leaky. Certain strains are less restrictive for one enzyme marker than another. This may be due to the different growth conditions required for derepression of the marker enzymes. Thus while invertase appears in cells within 30 min of a transfer from 5 to 0.1% glucose, acid phosphatase production requires a 1.5 hr phosphate starvation. The *sec*11 group is an extreme example. Cells of all 12 mutant alleles secrete nearly normal amounts of invertase at 37°C, but secrete no acid phosphatase. Although no organelles accumulate in *sec*11 at 37°C, internal invertase levels rise 7 fold, and about one third of the accumulated invertase is secreted when cells are returned to 25°C. The *sec*11 gene product may be difficult to convert to a completely thermosensitive form, or it may not be required absolutely for the secretory process.

More important than the occasional leaky strain is the fact that in most *sec* mutants the block to secretion and membrane permease incorporation is coupled with the accumulation of secretory enzymes and membrane-enclosed organelles. van Rijn, Linnemans and Boer (1975) have shown by histochemical staining of wild-type cells that acid phosphatase is contained within ER, Golgi-like structures and vesicles. The acid phosphatase that accumulates in *sec*1 cells at 37°C is contained within vesicles (Novick and Schekman, 1979); the ER and Berkeley bodies (Bbs) produced in other mutants also contain this enzyme (B. Esmon, P. Novick and R. Schekman, manuscript in preparation). The various membrane-enclosed structures produced at 37°C probably represent functional intermediates in the secretory pathway, since most of the mutants secrete the accumulated invertase upon return to 25°C.

The rates of invertase synthesis and export may also be coupled. In many of the mutants, 2–4 fold more invertase accumulates in an internal pool at 37°C than is secreted by wild-type cells in a comparable period (Table 3).

Some of the mutants accumulate more than one

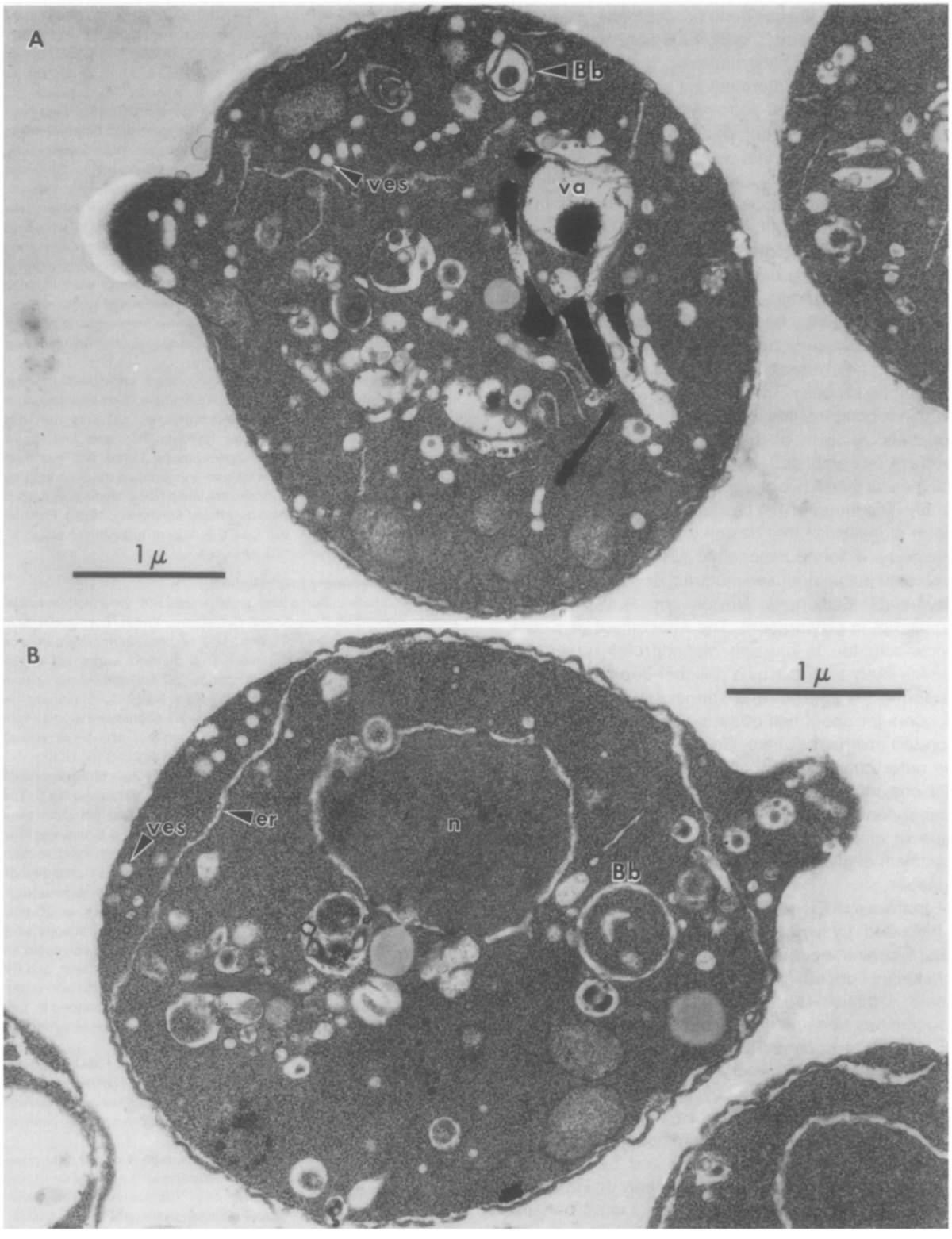


Figure 5. Thin Section Electron Micrographs of Cells Grown in YPD Medium at 25°C, Then Shifted to 37°C for 2 Hr
(A) HMSF 169 (*sec*14-3); (B) HMSF 178 (*sec*19-1). Symbols are as in Figures 2–4.

type of organelle. If a *sec* gene product acts at several stages in the pathway, a partial block might result in multiple structures. Alternatively, if the reactions which connect one intermediate to another are reversible, a block in the pathway could lead by mass action to the accumulation of an earlier intermediate. A third possibility is that unstable organelles may reversibly or irreversibly generate other structures.

The close resemblance of the structure in some *sec* mutants to ER suggests that the affected gene products are required at an early step in the pathway. The vesicle-accumulating mutants are probably defective in later steps. Bbs may be analogous to the Golgi apparatus; they may become distorted by the continued incorporation of new membrane and secretory material in the absence of discharge to a succeeding stage in the process.

An independent line of evidence supports these proposals. Analysis of the extent of glycosylation of invertase accumulated at 37°C in the *sec* strains indicates at least two stages in oligosaccharide assembly. The mutants that produce ER also accumulate a form of invertase that has only half as much carbohydrate as is found associated with the enzyme accumulated in the other *sec* mutants, or on the secreted enzyme (B. Esmon, P. Novick and R. Schekman, manuscript in preparation). The attachment of a core oligosaccharide to nascent mannoprotein chains in yeast is likely to occur by a dolichol-dependent reaction in the ER (Marriot and Tanner, 1979). This core accounts for about half of the carbohydrate found on secreted invertase (Lehle, Cohen and Ballou, 1979). The outer chain oligosaccharides are added by distinct enzymes (Raschke et al., 1973) in a dolichol-independent reaction (Parodi, 1979). This second stage of oligosaccharide assembly may occur after movement of glycoproteins from the ER to a Golgi-like organelle.

A pathway of ER → Bbs → vesicles → cell surface is indicated by cytological analysis of double *sec* mutant strains incubated at 37°C (P. Novick and R. Schekman, unpublished results). Further analysis should establish the order in which the *sec* gene products function and the relationship of this pathway to secretory and membrane protein maturation. The contribution of the secretory pathway to cell surface growth will be tested directly by the isolation and analysis of organelle and plasma membrane fractions from the *sec* mutant cells. By providing an enriched supply of intermediate organelles and by providing criteria for the authentic reconstruction of individual events in vitro, the *sec* mutants may aid in a biochemical dissection of the secretory pathway.

Experimental Procedures

Materials

S. cerevisiae isogenic haploid strains X2180-1A (*a*, gal2) and −1B (α, gal2) were from the yeast genetics stock center (Berkeley). NF1R, a spontaneous GAL$^+$ revertant, was derived from X2180-1A. HMSF 1 (*a*, *sec*1-1) was derived from X2180-1A (Novick and Schekman, 1979). Standard genetic techniques were used to construct SF182-3B (*a*, GAL$^+$) and SF150-5C (*a*, *sec*1-1, III *ACP*1-2, *pho*80-2). All other *sec* strains were derived from X2180-1A, NF1R or SF182-3B as described in the text.

YPD medium contained 1% Bacto-Yeast Extract, 2% Bacto-Peptone and 2% glucose; YP medium was the same with different levels of glucose. Phosphate-depleted YPD was prepared as described by Rubin (1973). Wickerham's minimal medium (Wickerham, 1946) was used with the following modifications: for phosphate-free medium, potassium chloride replaced potassium phosphate; for sulfate-free medium, chloride salts replaced all sulfate salts. Unless otherwise indicated, the carbon source was 2% glucose. Petri plates contained the indicated medium and 2% Difco agar. Liquid cultures were grown in flasks or tubes with agitation, and the experiments were initiated with exponentially growing cells from stock cultures at an A_{600} of 0.5–5. The absorbance of cell suspensions was measured in a 1 cm quartz cuvette at 600 nm in a Zeiss PM QII spectrophotometer; 1 A_{600} unit corresponds to 0.15 mg dry weight.

Other reagents were obtained as indicated: ethyl methanesulfonate, p-nitrophenolphosphate, glucose oxidase, O-dianisidine, peroxidase and cycloheximide were from Sigma; $H_2{}^{35}SO_4$ was from New England Nuclear; glutaraldehyde, osmium tetroxide and Spurr embedding medium were from Polysciences; Ludox AM was from Protex Wax (Oakland, California) and was purified as described by Price and Dowling (1977). Lyticase is a yeast lytic enzyme preparation useful in spheroplast formation (Scott and Schekman, 1980). Fraction II (30,000 U/mg; 1 unit will lyse 0.2 A_{600} of logarithmic phase S. cerevisiae in 30 min at 30°C) was used.

Isolation of Secretory (*sec*) Mutants

Stationary phase cultures were mutagenized with ethyl methanesulfonate as described (Novick and Schekman, 1979). For nitrous acid mutagenesis, stationary phase X2180-1A cells were collected on a nitrocellulose filter, washed twice with distilled water and resuspended in 5 ml of 0.5 M sodium acetate (pH 4.8) and 20 mg sodium nitrate. After 10 min at 30°C, 5 ml of 2.7% Na_2HPO_4 containing 1% yeast extract were added, and the cells were filtered and washed with water. The viability was 37%. The cells were then allowed to recover from mutagenesis by growth in YPD medium at 25°C for 16 hr.

In the largest enrichment experiment, 1.85 A_{600} of mutagenized cells were grown in 50 ml of YPD medium to a total A_{600} = 25.5. The culture was then incubated at 37°C for 3 hr and the cells were collected by filtration, washed and resuspended in 0.5 ml water. The cell sample was layered on 12.5 ml of a mixture containing Wickerham's salts (Wickerham, 1946) and 60% (v/v) of the purified stock Ludox suspension, in a Falcon 17 × 100 mm polypropylene tube. After centrifugation in a Sorval SS34 rotor (22,000 × g, 20 min, 4°C), the tube was punctured at the bottom and 1 ml fractions were collected. The A_{600} of the fractions was measured and corrected for the A_{600} of corresponding fractions from a cell-free gradient, and the densest 2% of the cells (1.5 A_{600} total) was diluted 2 fold with water. The cells were then centrifuged, the pellet was resuspended in 1 ml of water and diluted 400 fold, and 0.1 ml portions were spread on 200 YPD medium agar plates.

The plates were incubated at room temperature (20–25°C) for 2 days and colonies were replica-plated onto two YPD plates each; one replica was incubated at room temperature, the other at 37°C. After 26 hr the replicas were compared and temperature-sensitive colonies were picked from the master plate.

The temperature-sensitive clones were replica-plated onto phosphate-free minimal medium plates to derepress the synthesis of acid phosphatase, and after 10 hr at 24° or 37°C the replicas were stained for secreted acid phosphatase (Hansche, Beres and Lange, 1978).

The clones which showed temperature-sensitive secretion of phosphatase were screened for conditional secretion and internal accumulation of invertase. Cultures grown at 25°C in YP + 5% glucose medium were shifted to 37°C for 30 min, after which 2 A_{600} units of cells were sedimented for 1.5 min in a clinical centrifuge. The cell pellets were then resuspended in 2 ml of fresh YP + 0.1% glucose medium and cultures were incubated at 37°C for an additional 90 min. The cells were sedimented again and the pellets were resus-

pended in 10 mM NaN_3 at 0°C. Cell wall and internal invertase levels were measured as previously described (Novick and Schekman, 1979). Strains which secreted less invertase than the X2180 control were designated *sec* mutants; those that accumulated internal levels higher than the X2180 control were assigned to class A, and all others were put aside for further analysis.

Analytical Procedures

Cell number was determined with a hemocytometer; buds were counted as cells. Density was determined by sedimentation of 8 A_{600} units of cells on a Ludox gradient as described above. The region of highest cell concentration was estimated visually and a 0.5 ml sample was removed by puncturing the centrifuge tube with a syringe. Sample density was measured by weighing a 100 μl portion.

External (cell wall-bound) acid phosphatase was assayed at 37°C as described by van Rijn, Boer and Steyn-Parvé (1972); units of activity are nmole of p-nitrophenol released per min. External invertase was assayed at 37°C as described by Goldstein and Lampen (1975); units of activity are μmole of glucose released per min. Internal invertase was determined by assaying spheroplast lysates, prepared as previously described (Novick and Schekman, 1979). Sulfate permease was assayed at 37°C in 50 μM $(NH_4)_2SO_4$ as described by Breton and Surdin-Kerjan (1977); units of activity are nmole of SO_4^{2-} uptake per min. Radioactivity was measured in a Searle Delta 300 liquid scintillation counter.

Samples were prepared for electron microscopy by the procedure of Byers and Goetsch (1975).

Acknowledgments

We thank Susan Ferro and Frank Gadzhorn for help in the mutant screening process. We also thank Alice Taylor for her continued expert assistance with electron microscope techniques. This work was supported by grants from the NSF and the NIH.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 22, 1980; revised May 27, 1980

References

Bassford, P. and Beckwith, J. (1979). *Escherichia coli* mutants accumulating the precursor of a secreted protein in the cytoplasm. Nature *277*, 538–541.

Bassford, P. J., Silhavy, T. J. and Beckwith, J. R. (1979). Use of gene fusion to study secretion of maltose-binding protein into *Escherichia coli* periplasm. J. Bacteriol. *139*, 19–31.

Breton, A. and Surdin-Kerjan, Y. (1977). Sulfate uptake in *Saccharomyces cerevisiae*: biochemical and genetic study. J. Bacteriol. *132*, 224–232.

Byers, B. and Goetsch, L. (1975). Behavior of spindles and spindle plaques in the cell cycle and conjugation of *Saccharomyces cerevisiae*. J. Bacteriol. *124*, 511–523.

Chan, S. J., Keim, P. and Steiner, D. F. (1976). Cell-free synthesis of rat preproinsulins: characterization and partial amino acid sequence determination. Proc. Nat. Acad. Sci. USA *73*, 1964–1968.

Dodyk, F. and Rothstein, A. (1964). Factors influencing the appearance of invertase in *Saccharomyces cerevisiae*. Arch. Biochem. Biophys. *104*, 478–486.

Field, C. and Schekman, R. (1980). Localized secretion of acid phosphatase reflects the pattern of cell surface growth in *Saccharomyces cerevisiae*. J. Cell Biol. *86*, in press.

Fincham, J. R. (1977). Allelic complementation reconsidered. Carlsberg Res. Commun. *42*, 421–430.

Goldstein, A. and Lampen, J. O. (1975). β-D-Fructofuranoside fructohydrolase from yeast. Meth. Enzymol. *42*, 504–511.

Hansche, P. E., Beres, V. and Lange, P. (1978). Gene duplication in *Saccharomyces cerevisiae*. Genetics *88*, 673–687.

Henry, S. A., Atkinson, K. D., Kolat, A. I. and Culbertson, M. R. (1977). Growth and metabolism of inositol-starved *Saccharomyces cerevisiae*. J. Bacteriol. *130*, 472–484.

Katz, F. N., Rothman, J. E. Knipe, D. M. and Lodish, H. F. (1977). Membrane assembly: synthesis and intracellular processing of the vesicular stomatitis viral glycoprotein. J. Supramol. Structure *7*, 353–370.

Lehle, L., Cohen, R. E., and Ballou, C. E. (1979). Carbohydrate structure of yeast invertase. J. Biol. Chem. *254*, 12209–12218.

Marriot, M. and Tanner, W. (1979). Localization of dolichyl phosphate- and pyrophosphate-dependent glycosyl transfer reactions in *Saccharomyces cerevisiae*. J. Bacteriol. *139*, 565–572.

Matile, P., Cortat, M., Wiemken, A. and Frey-Wyssling, A. (1971). Isolation of glucanase-containing particles from budding *Saccharomyces cerevisiae*. Proc. Nat. Acad. Sci. USA *68*, 636–640.

Moor, H. (1967). Endoplasmic reticulum as the initiator of bud formation in yeast (*Saccharomyces cerevisiae*). Arch. Mikrobiol. *57*, 135–146.

Novick, P. and Schekman, R. (1979). Secretion and cell surface growth are blocked in a temperature-sensitive mutant of *Saccharomyces cerevisiae*. Proc. Nat. Acad. Sci. USA *76*, 1858–1862.

Palade, G. (1975). Intracellular aspects of the process of protein synthesis. *Science 189*, 347–358.

Parodi, A. J. (1979). Biosynthesis of yeast mannoproteins: synthesis of mannan outer chain and of dolichol derivatives. J. Biol. Chem. *254*, 8343–8352.

Price, C. A. and Dowling, E. L. (1977). On the purification of the silica sol Ludox AM. Anal. Biochem. *82*, 243–245.

Raschke, W. C., Kern, K. A., Antolis, C. and Ballou, C. E. (1973). Genetic control of yeast mannan structure. J. Biol. Chem. *248*, 4660–4666.

Rubin, G. M. (1973). The nucleotide sequence of *Saccharomyces cerevisiae* 5.8S ribosomal ribonucleic acid. J. Biol. Chem. *248*, 3860–3875.

Scott, J. and Schekman, R. (1980). Lyticase: endoglucanase and protease activities that act together in yeast cell lysis. J. Bacteriol. *142*, 414–423.

Tkacz, J. S. and Lampen, J. O. (1972). Wall replication in *Saccharomyces* species: use of fluorescein-conjugated concanavalin A to reveal the site of mannan insertion. J. Gen. Microbiol. *72*, 243–247.

Tkacz, J. S. and Lampen, J. O. (1973). Surface distribution of invertase on growing *Saccharomyces* cells. J. Bacteriol. *113*, 1073–1075.

van Rijn, H. J. M., Boer, P. and Steyn-Parvé, E. P. (1972). Biosynthesis of acid phosphatase of baker's yeast. Factors influencing its production by protoplasts and characterization of the secreted enzyme. Biochim. Biophys. Acta *268*, 431–441.

van Rijn, H. J. M., Linnemans, W. A. M. and Boer, P. (1975). Localization of acid phosphatase in protoplasts from *Saccharomyces cerevisiae*. J. Bacteriol. *123*, 1144–1149.

Wickerham, L. J. (1946). A critical evaluation of the nitrogen assimilation tests commonly used in the classification of yeasts. J. Bacteriol. *52*, 293–301.

Wood, W. and King, J. (1979). Genetic control of complex bacteriophage assembly. In Comprehensive Virology, *13*, H. Frankel-Conrat and R. Wagner, eds. (New York and London: Plenum Press), pp. 581–624.

A Yeast Mutant Defective at an Early Stage in Import of Secretory Protein Precursors into the Endoplasmic Reticulum

Raymond J. Deshaies and Randy Schekman

Department of Biochemistry, University of California, Berkeley, California 94720

Abstract. We have devised a genetic selection for mutant yeast cells that fail to translocate secretory protein precursors into the lumen of the endoplasmic reticulum (ER). Mutant cells are selected by a procedure that requires a signal peptide–containing cytoplasmic enzyme chimera to remain in contact with the cytosol. This approach has uncovered a new secretory mutant, *sec*61, that is thermosensitive for growth and that accumulates multiple secretory and vacuolar precursor proteins that have not acquired any detectable post-translational modifications associated with translocation into the ER. Preproteins that accumulate at the *sec*61 block sediment with the particulate fraction, but are exposed to the cytosol as judged by sensitivity to proteinase K. Thus, the *sec*61 mutation defines a gene that is required for an early cytoplasmic or ER membrane–associated step in protein translocation.

THE first step in the biogenesis of proteins destined for the secretory pathway is their insertion into the membrane of the endoplasmic reticulum (ER).[1] This process has been studied intensively in mammalian cells through the use of an in vitro assay that faithfully reproduces cotranslational translocation of secretory proteins into the lumen of the ER (2). Dissection of the components required for this activity has revealed the existence of both soluble and membrane-bound factors that participate in protein translocation. The signal recognition particle is a soluble ribonucleoprotein particle consisting of six polypeptides (54) and one molecule of 7SL RNA (55). The signal recognition particle binds to the signal sequence of a nascent preprotein (28, 56), thereby forming a complex that interacts with an integral membrane protein of the ER known as docking protein or signal recognition particle receptor (13, 33). This targeting event is followed by cotranslational translocation of the preprotein into the ER lumen. Either during or shortly after the translocation event, the signal sequence is cleaved by the enzyme signal peptidase (10) and core oligosaccharides are transferred to specific asparagine residues (44) of the translocated polypeptide. The mechanism of protein permeation across the hydrophobic core of the ER membrane is not understood. Experiments with intermediates artificially blocked at various stages of membrane penetration suggest that this process is mediated by proteins, though they have yet to be identified by the existing assays (14).

Recently, several groups have reconstituted protein translocation into the yeast ER in vitro (15, 42, 57). A yeast translation extract programmed with prepro-α-factor mRNA directs the synthesis of an intact precursor which can insert co- or posttranslationally into yeast microsomes and become core-glycosylated. The existence of a posttranslational reaction has allowed these investigators to demonstrate that protein translocation into the yeast ER is energy dependent (15, 43, 57). In addition, fractionation experiments suggest that the import reaction requires cytosolic components (58). The reconstitution of yeast protein translocation in vitro presents an opportunity to combine a biochemical analysis of protein translocation with a genetic approach aimed at identifying genes whose products participate in the reconstituted reaction.

Among a large collection of temperature-sensitive, secretion-defective mutants of *Saccharomyces cerevisiae* isolated in this laboratory (12, 36, 37), members of only two complementation groups (*sec*53 and *sec*59) affect early events in protein secretion, though neither mutation blocks protein translocation into the ER (10a). In an attempt to identify genes required for the translocation event, we have developed a direct selection for temperature-sensitive (Ts), import mutants of *Saccharomyces cerevisiae*.

The protocol described in this report is similar to that used by Oliver and Beckwith (39) to isolate mutants of *Escherichia coli* defective in the export of periplasmic and cell wall proteins from the cytoplasm. Strains expressing abortively translocated fusions of the periplasmic maltose-binding protein (MBP) to β-galactosidase produced much less β-galactosidase activity than strains harboring similar fusions with mutations in the maltose-binding protein signal sequence. By selecting for mutants that expressed high levels of β-galactosidase activity from the wild-type fusion protein (presumably owing to retention of the fusion protein in the cytoplasm), Oliver and Beckwith (39) were able to isolate mutations in

1. *Abbreviations used in this paper*: CPY, carboxypeptidase Y; ER, endoplasmic reticulum; proCPY, procarboxypeptidase Y; Ts, temperature sensitive; YPD, 1% yeast extract, 2% peptone, 2% dextrose.

two genes (*sec*A and *sec*B) that blocked export and maturation of multiple secretory precursors.

We have modified this strategy to select positively for mutants of *Saccharomyces cerevisiae* that are defective in protein translocation. In this report, we describe the isolation and phenotypic characterization of mutants that define two complementation groups (*sec*61 and *sec*62). Mutations in both genes cause temperature-sensitive growth and accumulation of α-factor precursor. Detailed analysis of *sec*61 strains has revealed that this mutation results in the accumulation of cytoplasmically exposed and unmodified precursors of multiple secretory and vacuolar proteins.

Materials and Methods

Strains, Growth Conditions, and Materials

The bacterial and yeast strains used in this study are listed in Table I. All plasmids were propagated in HB101, except those sensitive to restriction by BclI, which were propagated in NK 5772. Yeast strains were constructed by standard genetic techniques (48). Original mutant isolates were backcrossed at least three consecutive times to RDB 103 to test for cosegregation of the Ts growth and α-factor accumulation phenotypes. All experiments were performed with these backcrossed derivatives. Haploid *sec*61 strains were recovered from the original *sec*61 diploid isolates by mating the diploid mutant to diploid S395D-1, sporulating this tetraploid strain, and then sporulating a Ts *MAT*a/*Mat*α diploid spore clone derived from the tetraploid. Haploid *MAT*α mutant strains were then backcrossed with RDB 103.

YPD liquid broth contained 1% Bacto-Yeast extract, 2% Bacto-Peptone (Difco Laboratories, Detroit, MI), and 2–5% glucose. Wickerham's minimal medium (59) was used with 2–5% glucose. Solid media were supplemented with 2% Bacto-Agar. For pulse labeling of cells with $[^{35}S]SO_4^{2-}$, sulfate salts were replaced by chloride salts, and ammonium sulfate was supplemented at 200 μM for overnight growth and at 0–10 μM during radiolabeling. Liquid cultures were grown in flasks with vigorous agitation, and experiments were initiated with cells in logarithmic phase. The optical density at 600 nm (OD_{600}) of dilute cell suspensions was measured in 1-cm quartz cuvettes using a Zeiss PMQII spectrophotometer (Carl Zeiss, Inc., Thornwood, NY); 1 OD_{600} of cells corresponds to 0.15 mg of dry weight.

The following reagents were obtained as indicated: histidinol, tunicamycin, ethyl methanesulfonate, concanavalin A(Con A)-Sepharose 4B, α-methylmannoside, NADPH, cytochrome *c*, proteinase K, protein A, bovine serum albumin, and phenylmethylsulfonyl fluoride (PMSF) were obtained from Sigma Chemical Co., St. Louis, MO; restriction endonucleases, DNA modification enzymes, and nuclease-treated wheat germ in vitro translation components were from Bethesda Research Laboratories, Gaithersburg, MD; T4 DNA ligase and glyceraldehyde-3-phosphate (diethylacetyl form) were from Boehringer Mannheim Biochemicals, Indianapolis, IN; Amplify, [^{35}S]methionine (1,200 Ci/mmol), and ^{125}I-NaI (highest specific activity) were from Amersham Corp., Arlington Heights, IL; SDS-polyacrylamide gel electrophoresis reagents (electrophoresis grade) were from Bio-Rad Laboratories, Richmond, CA; nitrocellulose was from Schleicher & Schuell, Inc., Keene, NH; protein A-Sepharose was from Pharmacia Fine Chemicals, Piscataway, NJ; carrier-free $[^{35}S]Na_2SO_4$ was from ICN Radiochemicals, Irvine, CA; IgG Sorb was from The Enzyme Center, Boston, MA; and nonfat dry milk was from Safeway, North Berkeley, CA. Lyticase (fraction II, 30,000–90,000 U/ml) was prepared as described by Scott and Schekman (46). Invertase (45) and carboxypeptidase Y (50) antisera were prepared as described previously. Anti-α-factor sera was generously provided by J. Rothblatt, European Molecular Biology Laboratory, Heidelberg, Federal Republic of Germany (42). Rabbit reticulocyte lysate was prepared as described (1).

Construction of *HIS4* Gene Fusions

A *SUC2-HIS4* gene fusion was constructed in the yeast shuttle vector YCp50, which is a low copy number plasmid that contains *URA3* as a selectable marker and *CEN4 ARS1* for mitotic stabilization and replication competence, respectively (51). In the first step, the 1.6-kb Eco RI-Bam HI fragment from the 5′ portion of the *SUC2* gene (from pRB58 [5]) was inserted between the unique Eco RI and Bam HI sites in YCp50 to generate $YCp50^{EB}$. Next, the 1.6-kb Xho II fragment of pYAH-I2 (40), which contains the 5′ portion of *HIS4*, was inserted into the Bam HI site of $YCp50^{EB}$ to generate YCp501. The remainder of the *HIS4* gene was introduced by replacing the 1,350-bp Cla I-Sph I fragment of YCp501 with the corresponding 2.7-kb fragment from pYAH-I2 resulting in plasmid YCp502, which contains almost the entire *HIS4* gene (except for the first 33 codons of *HIS4*A) fused out of frame to the 5′ half of the *SUC2* gene. An in-frame *SUC2-HIS4* fusion was created by inserting the 250-bp Bam HI-Bcl I fragment of *SUC2* into the unique Bam HI site of YCp502, yielding YCp503. Unfortunately, this fusion plasmid failed to complement *his4C* mutations, even though a hybrid protein of the proper size was synthesized constitutively.

We reasoned that an increase in the copy number of the *SUC2-HIS4* fusion might allow complementation of *his4C* strains. Thus, the fusion gene was introduced into a multicopy 2 μm-based vector. The junction sequences and *HIS4* coding portion of the *SUC2-HIS4* fusion were transferred to

Table I. Bacterial and Yeast Strains

Strain	Genotype	Source or reference
Saccharomyces cerevisiae		
RDB 103	*leu2*-3,-112 *ade2 MAT*a	This study
RDM 15-5B	*leu2*-3,-112 *ade2 ura3*-52 *pep4*-3 *sec61*-2 *MAT*α	This study
RDM 15-9B	*ade2 pep4*-3 *Mat*α	This study
RDM 15-10D	*leu2*-3,-112 *ade2 sec18*-1 *MAT*α	This study
RDM 15-3A	*leu2*-3,-112 *his4 pep4*-3 *sec18*-1 *sec61*-2 *MAT*α	This study
FC2-12B	*leu2*-3,-112 *ura3*-52 *trp1*-1 *his4*-401 *HOL*1-1 *MAT*α	R. Parker (40)
DYFC2-12B	*leu2*-3,-112/*leu2*-3,-112 *ura3*-52/*ura3*-52 *trp1*-1/*trp1*-1 *his4*-401/*his4*-401 *HOL*1-1/*HOL*1-1 *MAT*α/*MAT*α	This study*
S395D-1	*his1*/*his1 leu1*/*leu1 trp2*/*trp2 MAT*a/*MAT*a	YGSC‡
PBY404C	*suc2*-Δ9 *MAT*a	P. Böhni
165/7	*his4*-25 (*his4*A$^-$) *MAT*a	R. Parker (40)
E331	*his4*-331 (*his4*B$^-$) *MAT*a	R. Parker (40)
S942-1Ca	*his4*-864 (*his4*C$^-$) *MAT*a	R. Parker (40)
Escherichia coli		
NK5772	*dcm*-6 *dam*-3 *gal*K2 *gal*T22 *met*B1 *leu*Y1 *tsx*-78 *thi*-1 *ton*A31 *mtl*-1	J. Kadonaga
HB101	F$^-$ *hsdS20* (r_B^-, m_B^-) *rec*A13 *ara*-14 *pro*A2 *lac*Y1 *gal*K2 *rps*L20 (S_m^r) *xyl*-5 *mtl*-1 *sup*E44	(30)

* Spontaneously derived from FC2-12B.
‡ Yeast Genetic Stock Center, University of California, Berkeley, CA.

pRB58 by digesting YCp503 to completion with Bam HI, and partially with Sal I, to generate a 3.4-kb Bam HI-Sal I fragment. This fragment was inserted in place of the 4.2-kb Bam HI-Sal I fragment (containing the 3' half of *SUC2*) of pRB58. The resulting plasmid, pSHE1, produced more fusion protein and complemented mutations in *his*4A, 4B, and 4C.

Expression of the cytoplasmic invertase-*HIS4* fusion protein was eliminated by replacing the extreme 5' portion of the *SUC2* coding region with the promoter elements and coding sequences for the prepro region of the yeast *MFα1* gene. This replacement was accomplished by inserting the 1.9-kb Bgl II-Bam HI fragment of pSEY210 (8) into the site vacated by complete digestion of pSHE1 with Bam HI, and partial digestion with Bcl I (the relevant Bcl I site is at the extreme 5' end of the *SUC2* insert; the excised fragment is 2.7 kb). The product of this manipulation was designated pαSHF8. In contrast to the *SUC2-HIS4* fusion, which contained amino acid residues 1–289 of secretory invertase, pαSHF8 encoded a fusion protein in which the first 88 amino acids of the prepro region of α-factor were fused in frame to invertase starting at amino acid 5 of the signal peptide sequence. This resulted in the deletion of four amino acids from the NH_2 terminus of the invertase portion of the *MFα1-SUC2-HIS4* fusion (see Emr et al., [8]).

Plasmids were introduced into yeast strains either by the spheroplast (17) or lithium acetate (19) procedure. Agarose gel electrophoresis, plasmid purification, fragment isolation, transformation of bacteria, and other recombinant DNA manipulations were performed by standard methods (30). All recombinant DNA modifying enzymes were used according to the suppliers' instructions.

Mutant Isolation and Screening

FC2-12B and DYFC2-12B cells containing pαSHF8 were grown to stationary phase in minimal medium supplemented with histidine, tryptophan, and leucine. 12 OD_{600} U of cells were harvested by centrifugation, washed with sterile 50 mM potassium phosphate buffer, pH 7.0, and resuspended at 2 OD_{600}/ml in potassium phosphate buffer. Ethyl methanesulfonate was added to a final concentration of 3%, and the cells were incubated with the mutagen for 60 min (round II) or 75 min (round I) at 30°C (% killing = 50% in round I, 73% in round II). The mutagen was quenched by adding an equal volume of sterile 12% sodium thiosulfate, and cells were collected by centrifugation, washed two times consecutively with potassium phosphate buffer and resuspended in 30 ml of minimal medium supplemented with leucine, tryptophan, and histidine. After a 20-h recovery period at 24°C, the mutagenized cells were centrifuged, resuspended to 5 OD_{600}/ml in potassium phosphate buffer and plated onto minimal medium supplemented with leucine, tryptophan, and 3 mM histidinol (1–3 × 10^6 cells per plate). After incubation at 30°C for 5–10 d, mutant clones were picked and steaked onto YPD plates. After 2 d at 30°C these patches were replica-plated onto YPD plates and individual replicas were incubated at 30 and 37°C. Clones that grew at 30°C but not at 37°C were picked and retested for Ts growth by streaking onto YPD plates at 37°C. Confirmed histidinol$^+$, Ts mutants were cured of the fusion plasmid by streaking them three times consecutively on nonselective medium (YPD). Uracil auxotrophs were isolated, rescreened for Ts growth, and retransformed with fresh pαSHF8 to assess the linkage of the Ts and histidinol$^+$ phenotypes to the original plasmid.

Ura$^-$ derivatives of the original isolates were screened by immunoblotting for the accumulation of intracellular forms of prepro-α-factor. Briefly, mutant colonies were inoculated into 3 ml of YPD and grown at 30°C to an OD_{600} of 1.5–15. 4 OD_{600} U of cells were collected and diluted to a volume of 3 ml with fresh YPD. These cultures were incubated for 2–3 h at 37°C and then diluted with an equal volume of ice-cold 20 mM NaN_3. The cells were collected by centrifugation, washed with ice cold 10 mM NaN_3, and resuspended in 0.3 ml Laemmli sample buffer (29) supplemented with 1 mM PMSF. Glass beads (~0.3 g of 0.3–0.5-mm diam) were added and the mixture was vortexed vigorously for 2 min, then immediately heated in boiling water for 5 min. Aliquots (24 μl) of these samples were applied to 12.5% SDS polyacrylamide gels and electrophoresis was performed as described by Laemmli (29).

Electrophoretically separated proteins were transferred to nitrocellulose filters (4), and the filters were processed essentially as described by Kaiser and Botstein (23), except that the blocking incubation was for 30 min at 24°C, and anti-α-factor antiserum (1/500 dilution) was used to probe the filters. An unfractionated lysate, prepared from *MATa* cells (PBY404C), was included at a final concentration of 3–5 mg of protein/ml during the antibody incubation to block antibody association with filter-bound proteins not related to α-factor. After decoration of bound antibodies with ^{125}I-labeled protein A, filters were exposed to X-ray film with an intensifying screen at −70°C.

Radiolabeling and Immunoprecipitation

Cells were grown overnight in minimal medium supplemented with 200 μM $(NH_4)_2SO_4$ to an OD_{600} of 0.5–1.0. An aliquot of cells was sedimented for 2–3 min at room temperature in a clinical centrifuge, washed with distilled water, and resuspended to 1–2 OD_{600}/ml in minimal medium supplemented with 0–10 μM $(NH_4)_2SO_4$. For radiolabeling of invertase, cells were first derepressed for invertase production by incubation in minimal medium (200 μM ammonium sulfate) plus 0.1% glucose for 30 min at 30 or 37°C, then collected and resuspended in minimal medium containing 0–10 mM $(NH_4)_2SO_4$ and 0.1% glucose. Radioactive sulfate ($H_2{}^{35}SO_4$) was added at a concentration of 200–300 μCi/OD_{600}, and incorporation was carried out for 20–40 min at 30 or 37°C. Labeling was terminated by the addition of an equal volume of ice-cold 20 mM NaN_3, and samples were chilled on ice for 5 min. Radioactive cells were sedimented in a clinical centrifuge, washed with ice-cold 10 mM NaN_3, and resuspended at 5–20 OD_{600}/ml in spheroplasting buffer, which contains 1.4 M sorbitol, 60 mM β-mercaptoethanol, 25 mM MOPS, pH 7.4, 10 mM NaN_3, 5 mM $MgCl_2$. Lyticase was added (25 U/OD_{600}) and cells were converted to spheroplasts during a 45-min incubation at 30°C. Spheroplasts were sedimented for 20 s in a microcentrifuge, resuspended to 5–10 OD_{600}/ml in 1% SDS, heated to 100°C for 4 min, and centrifuged for 10 min in a microcentrifuge (Oscar Fisher Co., Inc., Newburgh, NY) to remove insoluble material. Aliquots of the supernatant fractions were diluted to 1.0 ml in 200 mM NaCl, 12.5 mM sodium phosphate, pH 7.4 (PBS), 1% Triton X-100, 0.25% SDS. α-Factor (1 μl/OD_{600} cell equivalent), invertase (1.5 μl/OD_{600} cell equivalent) or carboxypeptidase Y (CPY) (2 μl/OD_{600} cell equivalent) antiserum was added in saturating amounts, and immunoprecipitations were carried out for 3–16 h at 4°C. For α-factor and invertase immunoprecipitations, samples were supplemented with an unfractionated lysate of nonradioactive cells (2 mg protein/ml) from strain PBY404C (*suc2*Δ, *MATa*). After the antibody incubation, protein A-Sepharose was added in saturating amounts (1 μl beads/μl antiserum) and the incubation was continued for 1–2 h at room temperature. Immune complexes were sedimented in a microcentrifuge, and washed sequentially with PBS, 1% Triton X-100, 0.1% SDS (two times), and 50 mM NaCl, 10 mM Tris, pH 7.4 (two times). Bound antigens were then dissociated by heating in Laemmli sample buffer, and samples were applied to 7.5% (for invertase and CPY) or 12.5% (for prepro-α-factor) SDS polyacrylamide gels. After SDS-PAGE, the gels were fixed and treated with Amplify. Radioactive proteins were visualized by exposure to Kodak X-OMAT AR film (Eastman Kodak Co., Rochester, NY) at −70°C. Autoradiograms were quantified by scanning with a Kratos model SD3000 spectrodensitometer coupled to a Kratos SDS300 density computer (Kratos Analytical Instruments, Ramsey, NJ) and Hewlett-Packard 3380A integrator (Hewlett-Packard Co., Palo Alto, CA).

Glycoprotein Precipitation with Con A-Sepharose 4B

Cells were labeled with [^{35}S]SO_4^{2-} and extracts were diluted 10 fold with concanavalin A reaction buffer (0.5 M NaCl, 20 mM Tris, pH 7.4, 2% Triton X-100) and split into two equal portions. One half was immunoprecipitated directly with antibodies against α-factor or CPY, and the other was treated with Con A-Sepharose 4B beads (30 μl packed volume of beads for the amount of whole-cell extract derived from 1 OD_{600} of cells) for 3 h at room temperature. Con A-coated beads were then sedimented in a microcentrifuge. The supernatant fraction (unbound) was removed, adjusted to 0.2% SDS, and heated in boiling water for 3 min; the beads (bound fraction) were washed two times with con A reaction buffer and Con A-bound glycoproteins were eluted by heating the beads in 200 μl 1% SDS at 100°C. The eluate was clarified by centrifugation and diluted to 1.0 ml in Con A reaction buffer. Unbound and bound fractions were then immunoprecipitated with α-factor or CPY antiserum as described in the preceding section.

Subcellular Fractionation and Proteolysis of sec18 sec61 Extracts

Fractionation of RDM 15-3A extracts was performed as described by Bernstein et al. (1) except that prior to spheroplast formation the culture was incubated for 60 min at 30°C in YPD containing 5% glucose. Immunoblotting was performed as described above in *Mutant Isolation and Screening*. NADPH cytochrome *c* reductase (25), glyceraldehyde-3-phosphate dehydrogenase (1), and total protein (31) were assayed as previously described. The latency of the NADPH cytochrome *c* reductase activity was determined by performing assays in the absence and presence of 0.1% Triton X-100.

Determinations of the proteolytic sensitivity of accumulated prepro-α-

factor and proCPY in lysates of *sec*18 *sec*61 cells were performed starting with spheroplasts prepared as described in the previous section. All subsequent operations were performed at 4°C. Spheroplasts (25 OD_{600} U) were layered over a 10-ml cushion of SPB and 1.9 M sorbitol, and sedimented at 5,000 rpm for 5 min in a Sorvall HB-4 rotor (DuPont/Sorvall, Newtown, CT). The spheroplast pellet was resuspended in 0.73 ml lysis buffer (0.3 M mannitol, 0.1 M KCl, 50 mM Tris, pH 7.5, 1 mM EGTA), transferred to a 2.0 ml Potter-Elvejem homogenizer tube, and lysed by three consecutive cycles of homogenization (1 min of homogenization followed by 1 min on ice) using a motor-driven Potter-Elvejem teflon pestle. After centrifugation at 2,000 rpm for 4 min in a Sorvall HB-4 rotor, the resulting cell-free extract was split into two 0.3-ml aliquots. One sample was adjusted to 0.4% Triton X-100, and 40-μl aliquots (without protease control) from each sample were transferred to 0.56 ml of 20% trichloroacetic acid (TCA). Proteinase K was added to the remainder of both samples at a final concentration of 0.3 mg/ml, and at 0, 1.5, 3, 10, and 20 min, 42-μl aliquots from each sample were quenched in 0.56 ml of 20% TCA (processing of the zero time points required 15 s). TCA precipitates were collected by centrifugation and washed with −20°C acetone. Precipitated proteins were solubilized in 50 μl of Laemmli sample buffer, and samples were diluted and immunoprecipitated with anti-α-factor and anti-CPY sera.

In Vitro Transcription and Translation

mRNA coding for secreted invertase was prepared by in vitro transcription of pGEM2-*SUC*2-23 (provided by J. Rothblatt), which contains the entire *SUC*2 gene inserted into pGEM2 downstream of the bacteriophage SP6 promoter (32). The template was cut with Pvu II, followed by phenol extraction and ethanol precipitation. Linear pGEM2-*SUC*2-23 (1.2 μg) was transcribed in a 20-μl reaction using SP6 polymerase as described by Rothblatt and Meyer (42). A 1-μl aliquot of the transcription mix was translated (10 μl reaction) in the BRL nuclease-treated wheat germ in vitro translation system, and the reaction was terminated by adding SDS to 2% and heating at 95°C for 4 min. Translation products were immunoprecipitated with anti-invertase serum as described above in *Radiolabeling and Immunoprecipitation*. The in vitro translation products shown in Fig. 6, lanes *7* and *9* each represent one-sixth of a 10-μl reaction.

Hybrid-selected α-factor mRNA was translated in a rabbit reticulocyte lysate as described by Bernstein et al. (1).

Results

A Selection Scheme for Yeast Translocation Mutants

Numerous experiments have demonstrated that a signal sequence can direct cytoplasmic proteins to the secretory pathway in vivo (7, 47, 49). Based on these data, we reasoned that if a signal sequence were attached to a cytoplasmic enzyme required for the formation of an essential nutrient, the hybrid protein would be directed to the ER. If the substrate of the enzyme were limited to the cytoplasm, the cells would not grow unless they were supplied with the nutrient. Mutations that block hybrid protein import into the ER would allow cells to grow on the substrate.

We chose the *HIS*4 gene because it encodes a trifunctional cytoplasmic polypeptide that can sustain amino-terminal protein fusion events and retain histidinol dehydrogenase activity associated with the carboxy-terminal domain (24, 40). Histidinol dehydrogenase catalyzes the last step in histidine biosynthesis, the conversion of histidinol to histidine. Cells that express wild-type *HIS*4 protein and contain the mutant allele *HOL*1-1 (this mutation increases the efficiency of histidinol uptake) are able to convert exogenously supplied histidinol to histidine.

We predicted that if a signal sequence were fused to the amino terminus of the *HIS*4 protein, it would be translocated into the lumen of the ER and glycosylated, as shown schematically in Fig. 1 *A*. Wild-type cells (bearing the *his*4-401 deletion which overlaps the A, B, and C regions of *HIS*4) expressing this plasmid-encoded fusion protein would not convert histidinol to histidine in that the histidinol dehydrogenase would be sequestered from the cytoplasm, and charged, polar molecules such as histidinol or histidine are unlikely to penetrate across the ER membrane. Therefore, these cells would not grow on medium containing histidinol in place of histidine. If cells containing this gene fusion were selected for growth on histidinol, mutants that mislocalized the fusion protein to the cytoplasm would convert the histidinol to histidine (Fig. 1 *B*) and grow.

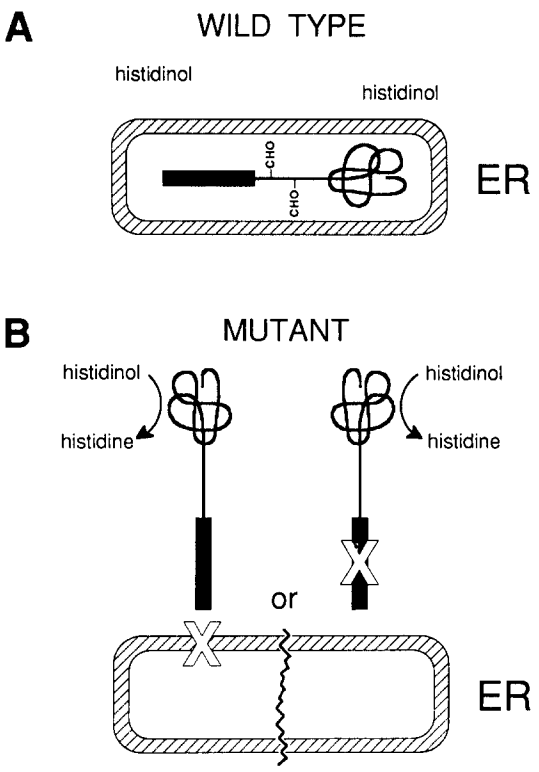


Figure 1. Predicted topology of a signal peptide–containing *HIS*4 hybrid protein. The predicted subcellular location and enzymatic properties of a histidinol dehydrogenase fusion protein are shown for both (*A*) wild-type and (*B*) mutant cells. (*CHO*) Asparagine-linked carbohydrate; (▬) a signal sequence; and (*folded black trace*) the catalytic domain of the fusion protein. (*X*) A mutation that inactivates either the signal sequence or some cellular component (membrane-bound or cytosolic) required for protein translocation.

Potentially this procedure would select for mutations that disrupted signal sequence function or mutations that disabled the cellular machinery responsible for targeting secretory proteins to the ER. These alternative possibilities are indicated by *X* marks in Fig. 1 *B*. Genetic tests (i.e., plasmid linkage) may be used to distinguish between these two classes of mutations. The strains described in this report contained recessive mutations that caused pleiotropic defects in the ER targeting apparatus.

In order to recover viable, translocation-defective cells, it was necessary to strike a balance between the lethal effect of secretory mutations and the requirement for sufficient cytosolic histidinol dehydrogenase activity. For this purpose, the growth temperature (30°C) represented a semipermissive condition where the mutant phenotype was only partially ex-

pressed. To identify mutants with more severe mislocalization defects, clones selected for growth on histidinol at 30°C were screened on rich medium for Ts growth at 37°C, presumably owing to an exaggeration of the partial defect expressed at 30°C.

Expression of HIS4 Hybrid Proteins

Two conditions were required for the approach to succeed: first, protein products of *HIS*4 gene fusions must retain histidinol dehydrogenase activity; and second, such a fusion protein must be directed to the yeast secretory apparatus in vivo, rendering the enzyme unable to supply histidine for growth. To test the first requirement, a gene fusion that encodes both a cytoplasmic and a signal peptide-containing hybrid was produced by ligating the 5′ half of the *SUC*2 coding region to a fragment containing the bulk of the *HIS*4 coding region (Fig. 2 *A*). The *SUC*2 gene codes for two different species of the yeast enzyme invertase (5): a cytoplasmic, unglycosylated form of the enzyme is expressed constitutively, and a secreted, highly glycosylated form is derepressed by growth in medium containing low concentrations of glucose. The primary structure of these two polypeptides differs only in the transitory presence of a signal peptide at the amino terminus of the secreted preenzyme. Yeast cells (*his*4Δ, *HOL*1-1) transformed with a multicopy plasmid containing the *SUC*2-*HIS*4 gene fusion (pSHE1) directed the synthesis of two different fusion proteins (data not shown). As expected, a nonglycosylated 130-kD species was produced constitutively. When transformants were shifted to derepression medium, a glycosylated fusion protein was also synthesized. To determine whether the cytoplasmic hybrid protein possessed histidinol dehydrogenase activity, pSHE1 transformants (*his*4Δ, *HOL*1-1) were assayed for their ability to grow on minimal medium containing histidinol and high concentrations of glucose. As a control, untransformed cells and cells transformed with vector sequences lacking the *HIS*4 insert were also analyzed. The results shown in Fig. 2 *B* demonstrated that the *SUC*2-*HIS*4 fusion allowed *his*4Δ, *HOL*1-1 cells to grow on histidinol. This capacity was dependent on the *HIS*4 insert, because clones transformed with vector sequences did not grow.

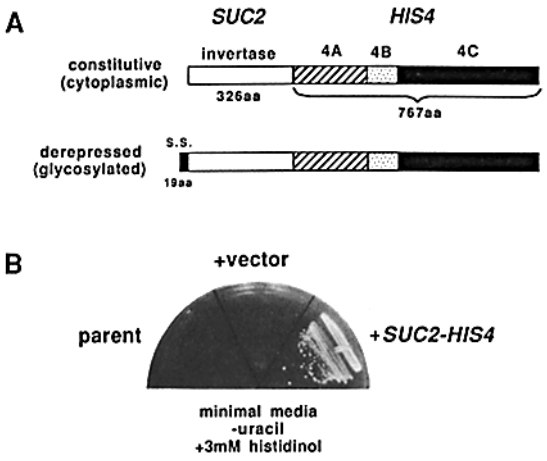


Figure 2. A *SUC*2-*HIS*4 gene fusion confers growth on histidinol. (*A*) The predicted structure of the protein products of the *SUC*2-*HIS*4 fusion are shown. *4A*, *4B*, and *4C* refer to separate domains of the trifunctional *HIS*4 protein; histidinol dehydrogenase activity is encoded by *4C*. The signal sequence of secretory invertase is shown as a small black rectangle at the NH_2 terminus of the derepressed fusion protein. (*B*) DYFC2-12B cells without plasmid (parent), transformed with vector sequences alone (+vector), or transformed with a multicopy plasmid containing the *SUC*2-*HIS*4 gene fusion (+*SUC*2-*HIS*4) were streaked onto minimal medium supplemented with 5% glucose, leucine, tryptophan, and 3 mM histidinol. These growth conditions should allow production of only the cytoplasmic fusion protein. The plate shown was incubated at 30°C for 3.5 d.

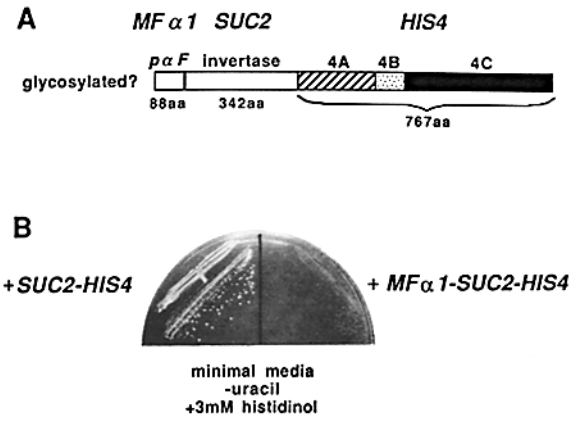


Figure 3. A *MFα1-SUC*2-*HIS*4 gene fusion does not confer growth on histidinol. (*A*) The predicted structure of the protein product of a *MFα1-SUC*2-*HIS*4 gene fusion is shown. The signal sequence-containing prepro region of α-factor is abbreviated as *pαF*. *4A*, *4B*, and *4C* are as described in Fig. 2 *A*. (*B*) DYFC2-12B cells transformed with multicopy plasmids containing either the *SUC*2-*HIS*4 (pSHE1) or *MFα1-SUC*2-*HIS*4 (pαSHF8) gene fusion were analyzed as described in Fig. 2 *B*.

A test of the second condition required elimination of the cytoplasmic invertase-histidinol dehydrogenase hybrid protein. This requirement was met by replacing the promoter of the *SUC*2 gene with upstream and coding sequences from the yeast *MFα1* gene (Fig. 3 *A*). *MFα1* encodes the precursor of the secreted mating pheromone α-factor (27). Multicopy plasmids (pαSHF8) bearing this tripartite fusion were introduced into yeast cells identical to those used in the previous experiment. In contrast to the pSHE1 transformants, cells containing pαSHF8 were not able to grow on minimal medium plus histidinol (Fig. 3 *B*).

Production of the glycosylated hybrid protein encoded by pαSHF8 was evaluated with transformed cells that were pulse-labeled with $[^{35}S]SO_4^{2-}$ in the presence or absence of tunicamycin, which inhibits asparagine-linked glycosylation. Extracts prepared from these cells were treated with invertase antiserum, which recognizes determinants encoded by the *SUC*2 portion of the fusion. Cells labeled in the absence of tunicamycin synthesized two species of fusion protein (Fig. 4, lane *1*). The predominant species migrated with a molecular mass of ~160 kD. In addition, a minor product of 140 kD was detected. When the transformant was labeled in the presence of tunicamycin (lane *2*), a single polypeptide that comigrated with the lower M_r form seen in lane *1* was made. The molecular mass of the unglycosylated material in lane *2* agreed with that predicted for an unmodified prepro-α-factor-invertase-*HIS*4 hybrid protein. Both polypeptides were plasmid encoded, in that cells containing vector se-

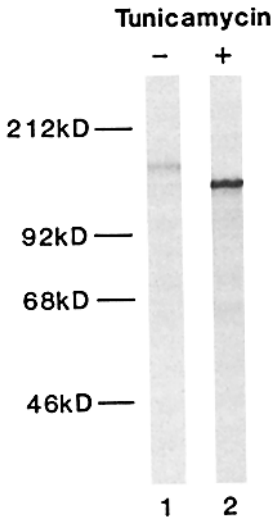


Figure 4. The *MFα1-SUC2-HIS4* fusion plasmid directs the synthesis of a glycoprotein. DYFC2-12B cells (2 OD_{600} per sample) transformed with pαSHF8 were labeled for 20 min at 30°C with $[^{35}S]SO_4^{2-}$ in the presence or absence of 10 μg/ml tunicamycin after a 15-min pretreatment with or without tunicamycin, respectively. Whole-spheroplast extracts were treated with anti-invertase serum and the immunoprecipitates were fractionated on a 7.5% SDS-polyacrylamide gel. Lanes *1* and *2* are individual tracks from a single polyacrylamide gel.

quences did not express any high M_r cross-reactive material (data not shown). Also, both products were under mating-type regulation, because neither *MATa* haploids nor *MATa/MATα* diploids transformed with pαSHF8 expressed any hybrid protein as expected for a gene under the control of *MATα* (data not shown). The nature of the 140-kD species seen in lane *1* was not examined further. In any event, this plasmid did not complement the chromosomal *his4* deletion. These data suggested that the pαSHF8 transformants were unable to grow on histidinol because the majority of hybrid protein was localized to the secretory pathway.

Isolation of Temperature-sensitive Histidinol Prototrophs

Inasmuch as the primary assumptions were confirmed by these experiments, the histidinol selection scheme was applied to yeast cells (DYFC2-12B; see Table I) containing the *MFα1-SUC2-HIS4* fusion plasmid. Transformants were mutagenized with ethyl methanesulfonate, allowed to recover overnight in minimal medium (-uracil) at 30°C, and plated on minimal medium (-uracil) with histidinol substituted for histidine. After incubation for 5–7 d at 30°C, plates were examined for the presence of histidinol prototrophs (histidinol$^+$). The frequency of mutations that allowed growth on histidinol was ~2.5×10^{-5} (Table II, round I). 440 histidinol prototrophs were then screened for Ts growth on rich medium at 37°C. Five clones exhibited both Ts and histidinol$^+$ phenotypes. Derivatives of all five isolates that had been cured of the plasmid were not able to grow on histidinol, confirming that the histidinol$^+$ phenotype was dependent on the plasmid. These cells also exhibited Ts growth on rich medium, demonstrating that thermosensitivity was caused by a chromosomal mutation. Retransformation of these clones with unmutagenized pαSHF8 restored growth on histidinol medium, demonstrating that this phenotype was also due to a chromosomal mutation, as opposed to plasmid-linked defects such as signal sequence mutations.

Secretion defects were tested directly by assaying for the accumulation of secretory precursors in plasmid-cured derivatives from four of the original isolates. Whole-cell extracts were fractionated by SDS-PAGE and immunoblotted with antiserum that reacts with prepro-α-factor. Three of the mutants accumulated a precursor form of α-factor that was not seen in wild-type cells.

A genetic relationship between the thermosensitive growth and α-factor accumulation phenotypes was evaluated by backcrossing two mutants to wild-type strains. Tetrads derived from these crosses exhibited low spore viability, and we discovered that this was because the original isolates were diploid. The unmutagenized parent strain was also diploid, possibly as a result of the transformation procedure. Haploid mutants were obtained from two of the isolates by mating with *MATa/MATα* diploids to form tetraploids, followed by two consecutive rounds of meiosis and tetrad analysis. Backcrosses of these haploid mutants to wild-type strains revealed that both Ts defects segregated as single mutations (two Ts and two wild-type spores in 19/19 tetrads analyzed in each cross). Thermosensitivity was inseparable from the prepro-α-factor accumulation property in that 10 out of 19 *MATα* spore clones were Ts and each accumulated prepro-α-factor. All nine wild-type progeny, however, failed to accumulate prepro-α-factor.

Heterozygous diploids, derived by mating both Ts mutants to wild-type strains, grew at 37°C, demonstrating that the Ts mutations were recessive. Precursor forms of carboxypeptidase Y and invertase detected in haploid mutant cells were absent in the heterozygotes, indicating that the accumulation phenotypes were recessive also. Because the parental strain was diploid, these recessive mutations may have been recovered by gene conversion or mitotic recombination events that occurred during or shortly after mutagenesis. Complementation analysis revealed that all five Ts mutations were allelic,

Table II. Histidinol Selection Scheme Enriches for Temperature-sensitive Translocation Mutants

	Round I*		Round II‡	
Stage of mutant isolation	Colonies analyzed/ total colonies	Fraction§ of total	Colonies analyzed/ total colonies	Fraction§ of total
Cells plated on histidinol medium	2×10^7	–	2.4×10^7	–
Histidinol prototrophs (histidinol$^+$)	492	2.5×10^{-5}	1,600	6.7×10^{-5}
Temperature-sensitive for growth	5‖/440	2.8×10^{-7}	58/600	6.4×10^{-6}
Accumulation of α-factor precursor¶	3/4	2.1×10^{-7}	7/40	1.1×10^{-6}

* DYFC2-12B diploid cells were used for mutant isolation.
‡ FC2-12B haploid cells were used for mutant isolation.
§ This value represents the number of colonies exhibiting a given phenotype (or the number expected to exhibit that phenotype, if all colonies were analyzed) divided by the number of cells plated on histidinol.
‖ All of these isolates are allelic.
¶ Immunoblots of whole cell lysates were probed with α-factor antiserum as described in Materials and Methods.

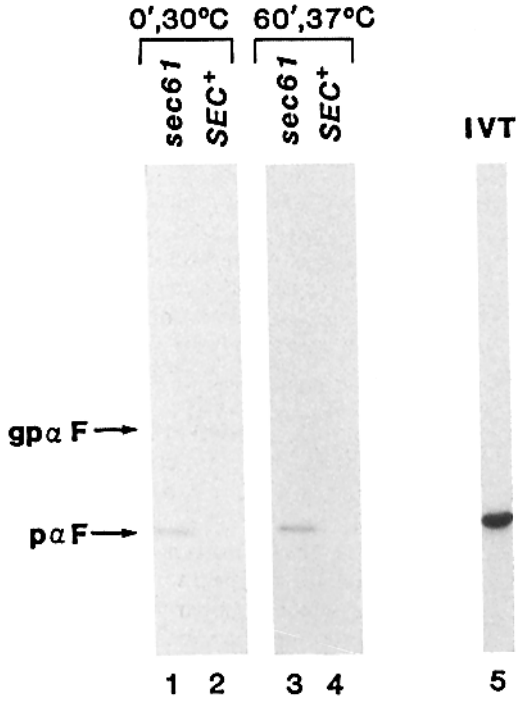


Figure 5. Unmodified prepro-α-factor accumulates at both the nonpermissive and semipermissive temperatures in *sec*61 strains. RDM 15-5B (*sec*61) and RDM 15-9B (*SEC*$^+$) cells were labeled with [^{35}S]SO_4^{2-} at 30°C for 30 min (lanes *1* and *2*) or at 37°C for 30 min after a 60-min incubation at 37°C (lanes *3* and *4*). Spheroplast pellet lysates were prepared and incubated with anti-α-factor serum. Immunoprecipitates were evaluated by SDS-PAGE on a 12.5% polyacrylamide gel. Each lane contains the material derived from 0.5 OD_{600} U of cells. Glycosylated prepro-α-factor in transit through the secretory pathway is designated *gpαF*; *pαF* refers to prepro-α-factor. *IVT* (lane *5*) is the in vitro translation product of prepro-α-factor mRNA. Lanes *1–5* were all derived from a single polyacrylamide gel.

although it was not certain that the mutations were of independent origin. The gene defined by these mutations was designated *sec*61.

To assess whether the proportion of Ts, histidinol$^+$ mutants among a population of mutagenized haploid cells might be substantially greater, the procedure was repeated using haploid FC2-12B cells transformed with pαSHF8. As expected, the frequency of "*sec*" mutants obtained from the histidinol selection was about fivefold greater when haploid cells were used (Table II, round II). All seven candidates obtained from round II complemented *sec*61. At least one additional complementation group was been identified among these isolates. The phenotype of strains bearing a mutant allele of this gene, *sec*62, is being investigated.

sec61 Cells Accumulate Unprocessed α-Factor Precursor

More information on the effect of the *sec*61 mutation was obtained by assessing the fate of prepro-α-factor expressed at the semipermissive and nonpermissive temperatures. Wild-type cells labeled for 30 min at either 30 or 37°C contained very little α-factor cross-reactive material (Fig. 5, lanes *2* and *4*) because of the rapid rate of α-factor secretion (22). At 30°C, a trace amount of core-glycosylated prepro-α-factor (*gpαF*) in transit through the early stages of the secretory pathway was detected (lane *2*).

When *sec*61 cells were labeled for 30 min at 30°C or 30 min at 37°C after a 60-min preshift to the nonpermissive temperature, a distinct species of α-factor precursor (*pαF*) was accumulated (Fig. 5, lanes *1* and *3*). This form of α-factor was not detected in wild-type cells at either temperature. The expression of a secretory defect at both 30 and 37°C was predicted by the selection scheme, which required that cells be partially defective at the growth temperature. Mutant cells also accumulated substantial amounts of α-factor precursor at 24°C. Additionally, *sec*61 strains grew slower than wild-type yeast at several temperatures, suggesting that there was no completely permissive temperature for *sec*61 function. Though a preshift to 37°C exaggerated the mutant phenotype, it did not completely inhibit α-factor secretion, in that even after a 60-min preshift to the nonpermissive temperature, newly synthesized biologically active pheromone was detected in the growth medium (not shown). The precursor that accumulated at the *sec*61 block comigrated with the in vitro translation product of hybrid selected *MFα1* mRNA electrophoresed on the same SDS-polyacrylamide gel (lane *5*). This result demonstrated that *sec*61 cells accumulated a form of α-factor that was not detectably modified either by glycosylation or proteolytic processing (22). One possible interpretation was that the *sec*61 mutation blocked an early event in protein translocation, prior to addition of core oligosaccharides in the lumen of the ER.

Multiple Unprocessed Secretory Proteins Accumulate in sec61 Cells

α-Factor shares early stages of the secretory pathway with other cell surface and vacuolar proteins (22), hence *sec*61 should interrupt the processing and localization of these proteins as well. This point was tested for the cell wall enzyme invertase and the vacuolar enzyme CPY by radiolabeling wild-type and mutant cells at 30 and 37°C, and treating spheroplast lysates with antisera reactive with each protein.

Evaluation of CPY precursor forms was complicated by the coincident electrophoretic mobilities of the unglycosylated precursor and the glycosylated mature species (50). This problem was circumvented by introduction of the *pep*4-3 mutation which blocks proenzyme cleavage (16). In *SEC pep*4-3 strains radiolabeled at 30 or 37°C (Fig. 6 *A*, lanes *2* and *4*), two forms of proCPY were seen: plCPY, a core glycosylated 67-kD precursor in transit through the ER, and a mature (*M*) 69-kD species that is found in the Golgi body and vacuole (50).

A novel form of proCPY was seen when *sec*61 *pep*4-3 strains were radiolabeled at 30 or 37°C after a 2-h incubation at the nonpermissive temperature. The new 58-kD species migrated as predicted for the primary translation product of the gene encoding CPY (53). At 30°C the *sec*61 defect was incomplete, in that both pl and mature forms of proCPY were also detected. After a 120-min preshift to 37°C, however, all of the newly synthesized proCPY accumulated at the *sec*61 block. With shorter 37°C preincubations (30 or 60 min), a fraction of the proCPY synthesized at 37°C escaped the *sec*61 block and appeared as the pl and mature forms (data not shown). These data suggested that the *sec*61 mutation was leaky, and required long temperature shifts to be fully expressed.

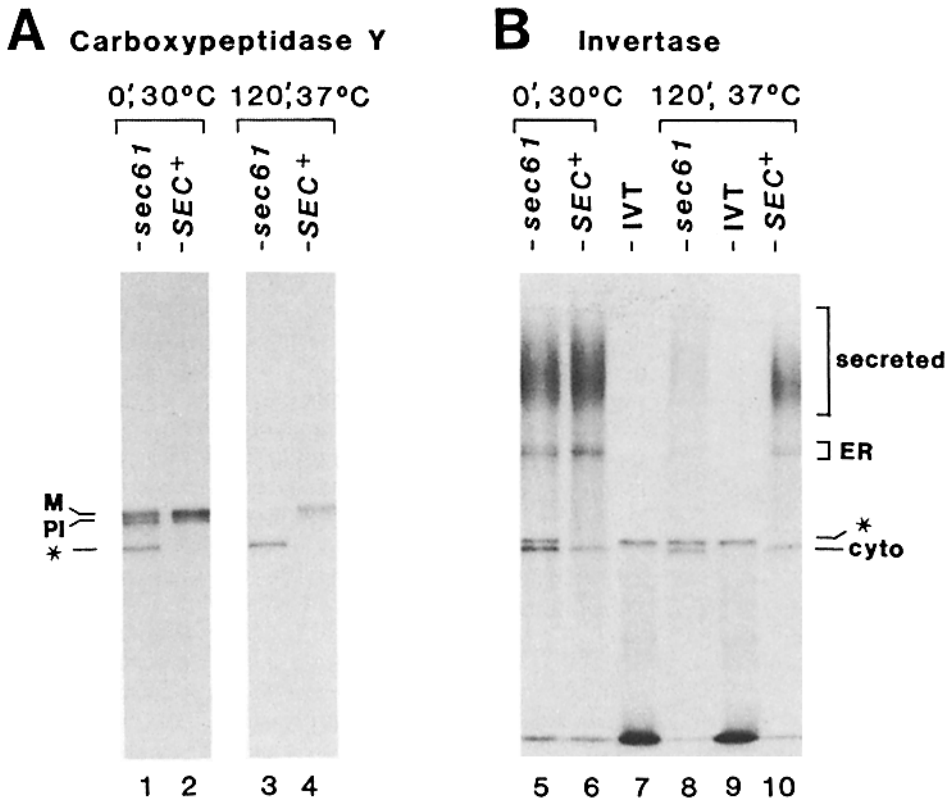


Figure 6. Unmodified precursors of invertase and CPY accumulate in *sec*61 cells. (*A*) The same extracts that were used for the experiment described in Fig. 5 were treated with anti-CPY serum. Immunoprecipitates were fractionated on a single 7.5% polyacrylamide gel. Lanes *1* and *2*, cells were labeled at 30°C for 30 min; lanes *3* and *4*, cells were labeled at 37°C for 30 min after a 120-min preincubation at 37°C. Lanes *1–4* contain the material derived from 0.2 OD_{600} U of cells. *PI*, *M*, and * refer to glycosylated proCPY in transit through the ER, mature proCPY in the Golgi body or vacuole, and proCPY accumulated at the *sec*61 block, respectively. (*B*) RDM 15-5B (*sec*61) and RDM 15-9B (*SEC*+) cells were labeled with $[^{35}S]SO_4^{2-}$ for 30 min at 30°C (lanes *5* and *6*) or for 30 min at 37°C after a 120-min incubation at 37°C (lanes *8* and *10*). Before each labeling, cultures were derepressed for invertase production in 0.1% glucose minimal medium at 30°C for 30 min (lanes *5* and *6*) or at 37°C for 30 min (lanes *8* and *10*). Whole-spheroplast lysates were prepared and treated with anti-invertase serum. Immune complexes were harvested and analyzed by SDS-PAGE on a 7.5% polyacrylamide gel. Lanes *5*, *6*, *8*, and *10* represent the total invertase derived from 0.5 OD_{600} U of cells. *IVT* (lanes *7* and *9*) is the in vitro translation product of secreted invertase mRNA. *ER*, *cyto*, and * refer to core glycosylated invertase in transit through the ER, cytoplasmic (unglycosylated) invertase, and secretory preinvertase accumulated by the *sec*61 mutation, respectively.

Additional information concerning the position of the *sec*61 block was revealed by analysis of radiolabeled invertase. Wild-type cells produced three discrete types of invertase at 30 or 37°C (Fig. 6 *B*, lanes *6* and *10*). Core-glycosylated intermediates in transit through the ER and highly glycosylated cell wall molecules were synthesized in response to glucose deprivation. A nonglycosylated cytoplasmic enzyme was made constitutively (9). Besides these species, *sec*61 mutant cells labeled at 30°C or after a 2-h incubation at 37°C accumulated an additional form (lanes *5* and *8*) that comigrated with the in vitro translation product of secretory invertase mRNA (Fig. 6 *B*, lanes *7* and *9*). The difference in M_r between the secreted and cytoplasmic primary translation products is due to the presence of a signal peptide at the amino terminus of the secretory preprotein (41). Comigration of *sec*61-specific invertase with the signal sequence-containing in vitro translation product indicated that the *sec*61 defect was imposed prior to signal peptide cleavage.

Based on the SDS gel mobility of precursors accumulated in *sec*61 cells, the block appeared to precede the addition of asparagine-linked core oligosaccharides, which are transferred to protein in the lumen of the ER. This prediction was tested directly by treating extracts from $[^{35}S]SO_4^{2-}$-labeled *sec*61 cells with the mannose-binding lectin Con A immobilized on Sepharose beads. Beads were recovered by centrifugation, pellet and supernatant fractions were treated with SDS, and the distribution of accumulated precursors was assessed by immune precipitation with α-factor and CPY antisera. The data in Fig. 7 *A*, lane *1* show the total complement of α-factor present in the *sec*61 extract derived from the same sample depicted in Fig. 5, lane *1*. When mixed with Con A-Sepharose and separated into pellet (lane *3*) and supernatant (lane *2*) fractions, glycosylated prepro-α-factor in transit through the ER was quantitatively recovered in the bound fraction, whereas prepro-α-factor accumulated by the *sec*61 mutation remains in the supernatant. The results obtained for proCPY were similar to those for α-factor. The data in Fig. 7 *B*, lane *4* shows the proCPY present in the total extract (same sample as Fig. 6, lane *1*). Mature and plCPY were bound to the Con A-Sepharose (lane *5*), although some material was lost in sample preparation. In contrast, the *sec*61-specific proCPY (*) was quantitatively recovered in the unbound fraction (lane *6*). Preinvertase accumulated in *sec*61 mutants behaved identically to proCPY and prepro-α-factor (data not shown). In all cases, the association was specific for mannose, as binding was prevented by the competitor, α-methylmannoside (data not shown).

sec61-Accumulated Preproteins Are Exposed to the Cytoplasm

The histidinol selection scheme demanded that a prepro-α-factor-invertase-*HIS*4 chimeric protein accumulate in a loca-

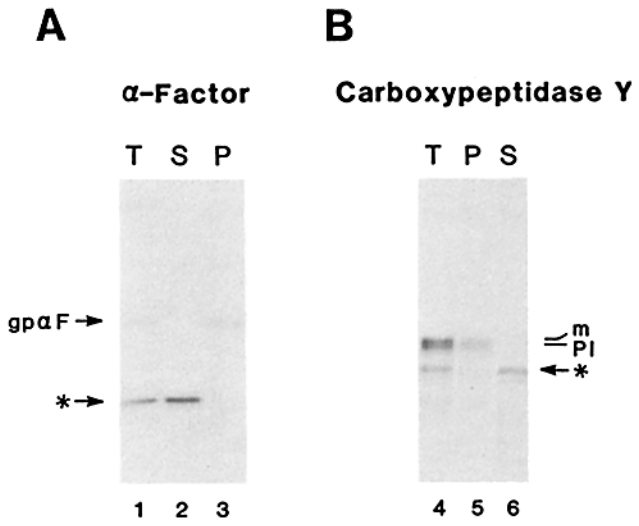


Figure 7. CPY and α-factor precursors accumulated at the *sec*61 block do not bind to Con A. (*A*) The $[^{35}S]SO_4^{2-}$ labeled extract (0.5 OD_{600} per lane) used to generate the sample shown in Fig. 5, lane *1* was directly immunoprecipitated with anti-α-factor serum (lane *1*), or was treated with Con A-Sepharose 4B and separated by sedimentation into bound (lane *3*) and free (lane 2) fractions prior to immunoprecipitation. *T*, *S*, and *P* are abbreviations for the total, supernatant, and pellet fractions. The position of the glycosylated α-factor precursor is indicated by *gpαF*; sec61-specific prepro-α-factor is indicated by *. (*B*) Same as in *A*, except the sample used (0.3 OD_{600} per lane) is identical to Fig. 6, lane *1*. Lane *4*, total CPY antigen; lane *5*, bound (pellet) proCPY species; lane *6*, free (supernatant) proCPY material. *PI*, *m*, and * are as described in Fig. 6 *A*.

tion accessible to histidinol. The most likely location of this fusion protein, and of preproteins whose localization was perturbed by the *sec*61 mutation, was in the cytoplasm or on the cytoplasmic face of the ER. The lack of detectable modifications on preproteins accumulated by the *sec*61 cells was consistent with this prediction. To test directly whether these precursors were exposed to the cytoplasmic compartment, extracts of $[^{35}S]SO_4^{2-}$-labeled *sec*61 cells were exposed to protease in the absence and presence of detergent. Proteolysis was terminated by the addition of TCA, and the quenched reactions were evaluated by immune precipitation with anti-α-factor and anti-CPY sera, followed by SDS-PAGE.

Extracts were prepared from a strain in which the disposition of precursor polypeptides and the integrity of ER vesicles could be evaluated simultaneously. For this purpose, a strain bearing mutant copies of the *sec*18 and *sec*61 genes was constructed. The *sec*18 mutation, which is nonpermissive at 30°C, causes core-glycosylated secretory proteins to accumulate within the lumen of the endoplasmic reticulum (9, 36). When *sec*18 *sec*61 mutants grown at 24°C were shifted to 30°C, a fraction of the newly synthesized prepro-α-factor was accumulated at the *sec*61 stage. The prepro-α-factor that escaped the *sec*61 block became core-glycosylated and arrested by the *sec*18 mutation. This glycosylated prepro-α-factor served as a reference for the lumen of the endoplasmic reticulum. The results in Fig. 8 *A* document the sensitivity of glycosylated prepro-α-factor and prepro-α-factor to proteinase K. When no protease was added (lanes *1* and *7*), glycosylated and unglycosylated forms of prepro-α-factor were seen. Upon exposure to protease for increasing lengths of time in the absence of detergent (lanes *2–6*), the amount of *sec*61 accumulated prepro-α-factor diminished gradually while the lowest mobility *sec*18 form (glycosylated prepro-α-factor) remained resistant to digestion. After a 20-min incubation with proteinase K, 90% of the prepro-α-factor was degraded, while 85% of the glycosylated form was resistant to proteolysis. The residual amount (10%) of prepro-α-factor that resisted protease action was similar to the fraction of NADPH cytochrome *c* reductase activity that was latent in this extract (see below). In the presence of detergent (lanes *8–12*), all species of prepro-α-factor were degraded rapidly; the digestion was essentially complete after 3-min (lane *10*).

As a control to estimate the proportion of lysed membranes in the extract, a mock-digested aliquot was centrifuged to obtain pellet and supernatant fractions. α-Factor immunoprecipitates of these fractions are displayed in lanes *13* and *14*. The extent of protection of *sec*18 prepro-α-factor was proportional to the amount of material that sedimented (compare lane *6* with lane *13*). Prepro-α-factor accumulated at the *sec*61 block, however, was almost completely protease sensitive in the absence of detergent (lane *6*), even though it sedimented quantitatively (lane *13*). Though clearly accessible to exogenous protease, the rate of prepro-α-factor digestion was accelerated two- to threefold by detergent. This may reflect increased exposure of this material to protease in the presence of detergent, implying that prepro-α-factor retained at the *sec*61 block was associated with some membrane-bound component, or directly with the phospholipid bilayer.

Protease sensitivity of accumulated prepro-α-factor was also examined in a homogenate prepared from a *sec*61 *SEC*18 strain. Although these samples did not contain the glycosylated control, the unglycosylated precursor was degraded at a rate similar to that seen in Fig. 8 *A*. Hence, proliferated ER membrane produced by the *sec*18 block (36) did not alter the behavior of the *sec*61 species.

Fig. 8 *B* documents the protease sensitivity of the *sec*18 and *sec*61 forms of proCPY present in the same extract. In the absence of detergent, proCPY that accumulated at the *sec*18 block was refractory to proteolysis over a 20-min incubation, while proCPY held at the *sec*61 block was completely digested within 1.5 min (lanes *2–6*). In the presence of detergent, all species of proCPY were susceptible to proteolysis, though the glycosylated proCPY was only partially digested (lanes *8–12*). The results in lane *13* demonstrate that the majority of proCPY present in this extract was sedimentable.

To ensure that the surface of the ER membrane was accessible to proteinase K, assays were performed to determine the latency of the cytoplasmically exposed ER membrane enzyme, NADPH cytochrome *c* reductase (25), to its substrate cytochrome *c*. Assays performed with the same extract used in these experiments indicated that only 10% of the reductase activity was latent.

Prepro-α-Factor Accumulated in sec61 Cells Is Particulate

The particulate nature of prepro-α-factor in extracts prepared from *sec*18 *sec*61 mutant cells (Fig. 8 *A*, lane *13*) was examined in more detail. Spheroplast lysates were subjected to differential centrifugation and subcellular fractions were assayed for NADPH cytochrome *c* reductase and glyceralde-

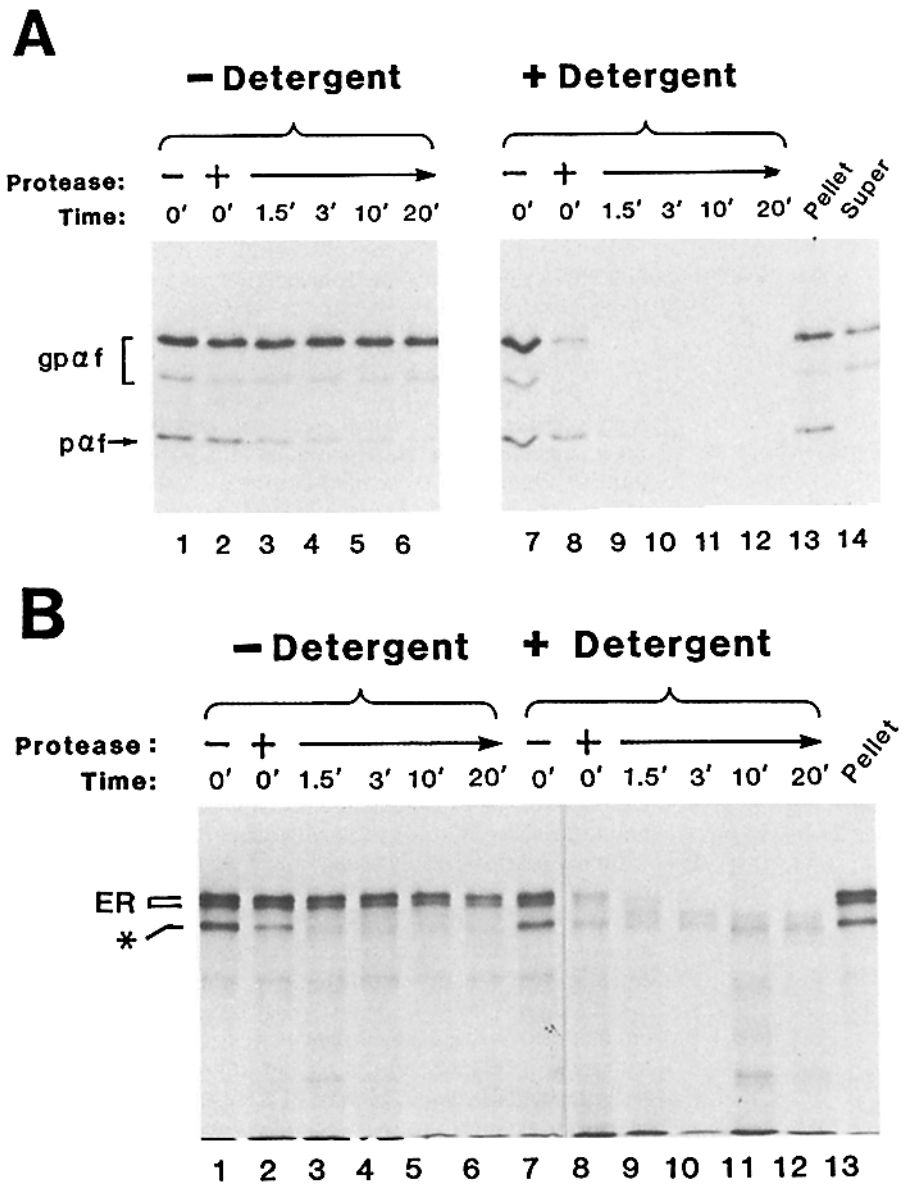


Figure 8. Proteolysis of prepro-α-factor and proCPY in *sec*18 *sec*61 cell extracts. RDM 15-3A cells were labeled and lysates were prepared as described in Materials and Methods. Aliquots of the lysate were mock digested (lanes *1* and *7*) or were treated with 300 μg/ml proteinase K in the absence (lanes *2–6*) or presence (lanes *8–12*) of 0.4% Triton X-100 for 0–20 min on ice. A separate aliquot was fractionated by centrifugation into pellet (lane *13*) and supernatant (lane *14*) fractions. (*A*) Samples were quenched with TCA, immunoprecipitated with anti-α-factor serum, and evaluated by SDS-PAGE on a 12.5% polyacrylamide gel. Each lane contains the amount of prepro-α-factor precipitated from 1.2 OD_{600} U of cells. α-Factor precursors accumulated by the *sec*18 and *sec*61 mutations are designated *gpαF* and *pαF*, respectively. (*B*) Aliquots from the proteolysis reactions were treated with anti-CPY serum and immunoprecipitates were analyzed by SDS-PAGE on a 7.5% polyacrylamide gel. Each lane contains total CPY antigen derived from 0.4 OD_{600} U of cells. ProCPY accumulated at the *sec*18 and *sec*61 stages is designated *ER* and *, respectively.

hyde-3-phosphate dehydrogenase, a cytoplasmic enzyme. Aliquots of each fraction were also evaluated by SDS-PAGE and immunoblotting with α-factor antiserum to monitor the fractionation properties of prepro-α-factor accumulated at the *sec*18 and *sec*61 blocks. As in the previous experiment, precursors accumulated by the *sec*18 mutation served as a marker for the lumen of the ER. The enzyme assay results are presented in Table III. As expected, the cytosolic marker was quantitatively recovered in the high-speed supernatant (HSS). NADPH cytochrome *c* reductase sedimented predominantly in the high speed pellet fraction; ~30% of the reductase activity sedimented in the low speed pellet fraction. An anti-α-factor immunoblot is shown in Fig. 9. The three lower mobility forms of α-factor precursor seen in the extract (lane *1*) corresponded to different glycosylated species (22) that were accumulated in *sec*18 mutants (compare the *sec*18 extract in lane *7* with the wild-type extract in lane *6*). The highest mobility band in lanes 1–5 corresponded to prepro-α-factor accumulated at the *sec*61 block, and comigrated with the in vitro translation product. The unglycosylated prepro-α-factor sedimented exclusively with the particulate fractions, whereas lumenal glycosylated prepro-α-factor was distributed between the high speed supernatant and pellet fractions, presumably due to some rupture of the ER membrane during cell lysis. The sedimentation of *sec*61-accumulated prepro-α-factor was not influenced by the *sec*18 mutation, in that extracts of *sec*61 mutants yielded similar results. In addition, proCPY accumulated in this mutant was recovered in the sedimentable fraction of a lysate (data not shown).

Discussion

A genetic selection has been devised to identify genes required for secretory protein translocation into the ER. This selection procedure demands that a signal peptide-containing cytoplasmic enzyme, whose location has been diverted to the secretory pathway, be retained in contact with the

Table III. Distribution of Marker Enzymes in sec18 sec61 Subcellular Fractions

Enzyme	Cellular location	Subcellular fraction	% Activity	Specific activity
Glyceraldehyde-3-P dehydrogenase	Cytoplasm	Extract	100	2.55
		HSS	112	5.32
		HSP	0.2	0.013
NADPH cytochrome *c* reductase	ER membrane	Extract	100	0.178
		HSS	5.6	0.026
		HSP	80.5	0.307

Subcellular fractions from the experiment depicted in Fig. 9 were assayed for their content of glyceraldehyde-3-P dehydrogenase and NADPH cytochrome *c* reductase activities as described in Materials and Methods. HSS and HSP, high-speed supernatant and pellet fractions, respectively. Each value presented is the average of the initial rates of activity measured in three independent trials. Specific activity is expressed as $\Delta A_{550\ nm}$ per minute per milligram of protein for NADPH cytochrome *c* reductase and $\Delta A_{340\ nm}$ per minute per milligram of protein for glyceraldehyde-3-P dehydrogenase.

cytosol because of a defect in the cellular translocation machinery. Mutations in two genes have been isolated that cause temperature sensitive growth and accumulation of prepro-α-factor. Because the selection demands that mutant cells express a partial defect at a temperature that permits growth (30°C), accumulation of α-factor precursor is seen at 30°C and is enhanced at the restrictive temperature (37°C). These phenotypes are the result of single lesions, inasmuch as temperature sensitivity and secretory protein accumulation cosegregate when mutants are backcrossed to wild type strains.

*sec*61, the mutant described in this report, also accumulates preinvertase and procarboxypeptidase (proCPY). Presumably other secretory proteins that are essential for yeast cell growth are also blocked in this mutant. We have not yet examined membrane protein precursors, many of which do not contain classical amino-terminal signal sequences (18, 34, 52). These properties could alter the assembly pathway of a protein. This possibility may be tested by evaluating the integration of membrane proteins into the ER membrane of *sec*61 mutant cells.

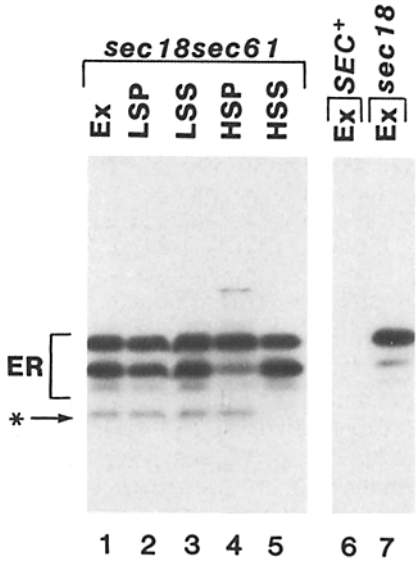


Figure 9. Fractionation of prepro-α-factor species accumulated in *sec*18 *sec*61 cells. Extracts of RDM 15-3A cells (*sec*18 *sec*61) were prepared and fractionated as described in Materials and Methods. Protein (30 μg) from each fraction was electrophoresed on a 12.5% polyacrylamide gel and immunoblotted with anti-α-factor serum. Lane *1*, whole extract; lane *2*, 660 - *g* pellet; lane *3*, 660 - *g* supernatant; lane *4*, 100,000 - *g* pellet; lane *5*, 100,000 - *g* supernatant. *Ex*, *LSP*, *LSS*, *HSP*, and *HSS* refer to extract, low-speed and high-speed pellet, and supernatant fractions, respectively. *ER* and * indicate forms of prepro-α-factor arrested by the *sec*18 and *sec*61 mutations. Unfractionated extracts of RDM 15-9B (*SEC*$^+$) and RDM 15-10D (*sec*18) strains were prepared, subjected to SDS-PAGE, and immunoblotted as described above. Lane *6*, whole extract from RDM 15-9B; lane *7*, whole extract from RDM 15-10D.

Preproteins accumulated at the *sec*61 block are not glycosylated, and their signal sequences have not been removed. These properties are expected for a molecule that has not yet engaged, or has only partly penetrated the ER membrane bilayer. Precursor forms of two secretory proteins are sedimentable but susceptible to exogenous proteolytic attack though the ER membrane remains largely intact.

α-Factor precursor made in an in vitro yeast protein synthesis reaction is soluble and becomes sedimentable only after translocation into ER vesicles (15). In contrast, both α-factor and CPY precursors sediment along with membranes in extracts of *sec*61 cells. This property could represent an intermediate stage in the translocation process in which precursors become firmly associated with the ER membrane (6), or could simply result from aggregation or nonspecific binding of precursors to membranes within mutant cells. Synthetic signal peptides and intact secretory precursor proteins have the capacity to insert into and at least partly through lipid monolayers and vesicle bilayers in vitro (3, 38). The physical basis of the sedimentability of precursor proteins may be established by fractionation of membranes from *sec*61 mutant cells.

Comparable blocks in secretion have been explored with *Escherichia coli*. Secretory precursors accumulate in thermosensitive mutants such as SecA, SecB, and SecY (21, 26, 39), or in wild-type cells that express high levels of maltose binding protein-β-galactosidase hybrid proteins (20). Cell fractionation experiments have shown that the precursors distribute between the periplasmic, membrane, and cytosolic fractions in proportions that are influenced by the monovalent ion concentration (20). In these conditions the preproteins may associate with the cytoplasmic membrane nonspecifically via the hydrophobic signal peptide.

A specific effect on the initial step in the secretory pathway, as opposed to a more general disruption of protein transport, is indicated by the pattern of precursors that appear in *sec*61 cells at the semirestrictive temperature, 30°C. Accumulation of core- or highly-glycosylated forms of α-factor, invertase, or CPY is expected in mutants with impaired intercompartmental protein transport from the ER or the Golgi body (9, 22, 50). In contrast, *sec*61 mutants show no such accumulation.

*sec*61 Mutant cells require 37°C preincubation periods of at least 2 h before the secretion block becomes complete. This long lag time suggests that the *sec*61 mutant protein is thermolabile for synthesis or association with an oligomeric complex, rather than thermosensitive for function. Alternatively, the mutant protein may be thermolabile in performing a modification that is required for the activity of another component of the translocation machinery. In either case, the defect would become more pronounced only as active species are replaced by inactive forms.

Two other yeast mutants, *sec*53 and *sec*59, are defective in an early stage of secretory protein biogenesis (12). Phenotypic characterization demonstrated that these Ts mutants accumulate underglycosylated, inactive precursors at the nonpermissive temperature (12). Initial protease protection experiments suggested that preinvertase is abortively translocated, remaining tightly associated with the ER membrane but accessible to trypsin in a homogenate (11). More refined methods for performing proteolysis experiments, however, have now shown that the partially glycosylated prepro-α-factor, invertase, and proCPY that accumulate in these mutants are completely protected against trypsin and protease K attack in the absence of, but not in the presence of detergent (10a). Because invertase accumulated in *sec*53 and *sec*59 mutants has had its signal peptide removed, is partially glycosylated and apparently resides within the lumen of the ER (10a), the *sec*61 defect must precede the block imposed by *sec*53 or *sec*59. In addition, the *sec*61 mutation complements both *sec*53 and *sec*59 mutations, indicating that *sec*61 defines a new function in the secretory pathway.

By characterizing the phenotypes of additional import mutants we may be able to reconstruct the events that occur during protein translocation in vivo (35). An examination of the properties of these mutants in the yeast in vitro protein translocation assay should further our understanding of the molecular mechanism of protein translocation into the yeast endoplasmic reticulum.

We thank Roy Parker for providing strains and plasmids, Jon Rothblatt for reading the manuscript and supplying us with anti-α-factor antiserum, Michael Culbertson for strains, Mitch Bernstein for hybrid-selected prepro-α-factor mRNA and rabbit reticulocyte lysate, Peter Böhni and Linda Hicke for constructing YCp50^{EB}, and Jasper Rine for his many valuable suggestions. We also thank Peggy McCutchan Smith for help in preparing the manuscript.

This work was supported by grants from the National Institutes of Health (GM-26755) and the National Science Foundation (DCB 02552).

Received for publication 17 February 1987, and in revised form 6 April 1987.

References

1. Bernstein, M., W. Hoffmann, G. Ammerer, and R. Schekman. 1985. Characterization of a gene product (Sec53p) required for protein assembly in the yeast endoplasmic reticulum. *J. Cell Biol.* 101:2374–2382.
2. Blobel, G., and B. Dobberstein. 1975. Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *J. Cell Biol.* 67:852–862.
3. Briggs, M. S., L. M. Gierasch, A. Zlotnick, J. D. Lear, and W. F. DeGrado. 1985. In vivo function and membrane binding properties are correlated for *Escherichia coli LamB* signal peptides. *Science (Wash. DC).* 228:1096–1099.
4. Burnette, W. N. 1981. "Western Blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* 112:195–203.
5. Carlson, M., and D. Botstein. 1982. Two differentially regulated mRNAs with different 5′ ends encode secreted and intracellular forms of yeast invertase. *Cell.* 28:145–154.
6. Connolly, T., and R. Gilmore. 1986. Formation of a functional ribosome-membrane junction during translocation requires the participation of a GTP-binding protein. *J. Cell Biol.* 103:2253–2262.
7. Emr, S. D., I. Schauer, W. Hansen, P. Esmon, and R. Schekman. 1984. Invertase-β-galactosidase hybrid proteins fail to be transported from the endoplasmic reticulum in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 4:2347–2355.
8. Emr, S. D., R. Schekman, M. C. Flessel, and J. Thorner. 1983. An *MFα1-SUC2* (α-factor-invertase) gene fusion for study of protein localization and gene expression in yeast. *Proc. Natl. Acad. Sci. USA.* 80:7080–7084.
9. Esmon, B., P. Novick, and R. Schekman. 1981. Compartmentalized assembly of oligosaccharides on exported glycoproteins in yeast. *Cell.* 25:451–460.
10. Evans, E. A., R. Gilmore, and G. Blobel. 1986. Purification of microsomal signal peptidase as a complex. *Proc. Natl. Acad. Sci. USA.* 83:581–585.
10a. Felman, R. I., M. Bernstein, and R. Schekman. 1987. Product of *sec*53 is required for folding and glycosylation of secretory proteins in the lumen of the yeast endoplasmic reticulum. *J. Biol. Chem.* 262:9332–9339.
11. Ferro-Novick, S., W. Hansen, I. Schauer, and R. Schekman. 1984. Genes required for completion of import of proteins into the endoplasmic reticulum in yeast. *J. Cell Biol.* 98:44–53.
12. Ferro-Novick, S., P. Novick, C. Field, and R. Schekman. 1984. Yeast secretory mutants that block the formation of active cell surface enzymes. *J. Cell Biol.* 98:35–43.
13. Gilmore, R., and G. Blobel. 1982. Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor. *J. Cell Biol.* 95:470–477.
14. Gilmore, R., and G. Blobel. 1985. Translocation of secretory proteins across the microsomal membrane occurs through an environment accessible to aqueous perturbants. *Cell.* 42:497–505.
15. Hansen, W., P. D. Garcia, and P. Walter. 1986. In vitro translocation across the yeast endoplasmic reticulum: ATP-dependent post-translational translocation of the prepro-α-factor. *Cell.* 45:397–406.
16. Hemmings, B. A., G. S. Zubenko, A. Hasilik, and E. W. Jones. 1981. Mutant defective in processing of an enzyme located in the lysosome-like vacuole of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA.* 78: 435–439.
17. Hinnen, A., J. B. Hicks, and G. R. Fink. 1978. Transformation of yeast. *Proc. Natl. Acad. Sci. USA.* 75:1929–1933.
18. Hoffmann, W. 1985. Molecular characterization of the *CAN1* locus in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 260:11831–11837.
19. Ito, H., Y. Fukuda, K. Murata, and A. Kumura. 1983. Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* 153:163–168.
20. Ito, K., P. J. Bassford, Jr., and J. R. Beckwith. 1981. Protein localization in *E. coli*: is there a common step in the secretion of periplasmic and outer-membrane proteins? *Cell.* 24:707–717.
21. Ito, K., M. Wittekind, M. Nomura, K. Shiba, T. Yura, A. Miura, and H. Nashimoto. 1983. A temperature-sensitive mutant of *E. coli* exhibiting slow processing of exported proteins. *Cell.* 32:789–797.
22. Julius, D., R. Schekman, and J. Thorner. 1984. Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. *Cell.* 36: 309–318.
23. Kaiser, C. A., and D. Botstein. 1986. Secretion-defective mutations in the signal sequence for *Saccharomyces cerevisiae* invertase. *Mol. Cell. Biol.* 6:2382–2391.
24. Keesey, J. K., R. Bigelis, and G. R. Fink. 1979. The product of the *his4* gene cluster in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 254:7427–7433.
25. Kubota, S., Y. Yoshida, H. Kumaoka, and A. Furumichi. 1977. Studies on the microsomal electron-transport system of anaerobically grown yeast. V. Purification and characterization of NADPH-cytochrome *c* reductase. *J. Biochem.* 81:197–206.
26. Kumamoto, C. A., and J. Beckwith. 1983. Mutations in a new gene, *sec*B, cause defective protein localization in *Escherichia coli*. *J. Bacteriol.* 154:253–260.
27. Kurjan, J., and I. Herskowitz. 1982. Structure of a yeast pheromone gene (*MFα*): a putative α-factor precursor contains four tandem repeats of mature α-factor. *Cell.* 30:933–943.
28. Kurzchalia, T. V., M. Wiedmann, A. S. Girshovich, S. Bochkareva, H. Bielka, and T. A. Rapoport. 1986. The signal sequence of nascent preprolactin interacts with the 54K polypeptide of the signal recognition particle. *Nature (Lond.).* 320:634–636.
29. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680–685.
30. Maniatis, T., E. F. Frisch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
31. Markwell, M. A. K., S. M. Haas, L. L. Bieber, and N. N. Tolbert. 1978. A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal. Biochem.* 7:206–210.
32. Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacterio-

phage SP6 promoter. *Nucleic Acids Res.* 12:7035–7056.
33. Meyer, D. I., E. Krause, and B. Dobberstein. 1982. Secretory protein translocation across membranes—the role of the "docking protein." *Nature (Lond.).* 297:647–650.
34. Nakayama, N., A. Miyajima, and K. Arni. 1985. Nucleotide sequences of *STE2* and *STE3*, cell type-specific sterile genes from *Saccharomyces cerevisiae. EMBO (Eur. Mol. Biol. Organ.) J.* 4:2643–2648.
35. Novick, P., S. Ferro, and R. Schekman. 1981. Order of events in the yeast secretory pathway. *Cell.* 25:461–469.
36. Novick, P., C. Field, and R. Schekman. 1980. Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. *Cell.* 21:205–215.
37. Novick, P., and R. Schekman. 1979. Secretion and cell-surface growth are blocked in a temperature-sensitive mutant of *Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA.* 76:1858–1862.
38. Ohno-Iwashita, Y., P. Wolfe, K. Ito, and W. Wickner. 1984. Processing of preproteins by liposomes bearing leader peptidase. *Biochemistry.* 23:6178–6184.
39. Oliver, D. B., and J. Beckwith. 1981. *E. coli* mutant pleiotropically defective in the export of secreted proteins. *Cell.* 25:765–772.
40. Parker, R., and C. Guthrie. 1985. A point mutation in the conserved hexanucleotide at a yeast 5′ splice junction uncouples recognition, cleavage, and ligation. *Cell.* 41:107–118.
41. Perlman, D., H. O. Halvorson, and L. E. Cannon. 1982. Presecretory and cytoplasmic invertase polypeptides encoded by distinct mRNAs derived from the same structural gene differ by a signal sequence. *Proc. Natl. Acad. Sci. USA.* 79:781–785.
42. Rothblatt, J. A., and D. I. Meyer. 1986*a*. Secretion in yeast: reconstitution of the translocation and glycosylation of α-factor and invertase in a homologous cell-free system. *Cell.* 44:619–628.
43. Rothblatt, J., and D. I. Meyer. 1986*b*. Secretion in yeast: translocation and glycosylation of prepro-α-factor *in vitro* can occur via an ATP-dependent post-translational mechanism. *EMBO (Eur. Mol. Biol. Organ.) J.* 5: 1031–1036.
44. Rothman, J. E., and H. F. Lodish. 1977. Synchronized transmembrane insertion and glycosylation of a nascent membrane protein. *Nature (Lond.).* 264:775–780.
45. Schauer, I., S. Emr, C. Gross, and R. Schekman. 1985. Invertase signal and mature sequence substitutions that delay intercompartmental transport of active enzyme. *J. Cell Biol.* 100:1664–1675.
46. Scott, J., and R. Schekman. 1980. Lyticase: endoglucanase and protease activities that act together in yeast cell lysis. *J. Bacteriol.* 142:414–423.
47. Sharma, S., L. Rodgers, J. Brandsma, M.-J. Gething, and J. Sambrook. 1985. SV40 T antigen and the exocytotic pathway. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1479–1489.
48. Sherman, F., G. R. Fink, and J. B. Hicks. 1983. Methods in Yeast Genetics: A Laboratory Manual. Revised edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
49. Simon, K., E. Perara, and V. R. Lingappa. 1987. Translocation of globin fusion proteins across the endoplasmic reticulum membrane in *Xenopus laevis* oocytes. *J. Cell Biol.* 104:1165–1172.
50. Stevens, T., B. Esmon, and R. Schekman. 1982. Early stages in the yeast secretory pathway are required for transport of carboxypeptidase Y to the vacuole. *Cell.* 30:439–448.
51. Stinchcomb, D. T., C. Mann, and R. W. Davis. 1982. Centromeric DNA from *Saccharomyces cerevisiae. J. Mol. Biol.* 158:157–179.
52. Tanaka, J. I., and G. R. Fink. 1985. The histidine permease gene (*HIP1*) of *Saccharomyces cerevisiae. Gene (Amst.).* 38:205–214.
53. Valls, L. A., C. P. Hunter, J. H. Rothman, and T. H. Stevens. 1987. Protein sorting in yeast: the localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide. *Cell.* 48:887–897.
54. Walter, P., and G. Blobel. 1980. Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. *Proc. Natl. Acad. Sci. USA.* 77:7112–7116.
55. Walter, P., and G. Blobel. 1982. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature (Lond.).* 229:691–698.
56. Walter, P., I. Ibrahimi, and G. Blobel. 1981. Translocation of proteins across the endoplasmic reticulum I. Signal recognition particle (SRP) binds to in vitro assembled polysomes synthesizing secretory protein. *J. Cell Biol.* 91:545–550.
57. Waters, M. G., and G. Blobel. 1986. Secretory protein translocation in a yeast cell-free system can occur post-translationally and requires ATP hydrolysis. *J. Cell Biol.* 102:1543–1550.
58. Waters, M. G., W. J. Chirico, and G. Blobel. 1986. Protein translocation across the yeast microsomal membrane is stimulated by a soluble factor. *J. Cell Biol.* 103:2629–2636.
59. Wickerham, L. J. 1946. A critical evaluation of the nitrogen assimilation tests commonly used in the classification of yeasts. *J. Bacteriol.* 52: 293–301.

LIPID ANCHORING OF ARCHAEOSORTASE SUBSTRATES AND MID-CELL GROWTH IN HALOARCHAEA

Mohd Farid Abdul-Halim[1], Stefan Schulze[1], Anthony DiLucido[1], Friedhelm Pfeiffer[2], Alexandre W. Bisson Filho[3*] and Mechthild Pohlschroder[1*]

[1]University of Pennsylvania, Department of Biology, Philadelphia, PA 19104, USA.

[2]Computational Biology Group, Max-Planck-Institute of Biochemistry, 82152 Martinsried, Germany.

[3]Department of Biology, Rosenstiel Basic Medical Science Research Center, Brandeis University, Waltham, MA 02454.

[*]Corresponding authors. Emails: pohlschr@sas.upenn.edu (M.P.); bisson@brandeis.edu (A.W.B-F)

The archaeal cytoplasmic membrane provides an anchor for many surface proteins. Recently, a novel membrane anchoring mechanism involving a peptidase, archaeosortase A (ArtA) and C-terminal lipid attachment of surface proteins was identified in the model archaeon *Haloferax volcanii*. ArtA is required for optimal cell growth and morphogenesis, and the S-layer glycoprotein (SLG), the sole component of the *H. volcanii* cell wall, is one of the targets for this anchoring mechanism. However, how exactly ArtA function and regulation control cell growth and mor-phogenesis is still elusive. Here, we report that archaeal homologs to the bacterial phosphatidylserine synthase (PssA) and phosphatidylserine decarboxylase (PssD) are involved in ArtA-dependent protein maturation. *H. volcanii* strains lacking either HvPssA or HvPssD exhibited motility, growth and morphological phenotypes similar to those of Δ*artA*. Moreover, we showed the loss of covalent lipid attachment to SLG in the Δ*hvpssA* mutant and that proteolytic cleavage of the ArtA substrate HVO_0405 was blocked in the Δ*hvpssA* and Δ*hvpssD* strains. Strikingly, ArtA, HvPssA, and HvPssD GFP-fusions co-localized to the mid position of *H. volcanii* cells, strongly supporting that they are involved in the same pathway. Finally, we have shown that the SLG is also recruited to the mid cell prior to being secreted and lipid-anchored at the cell outer surface. Collectively, our data suggest haloarchaea use the mid cell as the main surface processing hotspot for cell elongation, division and shape determination.

Introduction

Microbial cell surface proteins play critical roles in many important biological processes, including bioenergetics, mediation of intercellular communication, nutrient uptake, surface adhesion, and motility. Cell surface proteins also play important roles in cell elongation and shape maintenance, but how this is achieved in archaea is not well understood (1).

The structural organization of cellular surfaces is one important readout of how cells coordinate growth, morphogenesis, and division. In both bacteria and eukaryotes, a multitude of growth modes have been characterized, with cells inserting new envelope material almost all along the cell surface (2), bipolarly (3), unipolarly (4), and

in some cases, different modes can be interchangeable (5, 6). In the case of archaea, which lack a peptidoglycan cell wall, glycosylated S-layer and other proteins are commonly the sole component of the cell envelope (7, 8), where they typically show a 2D crystal-like arrangement. This poses an interesting problem for archaeal surface protein organization, and currently there is no data about the mechanisms of archaeal cell elongation control (9).

While many proteins are anchored to the cell surface via transmembrane (TM) domain insertion into the membrane, some are anchored through covalent N-terminal attachment of a lipid moiety (8). Recently, a novel mechanism was discovered whereby proteins are anchored to the membrane through a lipid moiety covalently attached to a processed C-terminus. In archaeal cells, processing and lipid modification of these C-terminal anchored proteins are mediated by enzymes known as archaeosortases, with archaeosortase A (ArtA) of the model archaeon *Haloferax volcanii* being the most studied example (10–13). Proteins recognized and processed by *H. volcanii* ArtA contain a distinct C-terminal tripartite

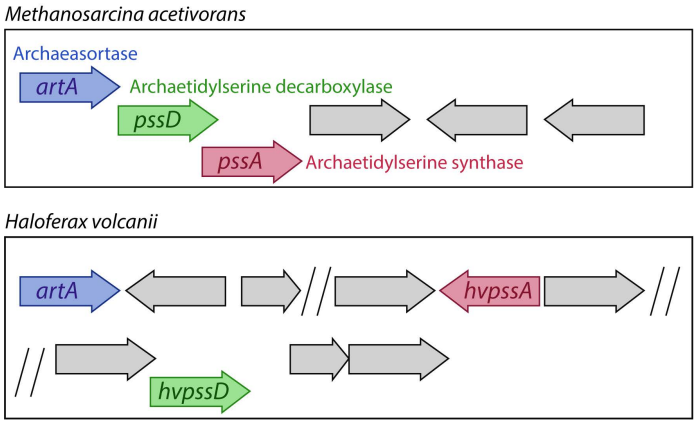


Figure 1: Schematic representation of *artA*, *pssA*, and *pssD* distribution across Euryarchaeota. **(top)** *M. acetivorans* and **(bottom)** *H. volcanii* genomic organization of *artA*, and genes encoding homologs to PssA (HvPssA) and PssD (HvPssD).

structure consisting of a conserved PGF motif, followed by a hydrophobic domain and then a stretch of positively charged residues. Molecular biological and biochemical analyses determined that ArtA does indeed process a diverse set of proteins, including both Tat and Sec substrates, that have been shown to play roles in motility and mating (10, 12). Most notably, this includes the S-layer glycoprotein (SLG), which is the sole component of the *H. volcanii* cell wall.

A previous *in silico* study by Haft and coworkers noted that in *Methanosarcina acetivorans* C2A, *Methanosarcina mazei* Gö1, as well as several other methanogens, the *artA* gene is located next to the gene that encodes an archaeal homolog of bacterial phosphatidylserine synthase (PssA) (14). Based on the degree of sequence similarity and its substrate specificity, the archaeal PssA homolog belongs to the PssA subclass II, similar to PssA found in gram-positive bacteria such as *Bacillus subtilis*, as opposed to PssA of gram-negative bacteria, such as *Escherichia coli,* which belong to subclass I (15).

Work in *B. subtilis* has elucidated most of the biochemistry involved in the reaction catalyzed by PssA, which involves the transfer of a diacylglycerol moiety from a CDP-phosphatidyl lipid to L-serine to make phosphatidylserine (16, 17). Phosphatidylserine can subsequently be decarboxylated to phosphatidylethanolamine by the enzyme phosphatidylserine decarboxylase (PssD), which has been characterized from *Sinorhizobium meliloti* and *B. subtilis* (17, 18). However, unlike bacterial PssA, *in vitro* study of the archaeal PssA homolog from *Methanothermobacter thermautotrophicus* (MTH_1027) revealed that this protein catalyzes the transfer of the archaetidic acid moiety of CDP-archaeol onto the hydroxyl group of L-serine to form the polar lipid archaetidylserine (CDP-2,3-di-O-geranylgeranyl-sn-glycerol:L-serine O-archaetidyltransferase) (15).

Mirroring the phosphatidylethanolamine biosynthesis reaction in bacteria, it was postulated that archaetidylserine could also undergo decarboxylation to archaetidylethanolamine by an archaeal PssD homolog, a putative archaetidylserine decarboxylase.

Distant homologs to PssA and PssD are encoded in the *H. volcanii* genome, which we refer to as HvPssA (HVO_1143) and HvPssD (HVO_0146). In this study, we show that HvPssA and HvPssD are involved in ArtA-dependent C-terminal protein maturation, which involves proteolytic cleavage and lipid anchoring. An interplay between ArtA, HvPssA and HvPssD is further supported by their colocalization at mid cell. These analyses reveal, to the best of our knowledge for the first time, that cell elongation happens from mid cell in archaea.

Results

Synteny of *artA*, *pssA* and *pssD* genes in *Methanosarcina strains*.

Based on the juxtaposition of the genes encoding the archaeosortase (*artA*) and the putative membrane lipid biosynthesis archaetidylserine synthase (*pssA*) in *Methanosarcina* (14), we hypothesized that the *H. volcanii* homologs of PssA (HvPssA) might be involved in lipidation of ArtA substrates.

Upon further *in silico* analyses, we were intrigued that in several *Methanosarcina* species the *artA* and *pssA* genomic region flanked a homolog of the *pssD* gene, probably encoding archaetidylserine decarboxylase, suggesting they could be involved in consecutive steps of a proposed archaetidylethanolamine biosynthetic pathway. However, in *H. volcanii*, the genes encoding ArtA (HVO_0915), HvPssA (HVO_1143) and HvPssD (HVO_0146) are not clustered in the same genomic neighborhood (Fig. 1).
It should be noted that while the *Methanosarcina spp.* and *H. volcanii* PssA and PssD homologs share a significant similarity, these proteins have only 30-35% sequence identity to the *in vitro* characterized enzymes from. *M. thermoautotrophicus*, *S. meliloti*, and *B. subtilis* (15, 17, 18). Thus, it is possible that the *H. volcanii* proteins may act on variants of the substrates processed by these experimentally characterized homologs.

***H. volcanii* Δ*hvpssA* and Δ*hvpssD* cells exhibit growth, morphology, and motility phenotypes similar to those of the Δ*artA* strain**.

In order to determine whether HvPssA and/or HvPssD are involved in the archaeosortase-dependent processing pathway, we generated *H. volcanii hvpssA* and *hvpssD* deletion mutants (Fig. S1), using the pop-in pop-out method (19).

We had previously shown that the *H. volcanii* Δ*artA* strain exhibits various severe phenotypic defects (e.g. poor growth, atypical morphology, impaired motility), perhaps due, at least in part, to defective processing of the SLG, an ArtA substrate (10). The Δ*hvpssA* and Δ*hvpssD* strains exhibit a growth defect similar to that of the Δ*artA* strain, as compared to

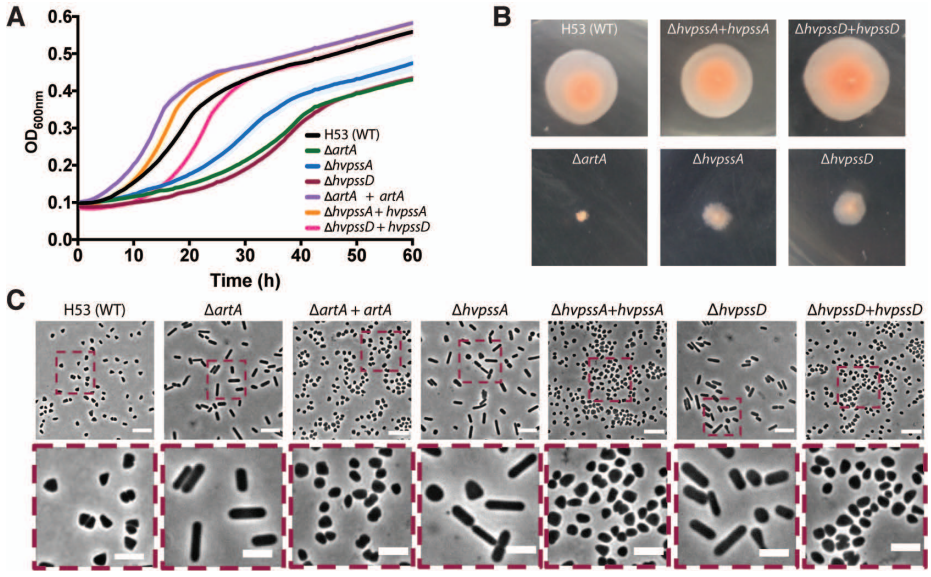


Figure 2: Absence of HvPssA or HvPssD leads to defects in growth, cell morphology, and motil-ity. **(A)** Growth curves: H53 (WT), Δ*artA*, Δ*hvpssA*, and Δ*hvpssD* cells were grown with shaking in 96-well plates with a total volume of 200 µl of liquid semi-defined CA medium and growth of six biological replicates was monitored at OD_{600} with recordings every 30 min. For com-plementation analysis, *artA, hvpssA* or *hvpssD* were expressed from pTA963, under the trypto-phan-inducible P*tnaA* promoter. Both the H53 and deletion strains were transformed with an empty pTA963 plasmid as a control. **(B)** Motility assay: H53, *artA*, *hvpssA*, or *hvpssD* deletion and complementation strains from individual col-ony on solid agar plates were individually stab-inoculated with a toothpick into semi-solid 0.35% agar in CA medium supplemented with trypto-phan, followed by incubation at 45 °C. **(C)** Phase-contrast images were taken from cells during mid-exponential growth phase (OD_{600} 0.3). Scale bars represent 5 µm.

the H53 parent strain used in these studies (Fig. 2A). For all genes, normal growth is rescued by complementation, expressing the deleted gene in trans from a plasmid. Moreover, the Δ*hvpssA* and Δ*hvpssD* strains are partially impaired in motility, the defect, however, being less severe than in Δ*artA*. While no halo is observed after 5 days of Δ*artA* incubation at 45°C, a small halo is formed by the Δ*hvpssA* and Δ*hvpssD* strains (Fig. 2B).

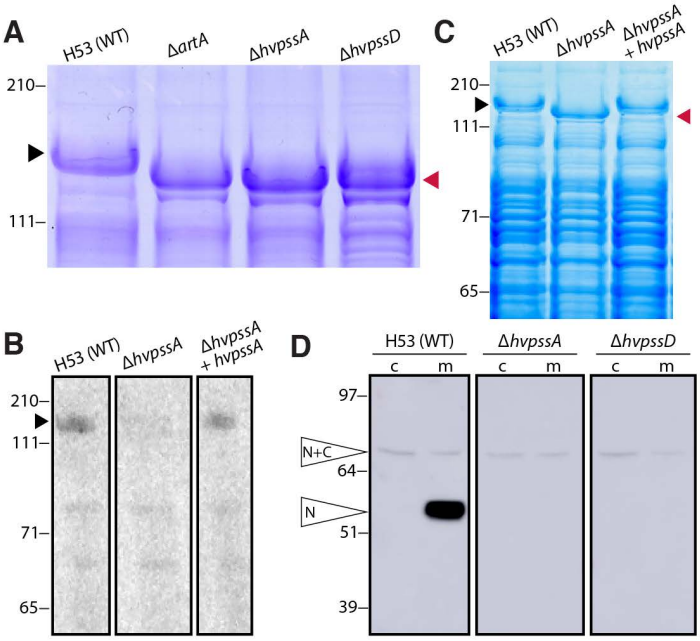


Figure 3: HvPssA and HvPssD are critical for HVO_0405 C-terminal processing and SLG lipidation. **(A)** Coomassie-stained LDS-PAGE gel of cell extracts from *H. volcanii* H53 (wt), Δ*artA*, Δ*hvpssA*, and Δ*hvpssD* strains. The Δ*artA*, Δ*hvpssA* and Δ*hvpssD* SLG (red arrow-head) exhibited a mobility shift compared to the wt SLG (black arrow-head). **(B)** Fluorography of protein extracts isolated from H53 (wt), Δ*hvpssA* and *hvpssA* complementation (Δ*hvpssA* + *hvpssA*) cells grown in the presence of 1 µCi/ml ^{14}C mevalonic acid. Significant la-beling of SLG (black arrowhead) is only detected in the WT and *hvpssA* complementation (Δ*hvpssA* + *hvpssA*) extracts. **(C)** Coo-massie-staining of the gel used for fluorography. The SLG mobility shift in Δ*hvpssA* (red arrowhead) is reverted upon *hvpssA* expression *in trans*. **(D)** Western blot analysis of cytoplasmic (c) and membrane (m) fractions of H53 (wt), Δ*hvpssA*, and Δ*hvpssD* strains expressing, *in trans,* HVO_0405-6xHis. The N-terminal domain of HVO_0405 was detected using anti-HVO_0405-N-term antibodies. HVO_0405 not processed by ArtA and N-terminal HVO_0405 protein processed by ArtA are labeled as "N+C" and "N", respectively. The C-terminal do-main, which carries a His-tag, has not been analyzed in this experi-ment. Numbers indicate molecular mass in kilodaltons.

Light microscopic examination of the parental strain at mid-exponential growth stages shows predominantly disk-shaped cells, while the Δ*artA* cells exhibit a predominantly rod-shaped phenotype (Fig. 2C). Again, the Δ*hvpssA* and Δ*hvpssD* strains show a similar but less severe phenotype. The vast majority of cells from cultures are rods, but we have observed a few disk-shaped cells in liquid cultures for each of these strains. This phenotype is fully complemented and disk-shaped cells are observed when HvPssA or HvPssD are expressed in trans in the Δ*hvpssA* and Δ*hvpssD* strains, respectively (Fig. 2C). Thus, Δ*hvpssA* and Δ*hvpssD* strains exhibit phenotypes similar to the Δ*artA* strain for three independent physiological effects, supporting the hypothesis that the encoded proteins are involved in the ArtA-dependent processing pathway.

HvPssA and HvPssD are required for SLG lipid modification.

To confirm that the drastic cell morphology transitions and the phenotypic similarity between the Δ*hvpssA*/Δ*hvpssD*/Δ*artA* mutants are due to the inhibition of the covalent lipid modification of ArtA substrates, we investigated the lipidation of the SLG mediated by ArtA (11). Initially, as an indirect analysis, we examined the effect of *hvpssA* and *hvpssD* deletions on SLG electrophoretic mobility in an LDS-PAGE gel, as a mobility shift is observed in an *artA* deletion strain (11).

While the similarity of Coomassie-stained band intensities for SLG isolated from the Δ*artA*, Δ*hvpssA* and Δ*hvpssD* strains and their parent strains indicate a similar SLG abundance, electrophoretic mobility demonstrates similar migration shifts of the SLG isolated from all three deletion strains as compared to SLG from the parent strain (Fig. 3A). To corroborate this observation, we set out to directly measure lipid modification of SLG in the Δ*hvpssA* strain.

These experiments proved that lipid labeling of the SLG with radiolabeled mevalonic acid, an archaeal lipid precursor, is severely impaired in the Δ*hvpssA* strain as compared to the parent strain (Fig. 3B) and that this phenotype can be complemented by expressing HvPssA *in trans*.

This is conclusive evidence for HvPssA being closely or even directly coupled with ArtA-dependent protein lipidation. Having run out of this label (and not being able to easily obtain more), we could not perform the equivalent experiment for the Δ*hvpssD* strain. However, because the phenotypes of the Δ*hvpssD* strain closely match those of the Δ*hvpssA* strain in all other assays, including the indirect gel-based assay for SLG modification (Fig. 3C), it is highly likely that HvPssD is also involved in ArtA-dependent SLG lipidation.

HvPssA and HvPssD are required for proteolytic ArtA-substrate processing.

To further support that HvPssA/HvPssD-dependent lipidation is required for ArtA-dependent C-terminal processing, we investigated the C-terminal proteolytic cleavage of a second ArtA substrate. As a reporter, we used the strain-specific domain fusion protein HVO_0405, with its centrally located cleavage site, because the cleaved and uncleaved versions of HVO_0405 can be easily distinguished by LDS-PAGE separation and subsequent immunostaining (12). In contrast, this ArtA substrate is cleaved in the corresponding parent strain as is evident from our Western blot analysis of membrane fractions using anti-HVO_0405 antibodies (Fig. 3D and Fig. S2).

Protein-internal serine residues as an alternative substrate for HvPssA.

Before we became aware that the gene encoding a PssD homolog is located between *artA* and *pssA* in several *Methanosarcina* strains, we considered the possibility that internal Ser residues of substrates might be targeted by HvPssA. This would reflect a restricted shift of substrate specificity from free serine to a protein-internal Ser residue, a shift feasible at only 30% protein sequence identity. This possibility was experimentally addressed. Ser to Ala replacement mutations in HVO_0405 supported this hypothesis (Fig. S3). However, with the detection of HvPssD as an additional player, this hypothesis became less likely, as this probable decarboxylase requires a free carboxyl group, which is lacking in internal Ser residues. However, the observed lack of substrate cleavage for HVO_0405 S463A could indicate that the processing site might be at or close to the mutated Ser, which is only five residues upstream of the PGF motif.

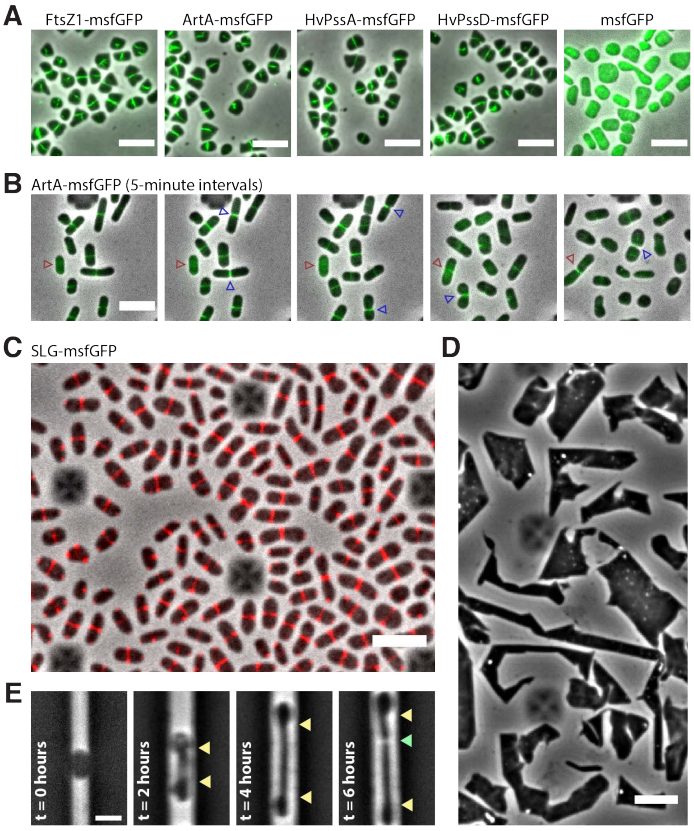


Figure 4: Mid-cell localization of the lipid-anchoring and processing ma-chinery in *H. volcanii*. (A) Snapshots of merged phase contrast (grey) and FITC (green) channels of cells expressing FtsZ1-msfGFP, ArtA-msfGFP, HvPssA-msfGFP, HvPssD-msfGFP and soluble msfGFP. Cells were immobilized under 0.5% agarose pads prepared with CA media. (B) Time lapses of cells growing inside CellAsic microfluidic device. Images of merged phase contrast (grey) and FITC (green) were taken every 5 minutes for 12 hours. Blue arrowheads indicate cell divi-sion events, while red arrowheads label one example of a cell elongat-ing only after the arrival of ArtA-msfGFP to mid cell. (C) Snapshot of SLG-msfGFP (red) mid-cell localization. (D) Phase-contrast images of *H. volcanii* cells under prolonged overexpression of the SLG-msfGFP fusion. (E) *H. volcanii* cells reshape and elongate preferentially at the mid cell during protoplasting recovery. Cells were loaded into the microfluidic chamber, S-layer was chemically removed by addition of 1 mg/mL Proteinase K and recovered with fresh media (t=0 hours). Yel-low arrowheads indicate cell area extended until cell division (green arrowhead). Scale bars represent 5 µm.

Mid-cell localization of ArtA, HvPssA, and HvPssD promote cell site specific lipidation *in vivo*.

Given the dependence of ArtA activity on HvPssA and HvPssD, we considered that ArtA activation required the direct or indirect interaction with HvPssA/HvPssD and/or the product of HvPssA/HvPssD catalysis. To test these hypotheses, we investigated the localization of ArtA, HvPssA and HvPssD in *H. volcanii* live cells. Strikingly, ArtA, HvPssA, and HvPssD tagged with msfGFP localize at mid cell (Fig. 4A). As controls, we also imaged a tagged version of FtsZ1, that has previously been shown to localize to mid cell and was speculated to participate in cell division (20), as well as msfGFP not fused to any protein. FtsZ1-msfGFP shows an almost identical localization pattern compared to ArtA, HvPssA and HvPssD, while the msfGFP protein by itself is not recruited to mid cell (Fig. 4A). Furthermore, time-lapses of cells cultivated within microfluidics suggest that ArtA/HvPssA/HvPssD proteins are recruited to mid cell right after daughter cells are born, and persist for most of the cell cycle, including during cytokinesis (Fig. 4B, blue arrowhead) (Movie S1).

Considering that deletions of *artA*, *hvpssA* and *hvpssD* each drastically perturbed growth (Fig. 2A), induced cells to stay in rod-like shape (Fig. 2C), and do not seem to play an essential role in cell division, we hypothesize lipid-anchoring of ArtA substrates specifically at mid cell might be important for cell elongation and morphogenesis. Interestingly, we also noticed a correlation between the presence of ArtA/HvPssA/HvPssD at the mid cell and actual cell elongation in the cell population (Fig. 4B, red arrowhead). To investigate this further, we expressed a second copy of the SLG tagged with msfGFP, a GFP variant shown to be fluorescent upon Sec-dependent transport in bacteria (21). Interestingly, the SLG-msfGFP fusion accumulated at the mid cell site instead of localizing around the cell envelope (Fig. 4C, Movie S2). Additionally, overexpression of SLG-msfGFP fusion caused severe growth and morphological defects (Fig. 4D). These results suggest that the SLG-msfGFP fusion could be not properly secreted but still interact with the Sec system, thus blocking the secretion of untagged SLG and other Sec substrates; alternatively, secreted SLG-msfGFP may be interfering with the 2D proteinaceous crystal array. However, independent of whether this SLG construct was secreted or not, our results strongly suggest that nascent SLG is targeted specifically to the mid cell.

Lastly, we investigated the morphological transitions in *H. volcanii* protoplast cells during *de novo* S-layer synthesis. If the mid-cell confined ArtA/HvPssA/HvPssD are in fact promoting lipidation of recently secreted SLG molecules at the cell surface, then one would be able to observe mid-cell localized reshaping during protoplast recovery. As expected, protoplasts generated by the addition of Proteinase K either within microfluidics or in bulk cultures assume a round-like shape (Fig. 4E, left panel), suggesting the S-layer might be the structure that ultimately determines the archaeal cell shape. As the protease is washed out and replaced by fresh media, cells rapidly reshape exclusively from mid-cell position (Fig. 4E). Altogether, our observations suggest that the mid-cell area in *H. volcanii* cells is not only dedicated to cell division, but also a central hub for outbound cell extension and other cellular processes.

Discussion

Our data confirmed the hypothesized involvement of the lipid biosynthesis enzyme homologs HvPssA and HvPssD in the C-terminal post-translational modifications of ArtA substrates. With respect to physiological effects, Δ*hvpssA* and Δ*hvpssD* showed similar but slightly less severe phenotypes than those resulting from the deletion of *artA* (Fig. 2). We furthermore experimentally determined the effects of *hvpssA* and *hvpssD* deletions on proteolytic cleavage and lipid labeling (Fig. 3).

Since the gene encoding HVO_0405 resulted from the fusion of two previously independent genes, this protein provided us with an excellent tool for the analysis of ArtA-related proteolysis in *H. volcanii*. Not only the mature protein but also the released protein fragment is large and stable enough to be detected by immunostaining (12). This allowed us to clearly demonstrate that ArtA-dependent proteolytic cleavage is blocked when either *hvpssA* or *hvpssD* are deleted in *H. volcanii*. This block could be bypassed by plasmid-based gene complementation. These results strongly suggest that lipidation and proteolysis are intricately connected with proteolytic cleavage only occurring if the modifying lipid and/or HvPssA/HvPssD is present. By sequence homology, *H. volcanii* HvPssA

is predicted to be involved in lipid biosynthesis, specifically, the generation of the polar lipid archaetidylserine from CDP-archaeol. However, the functionally characterized homolog (from *M. thermautotrophicus*) is only distantly related (30%-35% sequence identity). We considered that HvPssA might have a related, but distinct, function, with HvPssA acting on a Ser residue within a protein rather than on free serine. Ser to Ala replacement mutations in HVO_0405 (Fig. S3) supported this hypothesis. However, with the confirmed involvement of HvPssD, which requires the availability of a free carboxyl group, this hypothesis became unlikely. Instead, in combination, HvPssA and HvPssD might be generating archaetidylethanolamine.

Interestingly, a recent characterization of a bacterial rhombosortase, a non-homologous analog of archaeosortase (22), showed a direct involvement of glycerophosphoethanolamine-containing moiety in the process. Analogously, lipid-attached ethanolamine may be directly involved in the membrane anchoring of ArtA-substrates. In this scenario, instead of being directly involved in the ArtA-mediated substrate cleavage and/or lipid anchoring, HvPssA and HvPssD catalyze the final steps in a pathway that generates archaetidylethanolamine, a substrate required by this process. This opens the way for another hypothesis regarding the ArtA reaction mechanism. In this scenario, ArtA acts similar to sortase A in bacteria: ArtA cleaves the substrate through thioesterification, forming a thioester acyl- enzyme intermediate, which is consistent with the identification of Cys-173 as an active site residue (13). The nucleophilic attack of an amine resolves this intermediate but instead of a pentaglycine branched lipid II, the reactive amine nucleophile is archaetidylethanolamine (Fig. 5A). Such a mechanism would directly result in a covalently modified protein C-terminus. While archaetidylethanolamine lipid was reported to be absent from *Halobacterium salinarum* or *Haloarcula marismortui* (23), it is present in *H. volcanii* and several other haloarchaea albeit with varying abundance (24).

As lipid analysis does not cope with protein-bound lipids, the low concentration of archaetidylethanolamine is not surprising, even though the SLG is highly abundant and archaetidylethanolamine may be used as its membrane anchor. Nevertheless, a detectable amount of archaetidylethanolamine in *H. volcanii* membrane suggest the functional roles of HvPssA and HvPssD: catalyzing the synthesis of archaetidylserine and its decarboxylation to archaetidylethanolamine, respectively. These enzymes perhaps associate with or even form a complex with ArtA, resulting in a majority of the synthesized archaetidylethanolamine to be immediately used to modify the SLG and other ArtA substrates for their membrane anchoring upon C-terminal processing. Thus, only a small amount may be left free in the membrane. The lipid attached to an EDTA-soluble fraction of the SLG has been analyzed by mass spectrometry and was identified as archaetidic acid (25). However, as the lipid has been released from the protein by alkaline hydrolysis, this procedure may have hydrolyzed and thus removed the ethanolamine headgroup.

While investigating the interdependence between ArtA and HvPssA/HvPssD, we observed the recruitment of these proteins to the mid cell in *H. volcanii* (Fig. 4A). Considering these data and the observed mid-cell localization of the SLG (Fig. 4C), we propose a model for S-layer assembly, lipidation and growth in haloarchaea (Fig. 5B). First, SLG is recruited to mid cell, where it is transported across the cytoplasmic membrane in a Sec-dependent manner (26). Following secretion, SLG is processed and linked to archaetidylethanolamine by ArtA, requiring HvPssA/HvPssD for archaetidylethanolamine synthesis and/or interaction with ArtA.

There are still key aspects of haloarchaeal growth and shape control that are not addressed by our model. For example, it is still not clear how the deletion of either *artA*, *hvpssA*, or *hvpssD* generates a rod-shaped cell

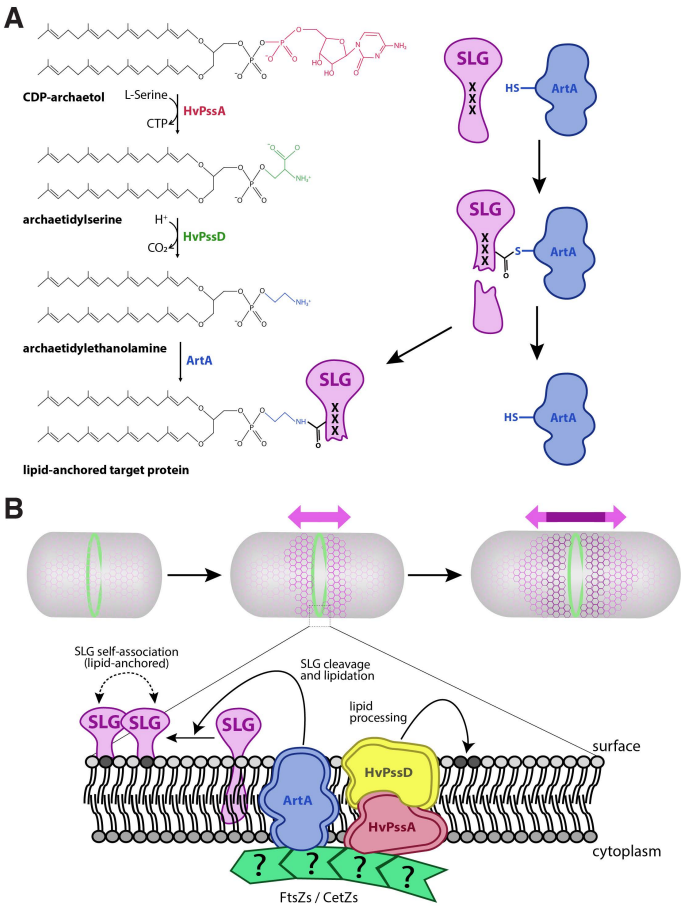


Figure 5: A model for lipid attachment and cell growth involving HvPssA, HvPssD, and ArtA. **(A)** In our speculative model, CDP-ar-chaeol is converted to archaetidylethanolamine in two steps involving HvPssA and HvPssD. ArtA acts as a peptidase and covalently links its active site cysteine to a newly generated C-terminus of its target protein, simultaneously releasing the C-terminal peptide. Then, the free amino group of ethanolamine attacks the thiocarboxylate, which marks the covalent attachment of the target protein to the ArtA active site cysteine. This results in covalent attachment of the lipid to the C-terminus of the target protein as a carboxamide, simultaneously re-leasing ArtA. The process of cleavage and lipidation is dependent on HvPssA and HvPssD, either by binding of archaetidylethanolamine to ArtA or by protein-protein interaction between ArtA and HvPssA/HvPssD. **(B)** Recruitment of ArtA/HvPssA/HvPssD to mid cell promotes anchoring of surface proteins and insertion of new SLG into the S-layer at mid cell, contributing to cell elongation and division.

population (Fig. 2C). Furthermore, although it has been shown that the S-layer is not essential in other archaea (27, 28), overexpression of our SLG-msfGFP fusion drastically impacted the morphology of *H. volcanii* cells (Fig. 4D) beyond the lack of SLG processing and lipidation (11).

Therefore, it is possible that our SLG-msfGFP fusion is actually blocking the transport of or interaction with other yet unknown surface proteins essential for shape maintenance. The concept of having different classes of surface-modifying proteins counteracting each other has been demonstrated in bacteria, where different classes of Penicillin Binding Proteins (PBPs) act on the peptidoglycan cell wall to control cell width homeostasis in rod-shaped cells (29).

Interestingly, just like ovococcoid bacteria that are capable of mid-cell elongation and also lack a clear dedicated elongation machinery, haloarchaea may be also employing cytoskeletal polymers to direct different sub-complexes for cell elongation and cell division (30). This evidence is even more striking for coccoid bacteria, for which a single point mutation in FtsZ is able to induce cell elongation in *Staphylococcus aureus* (31). However, haloarchaea might have conserved at least two distinct elongation modes in addition to cell division, generating disk-like and rod-like populations (32). This scenario would also corroborate the morphological malleability of haloarchaea, being capable of assuming unusual shapes like triangles and squares (33, 34).

In spite of the bacteria that use the localization of specialized proteins to promote mid-cell elongation, it is important to point out that these mechanisms are likely not evolutionary related to the proposed haloarchaeal S-layer lipidation and cell elongation. First, even though there are examples of bacterial species with S-layer that carry out peptidoglycan cell-wall synthesis at mid cell (6), their new S-layer material is inserted as patches distributed all around the cell (35, 36). Second, lack of conservation in the protein architecture between archaeal and bacterial S-layers argue that they may have emerged independently of each other (8, 37).

In conclusion, by applying a set of different experimental approaches, we have confirmed that two putative lipid biosynthesis enzymes, HvPssA and HvPssD, are involved in the proteolytic cleavage and lipid labeling of ArtA substrates specifically at mid cell. We have also proposed, to the best of our knowledge, the first molecular model for archaeal cell elongation.

Materials and Methods

Strains and growth conditions. The plasmids and strains used in this study are listed in Table S1. *H. volcanii* strain H53 and its derivatives were grown at 45°C in semi-defined casamino acid (CA) medium supplemented with tryptophan (50 µg ml^{-1} final concentration) (38). Cells were cultivated either in liquid medium (orbital shaker at 250 rpm) or on solid 1.5% agar. Difco agar and Bacto yeast extract were purchased from Becton, Dickinson, and Company. Peptone was purchased from Oxoid. To ensure equal agar concentrations in all plates, agar was completely dissolved in the media prior to autoclaving, and autoclaved media were stirred before plates were poured. *Escherichia coli* strains were grown at 37°C in NZCYM medium (Fisher Scientific) supplemented with ampicillin (100 µg/ml).

Plasmid preparation and *H. volcanii* transformation. DNA polymerase, DNA ligase, and restriction enzymes were purchased from New England BioLabs. Plasmids were initially transformed into *E. coli* DH5α cells. Plasmid preparations were performed using the QIAprep® Spin Miniprep (Qiagen) kits. Prior to *H. volcanii* transformation, plasmids were transformed into the *dam*$^{-}$ *E. coli* strain DL739. *H. volcanii* transformations were performed using the polyethylene glycol (PEG) method (38). All primers used to construct the recombinant plasmids are listed in Table S2.

Generation of chromosomal *hvpssA* and *hvpssD* deletions in H53. Chromosomal deletions were generated by homologous recombination (pop-in pop-out) as previously described (19). Plasmid constructs for use in the pop-in pop-out knockout process were generated by using overlap PCR as described previously (39) as it follows: approximately 700 nucleotides flanking the *hvpssA* gene were PCR amplified and cloned into the haloarchaeal suicide vector pTA131. The *hvpssA* upstream flanking region is amplified with primers FW_pssA_KO_XbaI and RV_pssA_up while the *hvpssA* downstream flanking region is amplified using FW_pssA_dw and RV_pssA_KO_XhoI (primers are listed in Table S2). The *hvpssA* upstream and downstream flanking DNA fragments were fused by PCR using primers FW_pssA_KO_XbaI and RV_pssA_KO_XhoI, followed by cloning into pTA131 digested with *Xba*I and *Xho*I. The insertion of the correct DNA fragment into the cloning site of the recombinant plasmid was verified by sequencing using the same primers. The final plasmid construct, pFH38, contained upstream and downstream *hvpssA* flanking regions and was transformed into the parental H53 *H. volcanii* strain. To confirm the chromosomal replacement event at the proper location on the chromosome, colonies derived from these techniques were screened by PCR using the FW_pssA_KO_XbaI and RV_pssA_KO_XhoI primers. The *hvpssA* deletion mutant generated in strain H53 was designated FH38 (Table S1). For the generation of plasmid construct for chromosomal *hvpssD* deletion, approximately 700 nucleotides flanking the *hvpssD* gene were PCR amplified and cloned into the haloarchaeal suicide vector pTA131. The upstream flanking region was amplified with primers FW_pssD_KO_XbaI and RV_pssD_up while the downstream flanking region was amplified using FW_pssD_dw and RV_pssD_KO_XhoI. The flanking DNA fragments were fused by PCR using primers FW_pssD_KO_XbaI and RV_pssD_KO_XhoI, followed by cloning into pTA131 digested with *Xba*I and *Xho*I. The insertion of the correct DNA fragment into the cloning site of the recombinant plasmid was verified by sequencing using the same primers. The final plasmid construct, pFH43, contained upstream and downstream *hvpssD* flanking regions and was transformed into the parental *H. volcanii* strain H53. Confirmation of *hvpssD* deletion on the chromosome was screened by PCR using the FW_pssD_KO_XbaI and RV_pssD_KO_XhoI primers. The *hvpssD* deletion mutant generated in strain H53 was designated FH63 (Table S1).

Construction of expression plasmids for HvPssA and HvPssD. To construct a tryptophan-inducible *H. volcanii hvpssA* gene with C-terminal His-tag, its coding region was amplified by PCR using primers FW_pssA_OE_NdeI and RV_pssA_OE_EcoRI_His (Table S2). Meanwhile, for the construction of *H. volcanii hvpssD* gene with C-terminal His-tag, its coding region was amplified by PCR using primers FW_pssD_OE_NdeI and RV_pssD_OE_EcoRI_His (Table S2). The PCR product was cloned into the expression vector pTA963 that had been digested with *Nde*I and *Eco*RI. This places the *hvpssA* or *hvpssD* gene under the control of the inducible tryptophanase promoter (p.*tna*). The recombinant pTA963 carrying the *hvpssA* gene was designated pFH39 and the pTA963 carrying *hvpssD* was designated pFH44. To complement the Δ*hvpssA* strain FH38, this strain was transformed with plasmid pFH39 to result in FH56. For complementation of Δ*hvpssD* strain FH63, this strain was transformed with plasmid pFH44 to result in FH70. The H53, Δ*hvpssA*, and Δ*hvpssD* strains were also transformed with the empty expression vector pTA963 which was used as a control.

Construction of expression plasmid for *csg*-msfGFP(SW) and msfGFP. To construct a tryptophan-inducible *H. volcanii csg* tagged with msfGFP, a sandwich fusion was created by intercalating msfGFP gene product amplified from synthetic fragment (oligos oHV81 and oHV82) in between the first 102bp (the secretion signal sequence of SLG, oligos oAB500 and oAB501) and the rest of the *csg* coding region (oligos oHV83 and oAB502). The 3 fragments were then assembled by Gibson assembly (40) together with the pTA962 plasmid (41), digested with NheI and BamHI and transformed into DH5α cells. The same process was followed for the cloning of msfGFP alone (oligos oHV101 and oHM68). The clones were subjected to validation by PCR and Sanger sequencing (Table S1).

Generation of chromosomal msGFP fusions in H26. C-terminal translational fusion constructs, including a pyrE2 cassette for selection, were generated by direct transformation of PCR fragment ensembles generated

by Gibson assembly (40), transformed directly in *H. volcanii* H26 cells and selected by growth in the absence of uracil. PCR fragments from ftsZ*1* (oligos oHV3 and oHV4 for upstream region and oHV8 and oHV9 for downstream region), *artA* (oligos oHM91 and oHM92 for upstream region and oHM93 and oHM94 for downstream region) , *hvpssA* (oligos oHV156 and oHV157 for upstream region and oHV158 and oHV159 for downstream region), and *hvpssD* (oligos oHV126 and oHV127 for upstream region and oHV128 and oHV129 for downstream region) were assembled to msfGFP (oligos oHM34 and oHM6) and the pyrE2 cassette (oligos oHV6 and oHV7). Chromosomal replacements were confirmed by PCR and Sanger sequencing (Table S1).

Immunoblotting. Liquid cultures were grown until mid-log phase (OD_{600} 0.2-0.5) and the cells were harvested by centrifugation at 3,800 x g for 5 min at room temperature. Cell pellets were resuspended and lysed in 1% (v/v) NuPAGE lithium dodecyl sulfate (LDS) sample buffer supplemented with 100 mM dithiothreitol (DTT) and stored at -20°C. Samples were electrophoresed on 4-12% Bis-Tris polyacrylamide gels (Invitrogen) with NuPAGE 3-(N-morpholino) propane sulfonic acid (MOPS) sodium dodecyl sulfate (SDS) running buffer (Invitrogen). Proteins were then transferred to polyvinylidene difluoride (PVDF) membranes (Millipore) using a semi-dry transfer apparatus at 15 V for 30 minutes (Bio-Rad). Subsequently, the membrane was washed twice in PBS, blocked for one hour in 3% bovine serum albumin (BSA) in PBS, and washed twice in PBS with 1% Tween-20 and once with PBS. For detection of the poly His-tag, the mouse anti-penta-His antibody (Qiagen; Catalog #34660) was used at a 1:2,000 dilution in 3% BSA in PBS with sodium azide. For the secondary antibody, HRP-conjugated Amersham ECL anti-mouse IgG from sheep (GE) was used at a 1:20,000 dilution in 10% nonfat milk in PBS. For detection of HVO_0405_Nterm (see Suppl Text S1 and associated methods), the rabbit anti-HVO_0405-N-term serum (12) was used as the primary antibody at a 1:10,000 dilution in 3% BSA in PBS with sodium azide. For the secondary antibody, HRP-conjugated Amersham ECL anti-rabbit IgG from donkey (GE) was used at a 1:60,000 dilution in 10% nonfat milk in PBS.

Lipid radiolabeling. The H53 parent strain carrying the vector control pTA963, and the Δ*hvpssA* deletion strain carrying either the *hvpssA* expression plasmid pFH39 or the vector control pTA963 were grown in 5 ml liquid CA medium. Upon reaching mid-log phase (OD_{600} ~0.5), 20 µl of each culture was transferred into 1 mL of fresh liquid CA medium supplemented with [^{14}C] mevalonic acid (resuspended in ethanol) at a final concentration of 1 µCi/ml. The cultures were harvested after reaching mid-log phase and proteins were precipitated from 1 ml cultures with 10% TCA, followed by a delipidation step to remove non-covalently linked lipid as described previously (42, 43). The delipidated proteins were separated by 7% Tris-Acetate (TA) LDS-PAGE. For analysis of the samples, the gel was dried onto blotting paper using a Gel Dryer (Bio-Rad model 583), exposed to a phosphor screen (Molecular Dynamics) for 3 weeks, and analyzed using a Typhoon imager (Amersham Biosciences).

Motility assays. The motility assays of *H. volcanii* H53 (parent), Δ*artA*, Δ*hvpssA*, and Δ*hvpssD* strains carrying the plasmid expressing the complementary gene (or pTA963 as control) were performed on 0.35% agar in CA medium supplemented with tryptophan as described previously (39). A toothpick was used to stab inoculate the agar, followed by incubation at 45°C. Halo sizes around the stab-inoculation site were measured after 3–5 days of incubation.

Growth curves. Growth curves were measured using a Biotek PowerWaveX2 microplate spectrophotometer. *H. volcanii* H53 (parent), Δ*artA*, Δ*hvpssA*, and Δ*hvpssD* strains carrying the plasmid expressing the complementary gene (or pTA963 as control) were first incubated in 5 ml liquid cultures in CA medium supplemented with tryptophan with continuous shaking at 45°C, until suitable OD_{600} values (0.2–0.5) were reached. Approximately 6 µl of each culture (adjusted to correct for OD_{600} differences) were then transferred into 194 µl of fresh CA medium supplemented with tryptophan (50 µg ml^{-1} final concentration) and grown to stationary phase, with OD_{600} recordings taken every 30 min.

Light microscopy. The *H. volcanii* strains H53 (parent), Δ*artA*, Δ*hvpssA*, and Δ*hvpssD* carrying the plasmid expressing the complementary gene (or pTA963 as control) were inoculated from colony to 5 ml CA liquid medium and grown until they reached mid-log phase (OD_{600} ~0.4-0.5). Serial liquid to liquid sub-inoculations were carried out by transferring 10 µl of the liquid culture to 5 ml fresh liquid CA medium with up to two transfers. Subsequently, 1 ml of each culture was concentrated by centrifugation at 4,911 × g for 1 min, pellets resuspended in 10 µl of liquid CA medium. Then, 10 µl of the concentrated cells was transferred to under a 0.5% agarose pad with CA medium and observed using a Nikon Eclipse TiE inverted TIRF microscope. ArtA-msfGFP time lapses were acquired by culturing *H. volcanii* cells inside Millipore ONIX CellAsic microfluidic plates as previously described (33). Images were taken under 5-minute intervals for 12 hours in both phase contrast and with 488 nm laser channels. *H. volcanii* protoplasts were generated within microfluidic channels by the addition of 1 mg/mL of Proteinase K (Invitrogen) in YPC medium until cells lost shape. Subsequently, the cells were washed with fresh YPC and time lapse was recorded with 10-minute intervals for 12 hours.

Acknowledgments

We want to thank Howard Goldfine, the Pohlschroder, Daldal and Garner labs for helpful discussions. We also want to thank Henry Miziorko, Ethan Garner and Jenny Zheng for access to equipment and reagents. MP and MAH were supported by the National Science Foundation grant 1817518. SS was supported by the German Science Foundation Postdoctoral Fellowship.

References

1. Bisson-Filho AW, Zheng J, Garner E. 2018. Archaeal imaging: leading the hunt for new discoveries. Mol Biol Cell 29:1675–1681.
2. de Pedro MA, Quintela JC, Höltje J V, Schwarz H. 1997. Murein segregation in *Escherichia coli*. J Bacteriol 179:2823–34.
3. May JW, Mitchison JM. 1986. Length growth in fission yeast cells measured by two novel techniques. Nature 322:752–754.
4. Brown PJB, de Pedro MA, Kysela DT, Van der Henst C, Kim J, De Bolle X, Fuqua C, Brun Y V. 2012. Polar growth in the Alphaproteobacterial order Rhizobiales. Proc Natl Acad Sci 109:1697 LP – 1701.
5. Sipiczki M, Takeo K, Grallert A. 1998. Growth polarity transitions in a dimorphic fission yeast. Microbiology 144:3475–3485.
6. Aaron M, Charbon G, Lam H, Schwarz H, Vollmer W, Jacobs-Wagner C. 2007. The tubulin homologue FtsZ contributes to cell elongation by guiding cell wall precursor synthesis in *Caulobacter crescentus*. Mol Microbiol 64:938–952.
7. Rodrigues-Oliveira T, Belmok A, Vasconcellos D, Schuster B, Kyaw CM. 2017. Archaeal S-layers: Overview and current state of the art. Front Microbiol 8:1–17.
8. Pohlschroder M, Pfeiffer F, Schulze S, Halim MFA. 2018. Archaeal cell surface biogenesis. FEMS Microbiol Rev 42:694–717.
9. Eun Y-J, Ho P-Y, Kim M, LaRussa S, Robert L, Renner LD, Schmid A, Garner E, Amir A. 2018. Archaeal cells share common size control with bacteria despite noisier growth and division. Nat Microbiol 3:148–154.

10. Abdul Halim MF, Pfeiffer F, Zou J, Frisch A, Haft D, Wu S, Tolić N, Brewer H, Payne SH, Paša-Tolić L, Pohlschroder M. 2013. *Haloferax volcanii* archaeosortase is required for motility, mating, and C-terminal processing of the S-layer glycoprotein. Mol Microbiol 88:1164–1175.
11. Abdul Halim MF, Karch KR, Zhou Y, Haft DH, Garcia BA, Pohlschroder M. 2016. Permuting the PGF signature motif blocks both archaeosortase-dependent C-terminal cleavage and prenyl lipid attachment for the *Haloferax volcanii* S-layer glycoprotein. J Bacteriol 198:808–815.
12. Abdul Halim MF, Stoltzfus JD, Schulze S, Hippler M, Pohlschroder M. 2017. ArtA-Dependent processing of a Tat substrate containing a conserved tripartite structure that Is not localized at the C-terminus. J Bacteriol 199:e00802-16.
13. Abdul Halim MF, Rodriguez R, Stoltzfus JD, Duggin IG, Pohlschroder M. 2018. Conserved residues are critical for *Haloferax volcanii* archaeosortase catalytic activity: Implications for convergent evolution of the catalytic mechanisms of non-homologous sortases from archaea and bacteria. Mol Microbiol 108:276–287.
14. Haft DH, Payne SH, Selengut JD. 2012. Archaeosortases and Exosortases Are Widely Distributed Systems Linking Membrane Transit with Posttranslational Modification. J Bacteriol 194:36 LP – 48.
15. Morii H, Koga Y. 2003. CDP-2,3-di-O-geranylgeranyl-sn-glycerol:L-serine O-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon *Methanothermobacter thermautotrophicus*. J Bacteriol 185:1181–1189.
16. Dutt A, Dowhan W. 1981. Characterization of a membrane-associated cytidine diphosphate-diacylglycerol-dependent phosphatidylserine synthase in bacilli. J Bacteriol 147:535–42.
17. Okada M, Matsuzaki H, Shibuya I, Matsumoto K. 1994. Cloning, sequencing, and expression in *Escherichia coli* of the *Bacillus subtilis* gene for phosphatidylserine synthase. J Bacteriol 176:7456–61.
18. Nishibori A, Kusaka J, Hara H, Umeda M, Matsumoto K. 2005. Phosphatidylethanolamine domains and localization of phospholipid synthases in *Bacillus subtilis* membranes. J Bacteriol 187:2163–2174.
19. Allers T, Ngo H-P, Mevarech M, Lloyd RG. 2004. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. Appl Environ Microbiol 70:943–53.
20. Duggin IG, Aylett CHS, Walsh JC, Michie K a., Wang Q, Turnbull L, Dawson EM, Harry EJ, Whitchurch CB, Amos L a., Löwe J. 2015. CetZ tubulin-like proteins control archaeal cell shape. Nature 519:362–5.
21. Dinh T, Bernhardt TG. 2011. Using superfolder green fluorescent protein for periplasmic protein localization studies. J Bacteriol2011/07/15. 193:4984–4987.
22. Gadwal S, Johnson TL, Remmer H, Sandkvist M. 2018. C-terminal processing of GlyGly-CTERM containing proteins by rhombosortase in Vibrio cholerae. PLOS Pathog 14:e1007341.
23. Daiyasu H, Kuma K-I, Yokoi T, Morii H, Koga Y, Toh H. 2005. A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition. Archaea 1:399–410.
24. Kellermann MY, Yoshinaga MY, Valentine RC, Wörmer L, Valentine DL. 2016. Important roles for membrane lipids in haloarchaeal bioenergetics. Biochim Biophys Acta 1858:2940–2956.
25. Kandiba L, Guan Z, Eichler J. 2013. Lipid modification gives rise to two distinct *Haloferax volcanii* S-layer glycoprotein populations. Biochim Biophys Acta 1828:938–43.
26. Irihimovitch V, Eichler J. 2003. Post-translational secretion of fusion proteins in the halophilic archaea *Haloferax volcanii*. J Biol Chem 278:12881–7.
27. Zhang C, Phillips APR, Wipfler RL, Olsen GJ, Whitaker RJ. 2018. The essential genome of the crenarchaeal model *Sulfolobus islandicus*. Nat Commun 9:4908.
28. Zhang C, Wipfler RL, Li Y, Wang Z, Hallett EN, Whitaker RJ. 2019. Cell structure changes in the hyperthermophilic crenarchaeon *Sulfolobus islandicus* Lacking the S-Layer. MBio 10:e01589-19.
29. Dion MF, Kapoor M, Sun Y, Wilson S, Ryan J, Vigouroux A, van Teeffelen S, Oldenbourg R, Garner EC. 2019. *Bacillus subtilis* cell diameter is determined by the opposing actions of two distinct cell wall synthetic systems. Nat Microbiol 4:1294–1305.
30. Pérez-Núñez D, Briandet R, David B, Gautier C, Renault P, Hallet B, Hols P, Carballido-López R, Guédon E. 2011. A new morphogenesis pathway in bacteria: unbalanced activity of cell wall synthesis machineries leads to coccus-to-rod transition and filamentation in ovococci. Mol Microbiol 79:759–771.
31. Pereira AR, Hsin J, Król E, Tavares AC, Flores P, Hoiczyk E, Ng N, Dajkovic A, Brun Y V, VanNieuwenhze MS, Roemer T, Carballido-Lopez R, Scheffers D-J, Huang KC, Pinho MG. 2016. FtsZ-dependent elongation of a coccoid bacterium. MBio 7:e00908-16.
32. Li Z, Kinosita Y, Rodriguez-Franco M, Nußbaum P, Braun F, Delpech F, Quax TEF, Albers S-V. 2019. Positioning of the motility machinery in halophilic archaea. MBio 10.
33. Walsh JC, Angstmann CN, Bisson-Filho AW, Garner EC, Duggin IG, Curmi PMG. 2019. Division plane placement in pleomorphic archaea is dynamically coupled to cell shape. Mol Microbiol.
34. Walsby AE. 1980. A square bacterium. Nature 283:69–71.
35. Comerci CJ, Herrmann J, Yoon J, Jabbarpour F, Zhou X, Nomellini JF, Smit J, Shapiro L, Wakatsuki S, Moerner WE. 2019. Topologically-guided continuous protein crystallization controls bacterial surface layer self-assembly. Nat Commun 10:2731.
36. Oatley P, Kirk JA, Ma S, Jones S, Fagan RP. 2018. Spatial organization of *Clostridium difficile* S-layer biogenesis. bioRxiv 405993.
37. Bharat TAM, Kureisaite-Ciziene D, Hardy GG, Yu EW, Devant JM, Hagen WJH, Brun Y V, Briggs JAG, Löwe J. 2017. Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. Nat Microbiol 2:17059.
38. Dyall-Smith M. 2009. The Halohandbook: Protocols for halobacterial genetics Ver 7.2.
39. Tripepi M, Imam S, Pohlschroder M. 2010. *Haloferax volcanii* flagella are required for motility but are not Involved in PibD-dependent surface adhesion. J Bacteriol 192:3093–3102.
40. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–5.
41. Allers T, Barak S, Liddell S, Wardell K, Mevarech M. 2010. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. Appl Environ Microbiol2010/01/22. 76:1759–1769.
42. Kikuchi A, Sagami H, Ogura K. 1999. Evidence for Covalent Attachment of Diphytanylglyceryl Phosphate to the cell-surface glycoprotein of *Halobacterium halobium*. J Biol Chem 274:18011–18016.
43. Konrad Z, Eichler J. 2002. Protein glycosylation in *Haloferax volcanii*: partial characterization of a 98-kDa glycoprotein. FEMS Microbiol Lett 209:197–202.

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 2626–2631, March 1997
Microbiology

A genetic system for Archaea of the genus *Methanosarcina*: Liposome-mediated transformation and construction of shuttle vectors

WILLIAM W. METCALF*†, JUN KAI ZHANG*, ETHEL APOLINARIO‡, KEVIN R. SOWERS‡, AND RALPH S. WOLFE*

*Department of Microbiology, B103 Chemical and Life Science Laboratory, 601 South Goodwin Avenue, University of Illinois, Urbana, IL 61801; and ‡Center of Marine Biotechnology, Suite 236, Columbus Center, 701 Pratt Street, University of Maryland Biotechnology Institute, Baltimore, MD 21202

Contributed by Ralph S. Wolfe, December 26, 1996

ABSTRACT **New methods that allow, for the first time, genetic analysis in Archaea of the genus *Methanosarcina* are presented. First, several autonomously replicating plasmid shuttle vectors have been constructed based on the naturally occurring plasmid pC2A from *Methanosarcina acetivorans*. These vectors replicate in 9 of 11 *Methanosarcina* strains tested and in *Escherichia coli*. Second, a highly efficient transformation system based upon introduction of DNA by liposomes has been developed. This method allows transformation frequencies of as high as 2×10^8 transformants per microgram of DNA per 10^9 cells or ≈20% of the recipient population. During the course of this work, the complete 5467-bp DNA sequence of pC2A was determined. The implications of these findings for the future of methanoarchaeal research are also discussed.**

Methanoarchaea play a key role in the global carbon cycle by recycling organic carbon from anaerobic environments into the atmosphere as methane gas. One of the most significant achievements in microbiology in recent years has been the recognition that Archaea, including the methanoarchaea, constitute a truly unique form of life, evolutionarily distinct from both eukaryotes and eubacteria (1, 2). Biochemical study of the methanogenic pathways of methanoarchaea has provided a wealth of information on novel enzyme mechanisms and cofactors (3, 4). However, apart from the methanogenic pathways, very little is known about these organisms. Because there is a dearth of genetic methods available for examination of methanoarchaea, the study of these organisms has been largely confined to biochemical and physiological approaches. Thus, relatively little is known about other aspects of their metabolism, including the pathways by which they synthesize their unique cofactors and other cellular components, as well as how they regulate gene expression in response to changes in their environment.

Development of methods for genetic analysis of methanoarchaea is crucial for increasing our understanding of how they adapt and survive in their limited environments. The recently published sequence of the *Methanococcus jannaschii* genome illustrates this point (5). Many unexpected genes were observed, whereas other, expected genes were absent. Close to 60% of the putative coding regions in the *Methanococcus jannaschii* genome were not significantly similar to anything previously known. Without methods for analyzing the function of these unknown genes and for identifying which of the genes are involved in known pathways, these observations will remain unexplained. Genetic analysis is the most direct method for addressing these questions.

After decades of effort, genetic analysis of methanoarchaea remains fairly primitive relative to the analyses of eubacteria and many eukaryotes. This is primarily due to the lack of tools commonly used for analysis of more tractable organisms. Because of their unique properties, there are few antibiotic, or other, selections that are effective on methanoarchaea (6). Thus, although both plasmids and transposons are known to exist in methanoarchaea, no usable plasmid vectors or selectable transposons have been developed (5, 7, 8). A transducing phage has been identified for *Methanobacterium thermoautotrophicum*, but it is impractical for routine use due to a very low burst size (about six phage per cell; ref. 9). Further, conjugation is not known to occur among the methanoarchaea. Lastly, transformation or transfection with purified DNA has yet to be demonstrated in most methanoarchaea and is usually inefficient in those where it has been shown (10–12).

Despite these drawbacks, some progress has been made in developing methods of genetic analysis applicable to methanoarchaea. The antibiotics puromycin, pseudomonic acid, and neomycin have been shown to be active against some methanoarchaea (13–15). The puromycin-resistance (Pur^R) gene of *Streptomyces alboniger*, *pac*, and the aminoglycoside phosphotransferase genes *aphI* and *aphII* have been modified to allow expression in *Methanococcus* species (15, 16). Both cassettes have subsequently been used in reverse genetic approaches to construct *Methanococcus* mutants (15–17). This method has been significantly improved by development of a highly efficient polyethylene glycol (PEG)-mediated transformation protocol for *Methanococcus maripaludis* (18).

It has long been known that traditional methods of chemical mutagenesis are effective on methanoarchaea. With these methods, a variety of mutants have been isolated in diverse organisms (11, 19–21). Unfortunately, both the reverse genetic and the chemical mutagenesis methods have significant limitations. The former method requires the gene of interest to be identified and cloned before mutagenesis, as well as an efficient transformation method; the latter method does not allow identification of the effected gene once a mutant is obtained. Development of a functional cloning vehicle would greatly facilitate mutant analysis, and, in addition, would allow other experiments, such as identification of gene regulatory elements. Such a vector was recently reported for use in *Methanococcus maripaludis*§. In this paper, we report the construction of a shuttle vector for gene cloning and analysis that is functional in a number of *Methanosarcina* species. Using this vector, we have developed a highly efficient protocol for liposome-mediated transformation of *Methanosarcina* that is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/942626-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

Abbreviations: PEG, polyethylene glycol; Pur^R, puromycin resistance.
Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. U78295).

†To whom reprint requests should be addressed. e-mail: bill_metcalf@qms1.life.uiuc.edu.

§Tumbula, D. L. & Whitman, W. B., Abstracts of the 96th American Society for Microbiology General Meeting, May 19–23, 1996, New Orleans, LA.

applicable to both plasmid transformation and reverse genetic experiments.

MATERIALS AND METHODS

Bacterial Strains, Media, and Growth Conditions. Standard conditions were used for growth of *Escherichia coli* strains (22). DH5α and DH5α/λ*pir* (23) were from S. Maloy (University of Illinois). The former was used for plasmid constructions with pBluescript or pTZ18R vectors, and the latter was used for constructions of *pir*-dependent replicons. *Methanosarcina acetivorans* C2A (DSM 2834), *Methanosarcina barkeri* Fusaro (DSM 804), *Methanosarcina barkeri* MS (DSM 800), *Methanosarcina barkeri* W, *Methanosarcina mazei* C-16 (DSM 3318), *Methanosarcina mazei* S-6 (DSM 2053), *Methanosarcina mazei* LYC (DSM 4556), *Methanosarcina thermophila* TM-1 (DSM 1825), *Methanosarcina siciliae* C2J, *Methanosarcina* spp. WH1 (DSM 4659), and *Methanosarcina* spp. WH2 were from laboratory stocks. *Methanosarcina* strains were grown in single cell morphology (24) at 35°C in HS-methanol-acetate or T_{100}-trimethylamine broth media under strictly anaerobic conditions (25). Plating of *Methanosarcina* strains was essentially as described (24, 26). In some cases HS-methanol-acetate medium solidified by addition of noble agar at either 1% (bottom agar) or 0.5% (top agar) was used. Puromycin was used at 1 μg/ml for *Methanosarcina* spp. All plating manipulations were carried out under strictly anaerobic conditions in an anaerobic glove box.

Plasmids. Plasmid pBluescript KS(+) was from Stratagene, pTZ18R and pSL1180 were from Pharmacia, pMip1 (16) was from J. Konisky (Rice University, Houston), pGP704 (23) was from J. Slauch (University of Illinois), and pXS2 (27) was from our laboratory stock. Plasmid pC2A is a naturally occurring plasmid from *Methanosarcina acetivorans*. Standard methods were used for isolation and manipulation of plasmid DNA from *E. coli* (28). Plasmid DNA was isolated from *Methanosarcina* species by a modification of the standard alkaline lysis method (28) in which the cell pellet was resuspended in either growth medium or 20 mM Tris·HCl (pH 8.0) with 0.85 M sucrose rather than the standard lysis buffer. Plasmid pJK1 was constructed by ligation of a 1773-bp *Eco*RI fragment from pMip1 into *Eco*RI-digested pBluescript KS(+) such that the *pac* gene was in the same orientation as the *lacZ*α peptide. Plasmid pJK3 was constructed by ligation of three fragments: (*i*) *Apa*I- and *Not*I-digested pBluescript KS(+), (*ii*) a 427-bp *Apa*I and *Rca*I-digested PCR fragment amplified from pJK1 using the primers 5′-GCTTGTACTCGGTCATGAGAATCACTCC-3′ and 5′-AGCGGATAACAATTTACACAGG-3′, and (*iii*) a 1001-bp *Rca*I and *Not*I-digested PCR fragment amplified from pJK1 using the primers 5′-GGAGTGATTCTCATGACCGAGTACAAGC-3′ and 5′-GTTTTCCCAGTCACGAC-3′. Plasmid pWM223 was constructed by ligation of *Bgl*II-digested pC2A into the *Bam*HI site of pBluescript KS(+). Plasmids pWM224 and pWM225 were constructed by ligation of *Eco*RI-digested pC2A into the *Eco*RI site of pBluescript KS(+). The two plasmids differ with respect to the orientation of the pC2A insert. In addition to the pC2A insert, pWM225 also carries an 80-bp *Eco*RI fragment of unknown origin. Plasmid pWM241 was constructed by ligation of *Spe*I-digested pC2A into the *Spe*I site of pBluescript KS(+). Plasmid pJK8 was constructed by ligation of the *Bam*HI to *Xho*I *pac* cassette of pJK3 into the same sites in pWM241. Plasmid pJK21 carries the *Kpn*I to *Eco*RV *pac* cassette of pJK8 ligated with *Hin*cII- and *Kpn*I-digested pTZ18R. Plasmid pJK24 was made by self-ligation of a 3523-bp *Asc*I-digested PCR fragment amplified from pJK21 with the primers 5′-CCCGGCGCGCCTCTAGAGGATGATTAATTTTAAG-3′ and 5′-CCCGGCGCGCCAGGTGGCACTTTTCGGGGAAATG-3′. Plasmid pWM303 was made by ligation of a 420-bp *Eco*RI to *Bam*HI fragment of pGP704 to a 2329-bp *Mfe*I- and *Bgl*II-digested PCR fragment amplified from pJK24 with the primers 5′-CCGAGATCTAAAAAAAAGCCCGCTCATTAGGCGGGCTGACAGTTACCAATGCTTAATC-3′ and 5′-CCGCCGCAATTGCCCCCAGTGAATTAAAAATATATAAAAAAAGG-3′. Plasmid pWM307 was constructed by ligation of the 5467-bp *Spe*I fragment from pJK8 with *Xba*I-digested pWM303. Plasmids pWM309, pWM311, pWM313, pWM315, pWM317, pWM319, and pWM321 were constructed by ligation of *Asc*I-digested pWM307 with *Asc*I-digested PCR fragments carrying the *lacZ*α and polylinker regions amplified from pMTL20, pMTL21, pMTL22, pMTL23, pMTL24, pMTL25, and pSL1180, respectively, (29), using the primers 5′-GCCGGCGCGCCTTAACCATTCGCCATTCAGGCTGC-3′ and 5′-GCCGGCGCGCCAATACGCAAACCGCCTCTCC-3′.

DNA Sequencing, Analysis, and Hybridization. The complete DNA sequence of pC2A was determined from double-stranded templates by automated dye terminator sequencing. Standard primers were used to generate junction sequences from pWM223, pWM224, and pWM225. Internal sequences were derived from deletion derivatives of pWM224 and pWM225 constructed using the Exo III/Mung Bean Deletion Kit (Stratagene). The pC2A junction sequences in pWM223, pWM224, and pWM225 were verified by sequencing of pC2A with synthetic oligonucleotide primers. Gaps in the sequence were determined from pWM224 or pWM225 with synthetic oligonucleotide primers. DNA sequencing and oligonucleotide synthesis were performed at the Genetic Engineering Facility (University of Illinois). DNA sequences were compiled and analyzed using the GCG package (Version 8, September 1994, Genetics Computer Group, Madison, WI). Chromosomal DNA isolation and DNA hybridization experiments were performed as described (28).

Transformation. *E. coli* was transformed by electroporation using an *E. coli* Gene Pulser (Bio-Rad) as recommended. Electroporation of *Methanosarcina* in high-resistance medium was essentially identical to the method used for *E. coli* except that 0.85 M sucrose was used as the electroporation buffer. Electroporation of protoplasts, PEG-lithium acetate-mediated, and PEG-mediated transformations were performed essentially as described (12, 18, 30) except that the buffers were made isoosmotic to the growth medium by addition of sucrose. Natural transformation was performed by mixing cells with DNA followed by incubation for various time periods. For liposome-mediated transformation, cells from log-phase cultures (OD_{650} between 0.2 and 0.5) were collected by centrifugation and resuspended in 0.85 M sucrose at a density of $\approx 1 \times 10^9$ cells per milliliter. DNA:liposome complexes were formed by mixing 2–25 μl DOTAP (Boehringer Mannheim) in 100 μl of 20 mM Hepes (pH 7.4) with 2 μg of plasmid DNA in 50 μl of 20 mM Hepes (pH 7.4), followed by a 15-min incubation at room temperature. A 1.0-ml portion of the resuspended cells was added to the DNA:liposome suspension and incubated for 4 hr at room temperature. With these cell and DNA concentrations, the maximum transformation frequency of *Methanosarcina acetivorans* C2A was achieved with 15 μl of DOTAP reagent. For all methods, cells were transferred to 10 ml of broth medium after transformation, incubated at 35°C for 12–16 hr, and then plated on medium with puromycin.

RESULTS

Cloning and Sequence Analysis of pC2A. The complete 5467-bp DNA sequence of the naturally occurring *Methanosarcina acetivorans* plasmid pC2A was determined to provide a rational basis for design of an *Methanosarcina–E. coli* shuttle vector (Fig. 1). The sequence is in agreement with the previously determined restriction map of pC2A (7). Four ORFs of greater than 120 aa were identified, and their putative products

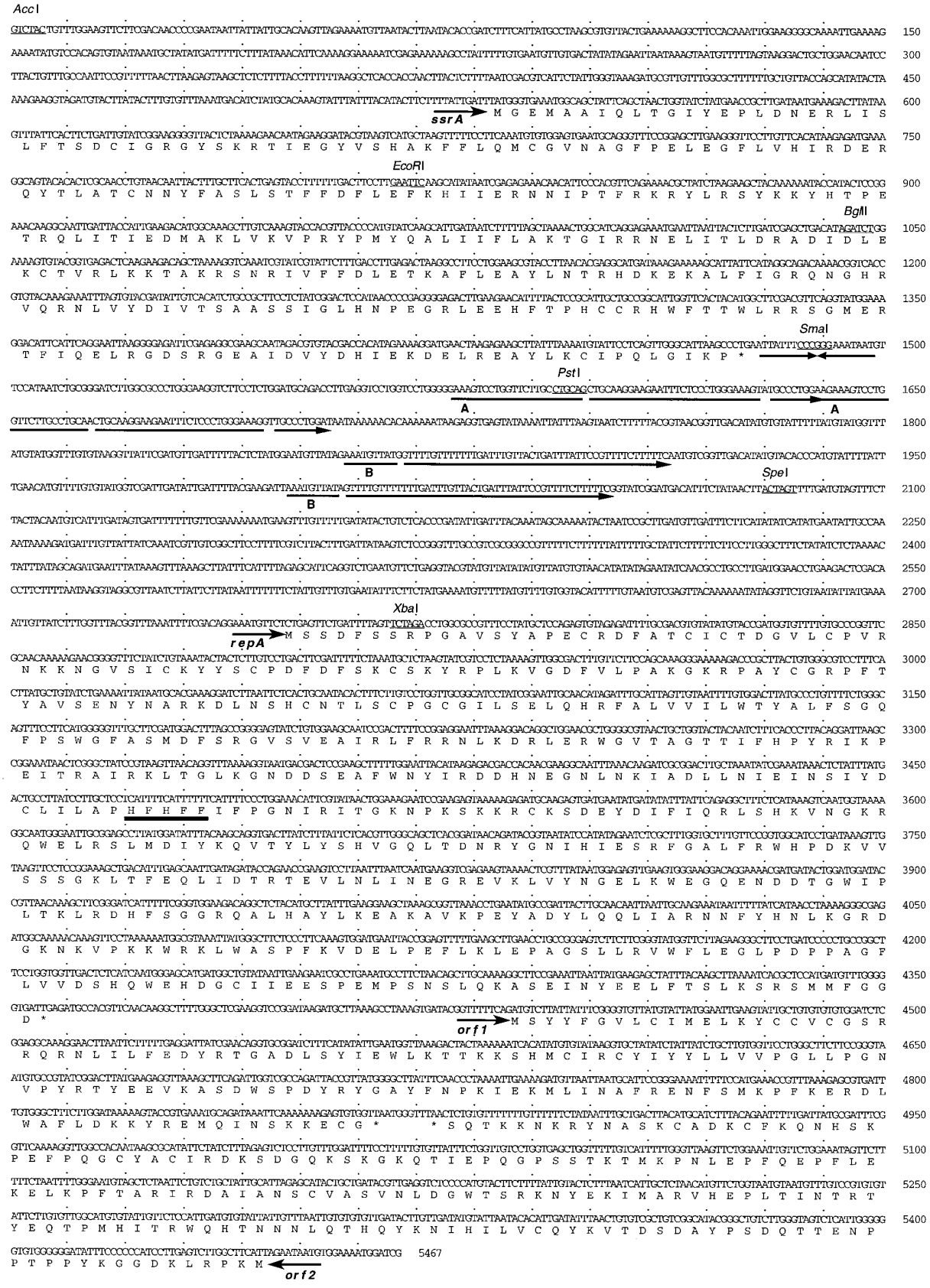


FIG. 1. The DNA sequence of pC2A. The complete 5467-bp DNA sequence is shown (also see Fig. 2). Selected restriction sites are underlined. Four ORFs, *ssrA*, *repA*, *orf1*, and *orf2*, of greater than 120 aa were observed in the sequence. Their orientation is shown by the small arrows. The amino acid sequences of the putative proteins encoded by these ORFs are shown below the corresponding DNA sequences. The HUHUU (U = bulky hydrophobic residues) motif conserved among the Rep proteins of rolling circle plasmids (31) was identified in the RepA protein (heavy underline). Two long direct repeats of 64 bp (A) and 56 bp (B), and a 20-bp perfect inverted repeat which may play a role in plasmid replication were noted in the intergenic region between *ssrA* and *repA*.

were examined for homology with other proteins in the GenBank and EMBL databases. Only one showed strong homology with other proteins. This putative protein shares extensive homology with a family of known site-specific recombinases (32). Its gene was therefore designated *ssrA*. A second ORF, designated *repA*, encodes a putative protein that shares limited homology with the replication initiation proteins from a family of phage and plasmids that replicate by a rolling-circle mechanism (31). The homologous region includes one of three conserved motifs believed to be directly involved in DNA replication (see Fig. 1). An inspection of the sequence surrounding the *repA* gene suggested that it may be translated using CTG as the initiation codon. This CTG codon is preceded by a consensus ribosome binding site (AGGAA), whereas alternate initiation codons are much further downstream and are not preceded by consensus ribosome binding sites. The putative proteins encoded by the other two ORFs, *orf1* and *orf2*, did not share significant homology with any other proteins in the databases. Three structural features that may be relevant for plasmid replication were noted in the pC2A DNA sequence. First, two very long direct repeats, 64 and 56 bp, were observed in the intergenic region between *ssrA* and *repA*. Second, a 20-bp perfect inverted repeat was observed immediately following the SsrA coding sequence.

Replication of pC2A Derivatives in *Methanosarcina*. A series of hybrid plasmids were constructed as potential *Methanosarcina–E. coli* shuttle vectors. The Pur^R gene of *Streptomyces alboniger*, *pac*, was chosen as selectable marker for these plasmids because (*i*) it confers Pur^R upon other methanoarchaea (16), (*ii*) a variety of *Methanosarcina* tested showed complete growth inhibition by puromycin at 0.5 μg/ml (data not shown), and (*iii*) a gene cassette with *pac* transcribed from a strong methanoarchaeal promoter is available (16). To facilitate further constructions, we modified this cassette by removal of numerous extraneous restriction sites, fusion of the *pac* gene directly to the start codon of *mcrB*, and introduction of additional flanking restriction sites. The resulting construct, pJK3, should be generally useful in methanoarchaea (Fig. 2).

Potential shuttle vectors were constructed by cloning pC2A in its entirety into the *E. coli* plasmid pBluescript KS(+). Each was disrupted at a different region of the pC2A replicon in the hope that at least one of the disruptions would not interfere with essential replication functions of the plasmid. The *pac* cassette from pJK3 was then added to each to provide a selectable marker for transformation experiments.

Initially, we attempted to transform both *Methanosarcina acetivorans* and *Methanosarcina barkeri* Fusaro by electroporation under a variety of conditions using several of the shuttle vector candidates. One of these experiments yielded a few Pur^R

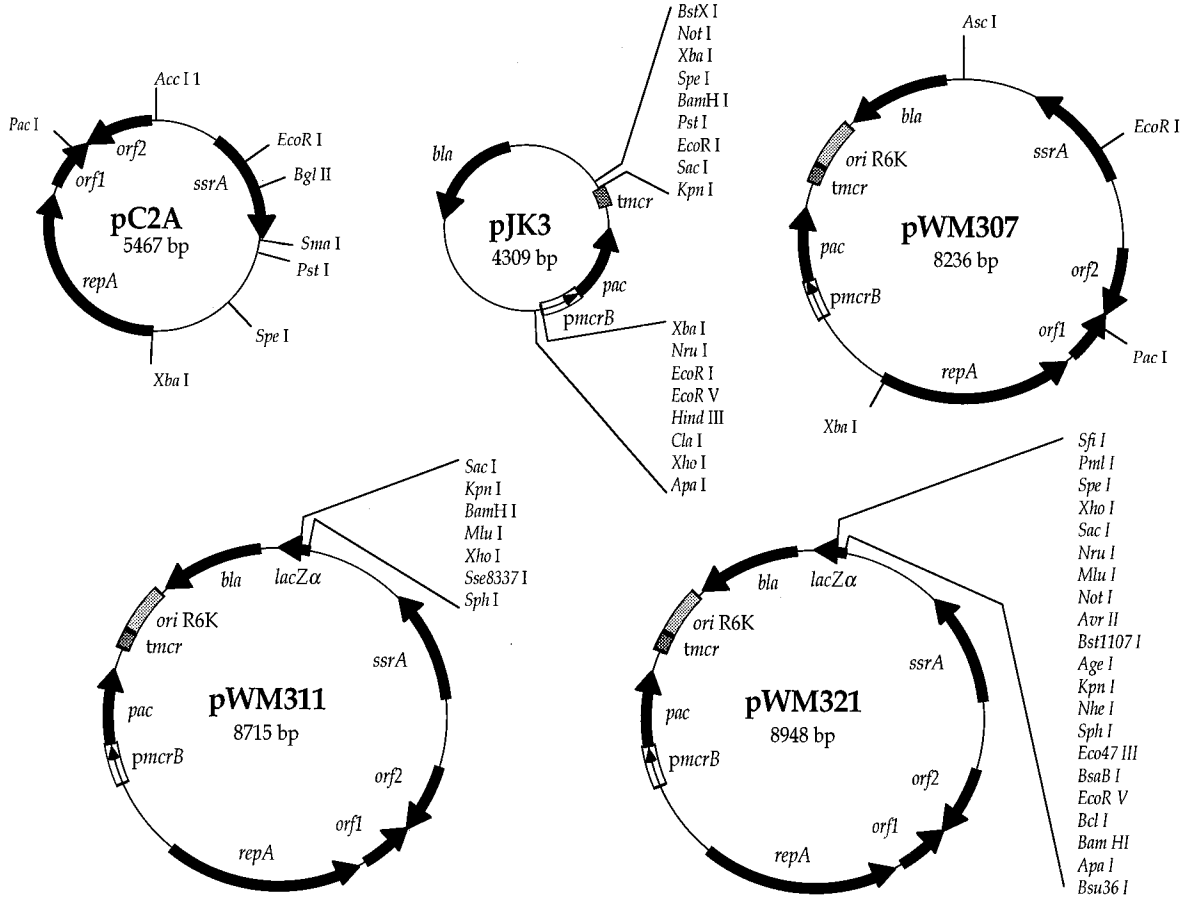


FIG. 2. Plasmids used in the study. pC2A is a naturally occurring plasmid from *Methanosarcina acetivorans*. The genes shown are described in the text. pJK3 is a pBluescript derivative carrying a modified *pac* gene cassette that confers Pur^R upon methanoarchaea. The promoter (p*mcrB*) and terminator (*tmcr*) of the *Methanococcus voltae* methyl reductase operon regulate expression of the puromycin acetyltransferase gene (*pac*) from *Streptomyces alboniger* in methanoarchaea. pWM307, pWM311, and pWM321 are *Methanosarcina–E. coli* shuttle vectors. The origin of replication from the plasmid R6K allows maintenance in *E. coli*, and manipulation of copy number by choice of an appropriate host strain. Each also carries the entire pC2A replicon for replication in *Methanosarcina*. Plasmids pWM309, pWM313, pWM315, pWM317, and pWM319 (not shown) are like pWM311 but carry different polylinkers within the *lacZ*α gene. The *lacZ*α gene allows blue-white screening of recombinant clones in *E. coli* for pWM309, pWM311, pWM313, pWM315, pWM317, and pWM319, but not in pWM321. The β-lactamase gene (*bla*) encodes resistance to penicillin derivatives in *E. coli*.

transformants of *Methanosarcina acetivorans* with the plasmid pJK8, in which pC2A is disrupted at the unique *Spe*I site upstream of *repA*. We were able to isolate intact pJK8 plasmid DNA from these transformants, and with it we could retransform both *E. coli* and *Methanosarcina acetivorans*. These data indicate that the Pur^R strains obtained in this experiment were true transformants and that pJK8 can replicate as a plasmid in either host.

Construction of Improved Shuttle Vectors and Optimization of Transformation Conditions. Plasmid pJK8 lacks many of the features desirable in a cloning vector. We modified pJK8 as described to generate pWM307, pWM309, pWM311, pWM313, pWM315, pWM317, pWM319, and pWM321 (Fig. 2). These plasmids provide a variety of useful features, including blue-white screening for recombinant clones (pWM309, pWM311, pWM313, pWM315, pWM317, and pWM319), symmetrical polylinkers (pWM313 and pWM315), and a large variety of unique restriction sites (pJK21). Also, because these plasmids use the *pir*-dependent R6K γ replication origin, their copy number can be modified from low to very high by using appropriate *E. coli* strains as hosts (33).

Using electroporation in high-resistance medium, the transformation frequency of *Methanosarcina acetivorans* with pJK8 was $\approx 10^2$ per microgram of DNA per 2×10^8 cells. A variety of modifications were tried to increase this frequency. These included varying the voltage, pulse length, and buffer composition, as well as the growth state and number of recipient cells. For these experiments, pWM307 was used because of its smaller size and fewer restriction endonuclease recognition sites. Despite these modifications, we were unable to increase the transformation frequency above $\approx 10^3$ per microgram of DNA per 2×10^8 cells (data not shown). Attempts to modify the PEG-mediated protocol used with *Methanococcus maripaludis*, and the PEG-lithium acetate-mediated protocol used with yeast were even less successful ($<10^1$ per microgram of DNA per 10^8 cells). Natural transformation never yielded transformants.

An alternative method, liposome-mediated transformation, resulted in a dramatic improvement in transformation frequency. Although there was significant variability in the exact number of transformants obtained in each experiment (which we believe is due to the difficulties inherent to plating these extremely oxygen-sensitive anaerobes), we were reproducibly able to achieve at least 10^7 transformants per microgram of DNA per 10^9 cells using *Methanosarcina acetivorans* as host. In some experiments, the transformation frequency was as high as 2×10^8 transformants per microgram of DNA per 10^9 cells, or *ca.* 20% of the recipient population.

Transformation of Other *Methanosarcina* Species by pWM307. Because a variety of *Methanosarcina* species are in routine use, we attempted to transform *Methanosarcina barkeri* Fusaro, *Methanosarcina barkeri* MS, *Methanosarcina barkeri* W, *Methanosarcina mazei* C-16, *Methanosarcina mazei* S-6, *Methanosarcina mazei* LYC, *Methanosarcina thermophila* TM-1, *Methanosarcina siciliae* C2J, *Methanosarcina* spp. WH1, and *Methanosarcina* spp. WH2 with pWM307. Eight of the 10 strains tested gave Pur^R transformants using the optimum liposome-mediated conditions determined with *Methanosarcina acetivorans* as host. Including *Methanosarcina acetivorans*, this represents strains from four of the five known *Methanosarcina* species. Only *Methanosarcina mazei* strains C-16 and LYC failed to yield Pur^R transformants. The transformation frequency in these strains was not determined, although it was clearly much lower than that achieved with *Methanosarcina acetivorans*. However, no attempt was made to optimize transformation conditions for these species.

To verify that these Pur^R strains were pWM307 transformants, we isolated total DNA from selected transformants and from the untransformed parental strains, and tested them for hybridization with pXS2. Plasmid pXS2 caries the *serC* gene from *Methanosarcina barkeri* Fusaro, which was shown to hybridize to all *Methanosarcina* tested (27), as well as the same β-lactamase (*bla*) gene present on pWM307. Therefore, pXS2 will hybridize to both the chromosome and shuttle vector present in each strain. As shown in Fig. 3, a common band corresponding to the chromosomal *serC* locus is seen in both transformed and untransformed isolates of each strain. In the Pur^R transformants only, a second band corresponding to pWM307 is also seen. This second band is the only one seen with pWM307 as probe, a result that verifies the identity of this

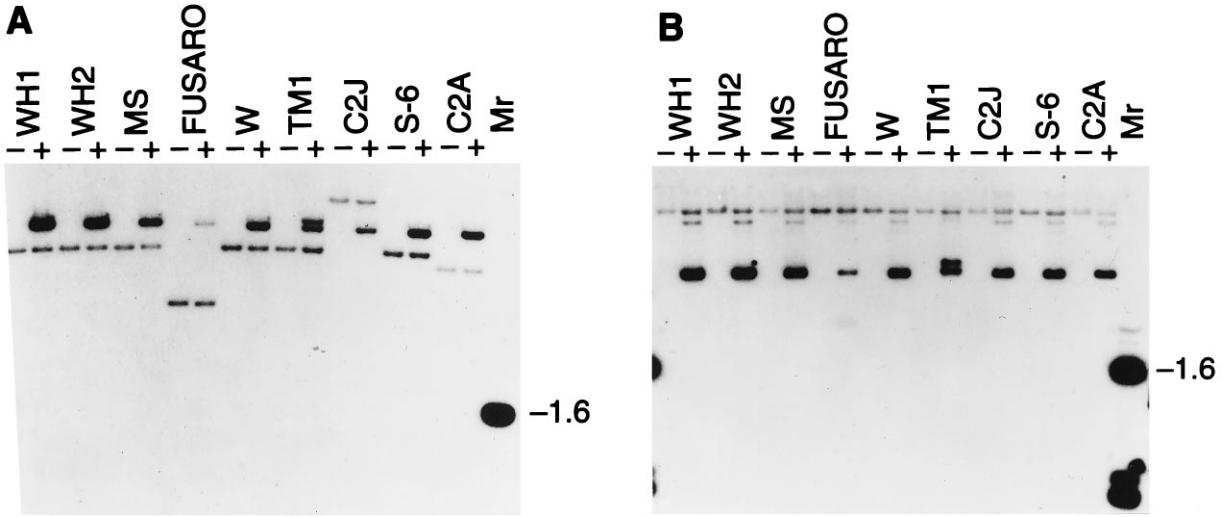


FIG. 3. Transformation of *Methanosarcina* species by pWM307. Total DNA was isolated from each strain and from Pur^R clones obtained from each after transformation with pWM307. *Eco*RI-digested (*A*) or undigested (*B*) DNA was electrophoresed, blotted, and hybridized to labeled pXS2 as described. Plasmid pXS2 hybridizes to both the chromosomal *serC* gene and the pWM307 *bla* gene. A band corresponding to the chromosomal *serC* locus is seen in both transformed and untransformed strains, whereas a band corresponding to pWM307 is seen only in the transformed strains. The pWM307 band migrates with a mobility much higher than the chromosomal *serC* band in undigested DNA, indicating that pWM307 is replicating as a plasmid in these transformants. The strains examined were *Methanosarcina* spp. WH1 (WH1), *Methanosarcina* spp. WH2 (WH2), *Methanosarcina barkeri* MS (MS), *Methanosarcina barkeri* Fusaro (Fusaro), *Methanosarcina barkeri* W (W), *Methanosarcina thermophila* TM1 (TM1), *Methanosarcina siciliae* C2J (C2J), *Methanosarcina mazei* S-6 (S-6), and *Methanosarcina acetivorans* C2A (C2A). +, Pur^R clones; −, untransformed parental strains. The 1.6-kbp band of the molecular weight markers (M_r) hybridizes to pXS2 vector sequences.

second band as the pM307 shuttle vector (data not shown). In undigested total DNA, the pWM307 band has a much greater mobility than the chromosomal *serC* band, indicating that pWM307 is replicating as a plasmid in these strains. In addition to the pWM307 band, *Methanosarcina thermophila* shows a second plasmid band of higher molecular weight. The nature of this second band is unclear at this time; however, we were able to isolate intact pWM307 from each transformant, including *Methanosarcina thermophila*, indicating that unmodified pWM307 was present in each strain tested.

DISCUSSION

Development of genetic techniques for routine analysis of methanoarchaea has been a long-time goal of researchers in the field. With the methods presented here, that goal has been achieved. The availability of a functional cloning vehicle and the application of efficient liposome-mediated DNA delivery methodology will allow genetic experiments that were not previously possible in methanoarchaea. The shuttle vectors reported here possess a variety of features desirable in a cloning vector. Because these vectors function in a number of *Methanosarcina* species, they should be usable without modification by laboratories working with different species within this genus. The shuttle vectors can be used for cloning *Methanosarcina* genes by complementation, and for identification of gene regulatory elements. These applications, however, represent only a small fraction of those possible with these vectors. A myriad of techniques involving plasmids are commonplace in modern molecular genetics. Most of these techniques should now be adaptable to methanoarchaeal research.

Equally important is the finding that liposome-mediated transformation can be highly efficient in methanoarchaea. The use of liposomes as a DNA delivery vehicle in Archaea has not previously been reported to our knowledge. This method is not restricted to transformation by autonomously replicating plasmids. In our experiments with *Methanosarcina acetivorans*, as many as 20% of the recipient cells were transformed by pWM307. Thus, experiments involving relatively infrequent events, such as homologous recombination or transposition, can be performed. We have recently used this method to construct *Methanosarcina* mutants by a reverse genetic approach.

It is likely that liposome-mediated transformation of *Methanosarcina* requires their growth in a single cell morphology (24). *Methanosarcina* species growing as single cells are bound solely by a membrane with a protein coat known as the S-layer. When these cells are suspended in an isoosmotic, Mg^{2+}-free medium, the S-layer is lost as well, exposing the cell membrane. Such exposed membranes probably promote liposome fusion and DNA delivery. We have achieved transformation without resuspending the cells in Mg^{2+}-free medium; however, under these conditions, the frequency is significantly reduced (data not shown). Liposome-mediated transformation may also be applicable to members of the genus *Methanococcus*. Like the single cell morphology of *Methanosarcina*, *Methanococcus* species have a cell structure composed of a membrane and S-layer protein coat, which can be removed by resuspension in appropriate buffer (12).

In combination, these methods represent a functional genetic system for the genus *Methanosarcina*. With the achievement of this goal, wide areas of the unique metabolism and physiology of methanoarchaea are now amenable to genetic analysis.

We thank S. Maloy, J. Konisky, and J. Slauch for providing bacterial strains and plasmids. This work was supported by National Institutes of Health Grant GM51334 and Department of Energy Grants DE-FG02–87ER13651 (R.S.W) and DE-FG02–93ER20106 (K.R.S.). W.W.M. is supported by a National Research Service Award postdoctoral fellowship (1 F32 GM16504–01A1).

1. Balch, W. E., Magrum, L. J., Fox, G. E., Wolfe, R. S. & Woese, C. R. (1977) *J. Mol. Evol.* **9,** 305–311.
2. Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4576–4579.
3. Deppenmeier, U., Muller, V. & Gottschalk, G. (1996) *Arch. Microbiol.* **165,** 149–163.
4. DiMarco, A. A., Bobik, T. A. & Wolfe, R. S. (1990) *Annu. Rev. Biochem.* **59,** 355–394.
5. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., *et al.* (1996) *Science* **273,** 1017–1140.
6. Hilpert, R., Winter, J., Hammes, W. & Kandler, O. (1981) *Zentralbl. Bakteriol. Mikrobiol. Hyg. Abt. 1 Orig. C* **2,** 11–12.
7. Sowers, K. J. & Gunsalus, R. P. (1988) *J. Bacteriol.* **170,** 4979–4982.
8. Hamilton, P. T. & Reeve, J. N. (1985) *Mol. Gen. Genet.* **200,** 47–59.
9. Meile, L., Abendschein, P. & Leisinger, T. (1990) *J. Bacteriol.* **172,** 3507–3508.
10. Bertani, G. & Baresi, L. (1987) *J. Bacteriol.* **169,** 2730–2738.
11. Micheletti, P. A., Sment, K. A. & Konisky, J. (1991) *J. Bacteriol.* **173,** 3414–3418.
12. Patel, G. B., Nash, J. H., Agnew, B. J. & Sprott, G. D. (1994) *Appl. Environ. Microbiol.* **60,** 903–907.
13. Possot, O., Gernhardt, P., Klein, A. & Sibold, L. (1988) *Appl. Environ. Microbiol.* **54,** 734–740.
14. Kiener, A., Rechsteiner, T. & Leisinger, T. (1986) *FEMS Microbiol. Lett.* **33,** 15–18.
15. Argyle, J. L., Tumbula, D. L. & Leigh, J. A. (1996) *Appl. Environ. Microbiol.* **62,** 4233–4237.
16. Gernhardt, P., Possot, O., Foglino, M., Sibold, L. & Klein, A. (1990) *Mol. Gen. Genet.* **221,** 273–279.
17. Berghoffer, Y. & Klein, A. (1996) *Appl. Environ. Microbiol.* **61,** 1770–1775.
18. Tumbula, D. L., Makula, R. A. & Whitman, W. B. (1994) *FEMS Microbiol. Lett.* **121,** 309–314.
19. Ladapo, J. & Whitman, W. B. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5598–5602.
20. Jain, M. K. & Zeikus, J. G. (1987) *Appl. Environ. Microbiol.* **53,** 1387–1390.
21. Kiener, A., Holliger, C. & Leisinger, T. (1984) *Arch. Microbiol.* **139,** 87–90.
22. Wanner, B. L. (1986) *J. Mol. Biol.* **191,** 39–58.
23. Miller, V. L. & Mekalanos, J. J. (1988) *J. Bacteriol.* **170,** 2575–2583.
24. Sowers, K. R., Boone, J. & Gunsalus, R. P. (1993) *Appl. Environ. Microbiol.* **59,** 3832–3839.
25. Sowers, K. R. & Schreier, H. J. (1995) *Archaea: A Laboratory Manual*, eds. Robb, F. T., Place, A. R., Sowers, K. R., Schreier, H. J., DasSarma, S. & Fleischmann, E. M. (Cold Spring Harbor Lab., Plainview, NY), Vol. 2.
26. Apolinario, E. A. & Sowers, K. R. (1996) *FEMS Microbiol. Lett.* **145,** 131–137.
27. Metcalf, W. W., Zhang, J.-K., Shi, X. & Wolfe, R. S. (1996) *J. Bacteriol.* **178,** 5797–5802.
28. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1992) *Current Protocols in Molecular Biology* (Wiley, New York), Vols. 1 and 2.
29. Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. (1988) *Gene* **68,** 139–149.
30. Elble, R. (1992) *BioTechniques* **13,** 18–20.
31. Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V. L., Pierson, L. S., Sternberg, N. & Leong, J. M. (1986) *EMBO J.* **5,** 433–440.
32. Ilyina, T. V. & Koonin E. V. (1992) *Nucleic Acids Res.* **20,** 3279–3285.
33. Metcalf, W. W., Jiang, W. & Wanner, B. L. (1994) *Gene* **138,** 1–7.

Week 9

Genetics and Geobiology

Metagenomics

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 2003, p. 2800–2809 Vol. 69, No. 5
0099-2240/03/$08.00+0 DOI: 10.1128/AEM.69.5.2800–2809.2003

The *ars* Detoxification System Is Advantageous but Not Required for As(V) Respiration by the Genetically Tractable *Shewanella* Species Strain ANA-3

Chad W. Saltikov,[1] Ana Cifuentes,[1] Kasthuri Venkateswaran,[2] and Dianne K. Newman[1]*

Department Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125,[1] *and Jet Propulsion Laboratory, Planetary Protection Technologies, Pasadena, California 91109*[2]

Received 14 October 2002/Accepted 26 February 2003

Arsenate [As(V); $HAsO_4^{2-}$] respiration by bacteria is poorly understood at the molecular level largely due to a paucity of genetically tractable organisms with this metabolic capability. We report here the isolation of a new As(V)-respiring strain (ANA-3) that is phylogenetically related to members of the genus *Shewanella* and that also provides a useful model system with which to explore the molecular basis of As(V) respiration. This gram-negative strain stoichiometrically couples the oxidation of lactate to acetate with the reduction of As(V) to arsenite [As(III); $HAsO_2$]. The generation time and lactate molar growth yield ($Y_{lactate}$) are 2.8 h and 10.0 g of cells mol of lactate^{-1}, respectively, when it is grown anaerobically on lactate and As(V). ANA-3 uses a wide variety of terminal electron acceptors, including oxygen, soluble ferric iron, oxides of iron and manganese, nitrate, fumarate, the humic acid functional analog 2,6-anthraquinone disulfonate, and thiosulfate. ANA-3 also reduces As(V) to As(III) in the presence of oxygen and resists high concentrations of As(III) (up to 10 mM) when grown under either aerobic or anaerobic conditions. ANA-3 possesses an *ars* operon (*arsDABC*) that allows it to resist high levels of As(III); this operon also confers resistance to the As-sensitive strains *Shewanella oneidensis* MR-1 and *Escherichia coli* AW3110. When the gene encoding the As(III) efflux pump, *arsB*, is inactivated in ANA-3 by a polar mutation that also eliminates the expression of *arsC*, which encodes an As(V) reductase, the resulting As(III)-sensitive strain still respires As(V); however, the generation time and the $Y_{lactate}$ value are two- and threefold lower, respectively, than those of the wild type. These results suggest that ArsB and ArsC may be useful for As(V)-respiring bacteria in environments where As concentrations are high, but that neither is required for respiration.

The contamination of groundwaters and surface waters with arsenic (As) is a major concern to public health in countries such as Bangladesh, China, Taiwan, Argentina, Chile, and the United States (40). Elevated As concentrations typically derive from the weathering of As-bearing minerals and/or from geothermal sources (2, 58). It is now known that a variety of microorganisms, including members of the *Eukarya*, *Archaea*, and *Bacteria*, influence As geochemistry in many locales throughout the world by virtue of their metabolism (31, 35, 53). These metabolic processes include oxidation (49, 58), reduction (1, 13), and methylation reactions (5) that strongly affect (and in some cases, control) As speciation in the environment. One process that is particularly intriguing is microbial respiration of arsenate [As(V); $HAsO_4^{-2}$]. In the absence of oxygen, microorganisms can gain energy by coupling the oxidation of organic material to As(V) reduction, resulting in the production of the highly toxic As compound, arsenite [As(III); $HAsO_2$]. As(V)-respiring organisms can affect water quality by catalyzing the mobilization of As(III) from sediments (1), as well as affect the biogeochemical cycles of other elements. For example, a significant proportion (~14%) of organic carbon remineralization to CO_2 within the hyper-saline waters of Mono Lake, Calif., has been linked to the activity of As(V)-respiring microorganisms (42).

* Corresponding author. Mailing address: Department Geological and Planetary Sciences, California Institute of Technology, Mailstop 100-23, Pasadena, CA 91125. Phone: (626) 395-6790. Fax: (626) 683-0621. E-mail: dkn@gps.caltech.edu.

To date, numerous phylogenetically diverse bacteria have been isolated that use As(V) as a terminal electron acceptor for respiratory growth, suggesting that this metabolic process may be ancient in origin (35, 53). As(V)-respiring organisms have been isolated from various sites including: a Superfund site contaminated with As (1), a seleniferous freshwater marsh in Nevada (41), an Australian goldfield (27), mud from a reed bed in Australia (28), a freshwater lake in Massachusetts (37), an alkaline hypersaline lake in California (6), geothermal pools within Yellowstone National Park (15, 19), an As-contaminated lake in Idaho (39), bovine rumen fluid, hamster feces, and termite hind guts (17). All of these As(V)-respiring strains are obligate anaerobes.

Until now, only three studies have investigated the molecular basis of As(V) respiration (23, 28, 35; D. K. Newman, C. W. Saltikov, E. Afkar, S. Tiwari, B. W. Kail, R. S. Oremland, F. M. M. Morel, and J. F. Stolz, unpublished data). Although the enzymology of this process is emerging, biochemical approaches alone will not be sufficient to determine how As(V)-respiration is regulated and/or functionally integrated with other cellular pathways that traffic in As. Of specific interest is the relationship between pathways that control As(V) respiration and those that control As resistance, given that an inescapable consequence of As(V) respiration is the buildup of toxic As(III). Arsenic detoxification by the products of the *ars* genes has been studied in great detail (reviewed in references 31 and 44), and the *ars* genes have been found in many organisms (50). The *ars* operon encodes a multisubunit As(III) efflux

TABLE 1. Strains and plasmids used in this study

Bacterial strain or plasmid	Genotype or markers and characteristics and uses[a]	Source or reference
E. coli		
DH10β	Host for *E. coli* cloning; F^- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*dlacZ*ΔM15 Δ(*codB-lac1*)*3 deoR recA1 endA1 araD139* Δ(*ara-leu*)*7697 galU galK* λ^- *rpsL* (Str^r)	Life Technology
UQ950	*E. coli* DH5α λ*pir* host for cloning; F^- Δ(*argF-lac*)*169* ϕ80d*lacZ58*(ΔM15) *glnV44*(AS) *rfbD1 gyrA96* (Nal^r) *recA1 endA1 spoT1 thi-1 hsdR17 deoR* λ*pir*$^+$	D. Lies, Caltech
β2155	Donor for bacterial conjugation; F′ *lacZ58*(ΔM15) *lac1*^{q} traD36 proA$^+$B$^+$/λ$^-$(?) *thrB1004 pro thi rpsL* (Str^r) *hsdS lacZ58*(ΔM15) Δ*dapA*::*erm* (Erm^r) *pir*$^+$::RP4-2-Tc::Mu (Km^r)	12
AW3110	W3110 Δ*ars*::*cam*	9
Shewanella spp.		
ANA-3	Isolated from an As-treated wooden pier piling in a brackish estuary (Eel Pond, Woods Hole, Mass.), contains *arsDABC*, respires on As(V) resistant to >5 mM arsenite	This study
ARSB1	*arsB* mutant derived from ANA-3; *arsB*::Kan(EZ::TN<KAN-2>)	This study
S. oneidensis MR-1	Manganese-reducing strain from Oneida Lake, N.Y., sediments, type strain	32
S. algae OK-1	ATCC 51192, type strain, isolated from red algae	51
S. amazonensis SB2B	Isolated from Amazon water, type strain	55
S. baltica 63	NCTC 10735, Japan, isolated from oil brine, type strain	60
Shewanella sp. strain MR-4	Isolated from the Black Sea water column, type strain	33
S. pealeana ANG-SQ1	Isolated from accessory nidamental gland of a squid, psychrotolerant, type strain	25
S. frigidimarina ACAM 591	Isolated from Antarctic Sea ice, type strain	7
S. putrefaciens 95	ATCC 8071, isolated from spoiled butter with surface taint, type strain	46
S. woodyi MS32	Isolated from seawater detritus, Alboran Sea, type strain	29
184	ATCC 8073, isolated from spoiled butter with surface taint	46
CL 256/73	NCTC 12093, isolate from human cerebrospinal fluid, type strain	18
CN32	Isolated from anaerobic subsurface core sample, New Mexico	14, 59
Plasmids and vectors		
pLAFR5	21.5-kb broad-host-range cosmid cloning vector; Tc^r, *lacZ*	22
EZ::TN<KAN-2>	1.2-kb transposon used for in vitro mutagenesis	Epicentre
pSALT1	pLAFR5-based 45-kb As(III)^r cosmid from ANA-3 genomic DNA library; contains *arsDABC*, Tc^r, confers resistance to As(III) and As(V)	This study
pSALT1-B10	As(III)-sensitive pSALT1; *arsB*::*kan*(EZ::TN<KAN-2>), Km^r Tc^r	This study
pSMV8	9.1-kb mobilizable suicide vector; *oriR6K mobRP4*, Gm^r	D. Lies, Caltech
parsB::*kan*	pSMV8 with Km^r Tn-interrupted *arsB* gene from pSALT1-B10	This study

[a] Str^r, streptomycin resistance; Nal^r, nalidixic acid resistance; Tc^r, tetracycline resistance; Gm^r, gentamicin resistance.

pump comprising a transmembrane oxyanion conducting channel, ArsB, that often associates with an ATPase subunit, ArsA. In addition, As(V) resistance is conferred by a small 16-kDa cytoplasmic As(V) reductase, ArsC, that reduces As(V) to As(III). Regulation of the *ars* operon is controlled by the As(III)-sensitive *trans*-acting repressor, ArsR, and the inducer-independent *trans*-acting repressor, ArsD. The *ars* operon functions to lower the intracellular As concentration, which permits survival in environments with high concentrations of As. Although it seems logical that the *ars* genes might be present and functional in As(V)-reducing bacteria, this has not been directly proven to date.

As an entry into exploring the relationship between As(V) respiration and As resistance at the molecular level, we report here the isolation and characterization of an As(V)-respiring facultative anaerobe. The focus of this study concerns whether the *arsB* and *arsC* genes are required for As(V) respiration.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. The strains and plasmids used in the present study are listed in Table 1. *E. coli* strains were grown in Luria-Bertani (LB) Miller medium (Difco). The growth conditions for various *Shewanella* strains are described below.

Isolation of ANA-3. ANA-3 was isolated in 1998 from an As-treated wooden pier located in a brackish estuary (Eel Pond [Woods Hole, Mass.]). Arsenate reducers were enriched by placing a 1-by-3-cm strip of the wood sample into a defined minimal medium (pH 7.2; $NaHCO_3$ [1.9 g/liter], KH_2PO_4 [0.2 g/liter], NH_4Cl [0.25 g/liter], KCl [0.5 g/liter], $CaCl_2 \cdot 2H_2O$ [0.1 g/liter], NaCl [1.0 g/liter], $MgCl_2 \cdot 6H_2O$ [0.4 g/liter], 1 ml of SL10 trace elements/liter, and 1 ml of vitamin solution [37]/liter) amended with lactate (10 mM), arsenate (5 mM), and sulfide (1 mM) and grown under anaerobic conditions by using the Hungate technique (30). As(V) reduction was visually observed by the formation of the yellow mineral As_2S_3 from the reaction of As(III) and S^{2-} (36). The sample was subcultured into fresh medium once As(V) reduction had occurred. Subculturing was repeated three times before finally plating on LB agar aerobically. A single colony was inoculated back into the defined minimal medium, and after several days a yellow precipitate developed. This isolate was named strain ANA-3.

ANA-3 was routinely grown in either LB medium or a minimal medium (pH 7.2) containing the following: 0.225 g of K_2HPO_4/liter, 0.225 KH_2PO_4/liter, 0.46 g of NaCl, 0.225 g of $(NH_4)_2SO_4$/liter, g of 0.117 $MgSO_4 \cdot 7H_2O$/liter, 2.24 g of sodium lactate/liter, 10 mM Na_2HAsO_4 or 20 mM $NaNO_3$, 4.2 g of $NaHCO_3$/liter, SL10 trace elements (1 ml), and vitamins (10 ml) (41). The medium was boiled under a stream of N_2-CO_2 (4:1), dispensed anaerobically into bottles flushed with the mixed gas, and autoclave sterilized. Sterile anaerobic sodium bicarbonate solution was added by injection after the tubes had cooled.

Phylogenetic analysis. Ten nanograms of purified genomic DNA (21) from liquid cultures was used as the template for PCR amplification. Universal 16S ribosomal DNA (rDNA) primers (Bact 11 and 1492) were used to amplify the 1.5-kb 16S rDNA fragment according to established protocols (46). Amplicons were sequenced directly after purification on Qiagen columns (Qiagen, Valencia, Calif.). The PCR product was sequenced by using the dideoxy chain termination method with the Sequenase DNA sequencing kit (U.S. Biochemical Corp., Cleveland, Ohio) and an ABI 373A automated sequencer (Perkin-Elmer Corp., Foster City, Calif.). The phylogenetic relationships of organisms covered in the present study were determined by comparison of individual 16S rDNA sequences to other existing sequences in the public database (GenBank [http://www.ncbi.nlm.nih.gov/]). Sequence alignments were obtained online from the Ribosomal Database Project (RDP [http://rdp.cme.msu.edu/html/]). Evolutionary trees were constructed by using PAUP*, version 4.0b10 (54), with the optimality criterion set to distance (minimum evolution). The Kimura two-parameter model was

TABLE 2. Growth characteristics of *Shewanella* sp. strain ANA-3

Electron donor or fermentation substrate (concn [mM])	Physiological parameter		
	Growth[a]	TEA[b] (concn [mM])	Reduction of TEA
Electron donors[c]		Oxygen	+
Acetate (10)	− (+)	Arsenate (10)	+
Citrate (10)	−	Fumarate (20)	+
Ethanol (10)	−	Selenate (5)	−
Formate (10)	−	Nitrate (5)	+
Fumarate (10)	−	MnO_2 (20)	+
Glucose (10)	−	$Fe(OH)_3$ (50)	+
Glycerol (5)	−	AQDS (5)	+
Lactate (10)	+ (+)	Sulfate (10)	−
Malate (10)	−	Thiosulfate (5)	+
Propionate (10)	−	Sulfite (5)	−
Pyruvate (10)	+ (+)	DMSO (10)	−
Succinate (10)	−		
Fermentation			
Lactate (10)	−		
Pyruvate (10)	−		

[a] +, Growth; −, no growth; parentheses indicate aerobic growth; no parenthesis indicates that aerobic growth on the substrate was not tested.
[b] 20 mM lactate was used as the electron donor. TEA, terminal electron acceptor; DMSO, dimethyl sulfoxide.
[c] 5 mM arsenate was used as the terminal electron acceptor.

used to estimate pairwise distances. Phylogenetic trees were inferred by neighbor-joining and tree bisection-reconnection branch-swapping algorithms. After a heuristic search was performed, bootstrap analysis was done with 1,000 replications. The final tree was assembled in Dendromaker (http://www.cib.nig.ac.jp/dda/timanish/dendromaker/home.html) and with Adobe Illustrator (Adobe Systems, Inc.). The GenBank nucleotide accession number for strain ANA-3 is AF136392.

Electron donors and acceptors. Various electron donors listed in Table 2 were screened for the ability to support growth on As(V) as the sole terminal electron acceptor. Lactate was used as the electron donor and sole carbon source for testing the electron acceptors listed in Table 2. Fumarate and As(V) reduction were determined by high-pressure liquid chromatography analysis (described below). The humic acid functional analog 2,6-anthraquinone disulfonate (AQDS) reduction was determined spectrophotometrically by monitoring absorption at 450 nm (38). Thiosulfate reduction was confirmed by colorimetric detection of hydrogen sulfide by the methylene blue method (10). Nitrate reduction was determined by monitoring the formation of nitrite by colorimetry with Griess reagent (sulfanilamide and *N*-naphthylethylenediamine in HCl) (52). Mineral reduction was monitored by observing the change in color of the iron oxide (from rust to dark brown) or the transformation of the manganese oxide from black to white. Minerals were prepared as described by Lovley and Phillips (26). Growth was inferred either by monitoring increases in CFU (per milliliter) or by visually inspecting increases in turbidity compared to controls without electron donor or acceptor.

Construction of the genomic library. Genomic DNA was prepared according to standard methods (3), partially digested with *Sau*3AI, and size fractionated on a 10 to 40% sucrose gradient (48). DNA fragments of 20 to 30 kb were ligated to the cosmid vector pLAFR5 previously digested with *Sca*I/*Bam*HI. After being packaged into phage by using Gigapack Gold XL (Stratagene, La Jolla, Calif.), the cosmid library was transduced into *Escherichia coli* β2155 (12), mated en masse into *E. coli* AW3110 (9), and plated onto LB agar containing tetracycline (15 μg/ml) and 5 mM sodium *meta*-arsenite (Sigma). Cosmid DNA from an As-resistant clone was isolated and transferred back into As(III)-sensitive AW3110 by electroporation to confirm that it conferred the As(III)-resistant phenotype. This cosmid was designated pSALT1. The region of pSALT1 that conferred As(III) resistance was mapped by in vitro transposon mutagenesis by using the EZ::TN<KAN-2> system (Epicentre) according to the manufacturer's instructions. Randomly mutagenized pSALT1 was electroporated into AW3110, and clones were screened for sensitivity to 5 mM As(III). After the flanking region of the transposon of an As(III)-sensitive clone (pSALT1-B10) was sequenced by using EZ::TN<KAN-2> primers (supplied by Epicentre), the nucleotide sequence was analyzed by BLAST searching (http://www.ncbi.nlm.nih.gov/BLAST/). A 5-kb region was sequenced by primer walking upstream and downstream of the initial sequence. The nucleotide sequence was assembled by using AssemblyLIGN (Accelrys) and submitted to the National Center for Biotechnology Information (accession no. AY161137).

Construction of *arsB* insertion mutation. An *arsB* gene replacement mutant was constructed from ANA-3 by exchanging the wild-type allele for the mutant allele of pSALT1-B10. PCR was used to generate a fragment with *Spe*I ends (underlined) from the mutated cosmid pSALT1-B10 by using the primers TNARSBF (GGACTAGTATGGGACGATTGATTAGGATGG) and TNARSBR (GGACTAGTGGTCGTGGCCGTTTACTCTTTA). The resulting 2.8-kb fragment contained the 1.2-kb kanamycin-resistant (Km^r) transposon flanked by ~800 bp of *arsB* on either side and was cloned into the *Spe*I site of the mobilizable suicide vector pSMV8 to generate p*arsB*::*kan*. The mutation was introduced into ANA-3 by conjugation from the *E. coli* donor strain β2155 containing p*arsB*::*kan*. Overnight cultures of the donor (800 μl) and ANA-3 (200 μl) were centrifuged together, resuspended in ~40 μl, and spotted onto LB agar containing 300 μM diaminopimelic acid. The mating reaction was incubated at 30°C for 6 h prior to plating onto LB medium plus kanamycin (50 μg/ml) without diaminopimelic acid. After overnight incubation at 30°C, 12 Km^r colonies were picked and tested for sensitivity to gentamicin (to indicate loss of pSMV8) and then analyzed for recombination of the mutant allele by PCR with primers TNARSBF and TNARSBR.

RT-PCR analysis. Overnight cultures of ANA-3 and ARSB1 grown in LB medium were diluted 1/25 into LB medium amended with 1 mM As(V). After incubation at 30°C for 4 h, 1 ml was used to isolate total RNA by using the Trizol reagent (Invitrogen). Crude RNA samples were DNase treated and cleaned up by using the Qiagen RNeasy Mini kit. Reverse transcription (RT) was performed with primers 16S-1492-R1 (GGTTACCTTGTTACGACTT), ARSA-R1 (GGCTTAATCGTTCAATACCAAT), and ARSC-R1 (TCACTACTTCACCGTCTTCCTT) and 1 μg of DNase-treated RNA. Control reactions consisted of (i) primer without RT and (ii) RT without primer. RT reactions were diluted 1/50 into sterile nuclease-free water, followed by PCR analysis with the corresponding reverse primers used in the RT reactions and the following forward primers: 16S-8-F1 (AGAGTTTGATCCTGGCTCAG), ARSA-F1 (GCTAGAAGAGGATTTACGCTCA), or ARSC-F1 (CCAACCATTATCCTCTACCTTG). PCR products were analyzed on 1% agarose gels.

Arsenate respiration experiments. Overnight cultures grown anaerobically on 20 mM lactate and 20 mM fumarate were used as the inocula for experiments to check for respiratory growth on 10 mM As(V). Cultures were centrifuged and rinsed twice in anaerobic minimal medium [without lactate and As(V)] and resuspended at ~10^8 cells/ml. Washed cells were inoculated into 100 ml of low-phosphate (~0.3 mM) minimal medium amended with arsenate (10 mM) and lactate (20 mM) at ~10^6 cells/ml and incubated anaerobically at 30°C without shaking. Control experiments with or without As(V) and/or lactate were also done to determine whether ANA-3 could grow in the absence of either a terminal electron acceptor or electron donor. Cultures were sampled periodically and analyzed for cell density by staining formaldehyde-fixed cells with 1 μg DAPI (4′,6′-diamidino-2-phenylindole)/ml, followed by filtration onto polycarbonate Nuclepore (Millipore Corp.) membranes (0.2 μm [pore size]). Stained cells were enumerated by epifluorescence microscopy on a Zeiss Axioplan (Carl Ziess MicroImaging, Inc.). Arsenic compounds [As(V) and As(III)], lactate, and acetate were quantified by high-pressure liquid chromatography (Waters) by using a Hamilton PRP-X300 column in series with a Bio-Rad Aminex HPX-87H column heated to 50°C. A mobile phase of phosphoric acid (30 mM) was set to 0.7 ml/min. Compounds were detected by UV at 210 nm.

Other *Shewanella* species were tested for the ability to respire As(V) by inoculating ~10^6 cells/ml into anaerobic LML medium (4) containing 5 mM As(V) as the electron acceptor and lactate as the carbon source and electron donor. Cultures were sampled before and after 24 h of incubation, and As(V) was measured by using the molybdenum blue assay (20).

Resistance to As(III). A microtiter plate assay was developed to determine aerobic As(III) sensitivity for various strains listed in Table 1. Overnight LB medium-grown cultures were diluted 100-fold into fresh LB medium amended with increasing As(III) concentrations. A total of 150 μl of each arsenic concentration was pipetted in quadruplicate into a 96-well microtiter dish and then incubated at 30°C and at 100 rpm for 24 h. Growth was monitored by measuring the optical density at 630 nm (OD_{630}) before and after the incubation period in a Dynex Opsys microplate reader (Dynex Technologies).

Anaerobic As(III) resistance in ANA-3 and the *arsB* mutant of ANA-3 (ARSB1) was tested by inoculating 1/100 of anaerobic starter cultures grown in 20 mM lactate and 20 mM fumarate into anaerobic Hungate tubes containing minimal medium supplemented with lactate (20 mM), fumarate (20 mM), and increasing As(III) concentrations. The OD_{600} was monitored periodically for

68 h. The maximum OD values reached during the incubation period were used for determining the resistance profiles on increasing As(III) concentrations.

RESULTS

Enrichment and isolation. By targeting the tidal interface of an old marine pier piling whose wood had been treated with an As-based preservative, we hypothesized that we would be able to enrich for a facultative anaerobe that could respire As(V). Our goal was to isolate an organism that we could develop into a model system for studying As(V) respiration and arsenic resistance. After several passages of the enrichment culture, followed by plating on LB agar, only coral-pink colonies were observed on the plates, and one was inoculated into anaerobic As(V) medium. Growth and As(V) reduction were observed within several days. This rod-shaped, 0.5-by-2.5-μm strain was designated ANA-3.

16S rDNA phylogeny. Phylogenetic analyses of the 16S rDNA sequence demonstrated that ANA-3 belongs to the *Proteobacteria*, gamma subdivision, genus *Shewanella*. Similarities among the 16S rDNA nucleotide sequences between ANA-3 and other *Shewanella* species in the RDP database are between 93 and 98%. A sequence variation of 1.7% was found between ANA-3 and *S. putrefaciens* ATCC 8071, and it was 2.8% between ANA-3 and *S. oneidensis* strain MR-1 (whose genome has been completely sequenced [16]). Of the strains we included in our analysis, we noticed the greatest sequence variation (6.3%) between ANA-3 and *S. hanedai*. A phylogenetic tree of the 16S rDNA sequences of strains from various sources (56) is shown in Fig. 1. ANA-3 is most closely related to *S. putrefaciens* based on the percent 16S rDNA similarity but does not cluster tightly with *S. putrefaciens* in the phylogenetic tree.

Arsenate respiration. To confirm that ANA-3 was capable of respiring As(V), we characterized the growth of ANA-3 when As(V) served as the sole terminal electron acceptor. Figure 2A shows a time course for As(V) respiration and growth. Cell density increased by several logs to a maximum of 1.6×10^8 cells/ml. Controls without either electron donor (lactate) or acceptor [As(V)] exhibited neither As(V) reduction nor cell growth (data not shown). The early stationary phase was reached by ~23 h. The generation time for ANA-3 was ~2.8 h. After 23 h the initial 10 mM concentration of As(V) was completely reduced to 10 mM As(III). Concurrently, 4.4 mM of lactate was oxidized to acetate (Fig. 2B). We observed a lactate molar growth yield ($Y_{lactate}$) of 10.2 g of cells/mol of lactate with ANA-3, assuming a cell dry weight of 2.8×10^{-13} g/cell (34). The oxidation of lactate and reduction of As(V) represents close to a 2:1 stoichiometric conversion of As(V) to As(III) and lactate to acetate as expected for the following reaction:

$$\underset{(\text{lactate}^-)}{CH_3\text{-}CHOH\text{-}COO^-} + 2HAsO_4^{2-} + 3H^+ \rightarrow \underset{(\text{acetate}^-)}{CH_3COO^-} + 2HAsO_2 + 2H_2O + HCO_3^-$$

where $\Delta G^\circ = -287.6$ kJ/mol of lactate (-71.7 kJ/mol electron).

In comparison, we tested the other *Shewanella* species listed in Table 1 for the ability to respire As(V), but none of them were able to do so. Arsenate thus does not appear to be a common electron acceptor for *Shewanella* species.

Other growth characteristics. ANA-3 respired on a variety of electron acceptors, including metal oxides of iron and manganese (Table 2). Among the carbon sources tested, lactate and pyruvate were the only electron donors that could support growth on As(V) (Table 2). Fermentation was not observed with either of these substrates, although ANA-3 was capable of metabolizing cysteine, evolving H_2S (data not shown). We observed the formation of As_2S_3 in anaerobic As(V)-reducing cultures of ANA-3 when cysteine was included in the medium as a reducing agent. The precipitation of As_2S_3 by As(V) reducing microorganisms has been described elsewhere (36). ANA-3 completely reduced As(V) when grown aerobically in LB medium supplemented with 5 mM As(V) and could grow in the presence of 10 mM As(III).

Identification of the ANA-3 *ars* operon. When ANA-3 was grown aerobically in LB medium, cell densities of $\sim 5 \times 10^9$ cells/ml were reached in overnight cultures. Similar cell densities were also observed in aerobically incubated LB medium-grown cultures supplemented with 5 mM As(III) or As(V). Given ANA-3's ability to resist the toxicity of As(V) and As(III) when grown in LB medium, we hypothesized that it might contain an *ars* operon. To test this, we identified a region of DNA from an ANA-3 genomic library that conferred high-level resistance to As(III) on other bacteria. Resistance to As(III) up to 10 mM was observed when the cosmid (pSALT1) containing this region was transformed into As(III)-sensitive strains of *E. coli* AW3110 (Fig. 3A) and *S. oneidensis* MR-1 (Fig. 3B). The *arsB* gene was shown to be essential for As(III) resistance. The *ars* deletion *E. coli* strain AW3110 harboring the mutagenized cosmid pSALT1-B10 no longer grew on LB agar plates containing 5 mM As(III). The genes on pSALT1 conferred As(III) resistance under aerobic conditions but were not sufficient to confer the ability to respire As(V), however, since pSALT1 was unable to promote growth on As(V) in addition to As(V) reduction when transformed into *S. oneidensis* strain MR-1 (data not shown).

Molecular analysis of pSALT1 revealed the presence of four genes, *arsDABC* (Fig. 4) but no *arsR* homolog within 5 kb upstream of *arsD* and 1 kb downstream of *arsC*. The lack of an *arsR* gene upstream of *arsD* was intriguing, since ArsR is a repressor for the expression of the *ars* operon and the *arsR* gene is commonly found immediately upstream of the *arsDABC* gene cluster (44). The *arsR* gene in ANA-3 may be distantly located from the *arsDABC* cluster or it is possible that this *ars* operon may be regulated in a different way. The ArsD, ArsA, ArsB, and ArsC of ANA-3 are predicted to be similar to those found on the *E. coli* plasmid R773 (Table 3) but only exhibit low amino acid sequence similarity to homologs found in the *S. oneidensis* MR-1 genome (Table 3). BLAST searching the GenBank database with the putative arsenic resistance proteins in the *S. oneidensis* MR-1 genome suggests that their closest relatives are found in *Pseudomonas aeruginosa* PAO1 (e.g., ArsR and ArsCs) and *Pyrococcus furiosus* DSM 3638 (e.g., ACR3) (Table 4). No ArsB-like homologs were found in the *S. oneidensis* MR-1 genomic database.

ANA-3 can be maintained on LB medium for multiple generations in the absence of As selection without losing its arsenic resistance. Because the cosmid library was generated

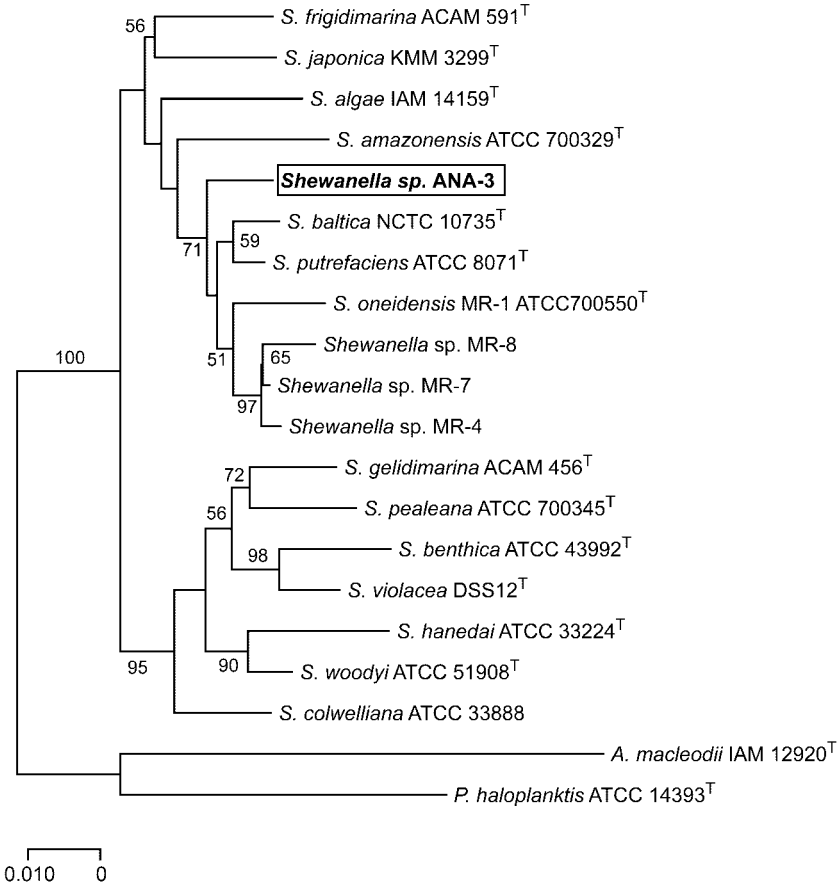


FIG. 1. Phylogenetic relationships among 16S rDNA sequences from *Shewanella* strains. *Shewanella* sp. strain ANA-3 is boxed and in boldface type. The phylogenetic tree was constructed according to the distance criterion. The scale represents the number of substitutions per site. The percentage of 1,000 bootstrap replicates that supported the branching order is shown near the relevant nodes. Nodes without bootstrap values occurred <50%. Outgroups included *Alteromonas macleodii* (X82145) and *Pseudoalteromonas haloplanktis* (X67024). GenBank accession numbers for *Shewanella* species are given parenthetically as follows: *S. figidimarina* (U85903), *S. japonica* (Af145921), *S. algae* (U91546), *S. amazonensis* (AF005248), *Shewanella* sp. strain ANA-3 (AF136392), *S. baltica* (AJ000214), *S. putrefaciens* (U91550), *S. oneidensis* (AF005251), *Shewanella* sp. strain MR-8 (AF005254), *Shewanella* sp. strain MR-7 (AF005253), *Shewanella* sp. strain MR-4 (AF005252), *S. gelidimarina* (U85907), *S. pealeana* (AF011335), *S. benthica* (X82131), *S. violacea* (D21225), *S. hanedai* (U91590), *S. woodyi* (AF003549), and *S. colwelliana* (AF170794).

from total genomic DNA and electrophoresis of the genomic DNA on a 0.7% agarose gel did not show any distinguishable plasmid bands, this suggests that the *ars* genes are either chromosomally encoded or on a highly stable megaplasmid. In addition, Southern blot analysis of ANA-3 genomic DNA (hybridized with an ANA-3 *arsB*) probe detected the presence of only one copy of *arsB*.

Mutagenesis of *arsB* and the effects on As(V) respiration and As(III) resistance. Although the *ars* system located on pSALT1 is not sufficient to confer respiratory As(V) reduction in *S. oneidensis* MR-1, it might still be necessary for growth on As(V), especially if high concentrations of As(III) are generated inside the cytoplasm. Therefore, we constructed a mutation in the *arsB* homolog in ANA-3 to determine whether *arsB* is also required for respiratory growth on As(V). Figure 4A shows the position of the transposon insertion in *arsB* introduced into ANA-3 to generate the strain ARSB1. ARSB1 was unable to reduce As(V) to As(III) under aerobic conditions (data not shown). We suspected that the mutation in *arsB* was polar to the downstream gene *arsC*, predicted to encode a small 17-kDa cytosolic As(V) reductase. To test this, we used RT-PCR to assay for the presence of *arsC*-specific message in RNA extracted from cells grown in the presence of 1 mM As(V). There was no detectable amount of *arsC* message in the *arsB* mutant strain ARSB1, unlike in wild-type ANA-3 grown under the same conditions (Fig. 4C). Controls for expression of the *arsA* and 16S rDNA genes were positive in both ANA-3 and ARSB1, confirming that the absence of *arsC*-message was due to a polar affect of the *arsB* mutation.

When we tested ARSB1 and ANA-3 for their ability to respire on As(V) in low-phosphate medium (~0.3 mM P_i), a significant difference in the phenotype of ARSB1 was observed compared to the wild-type ANA-3 (Fig. 5A). The generation time and $Y_{lactate}$ value for ARSB1 were 5.3 h and 3.6 g of cells/mol, respectively, ~2-fold longer and ~3-fold less than for the wild type. The cell density of ARSB1 reached a maximum of 3.8×10^7 cells/ml, ca. 75% lower than wild-type ANA-3, when 7 mM of As(V) had been reduced to As(III). Lower phosphate concentrations did not appear to affect the growth of wild-type ANA-3 on As(V), evident in the $Y_{lactate}$

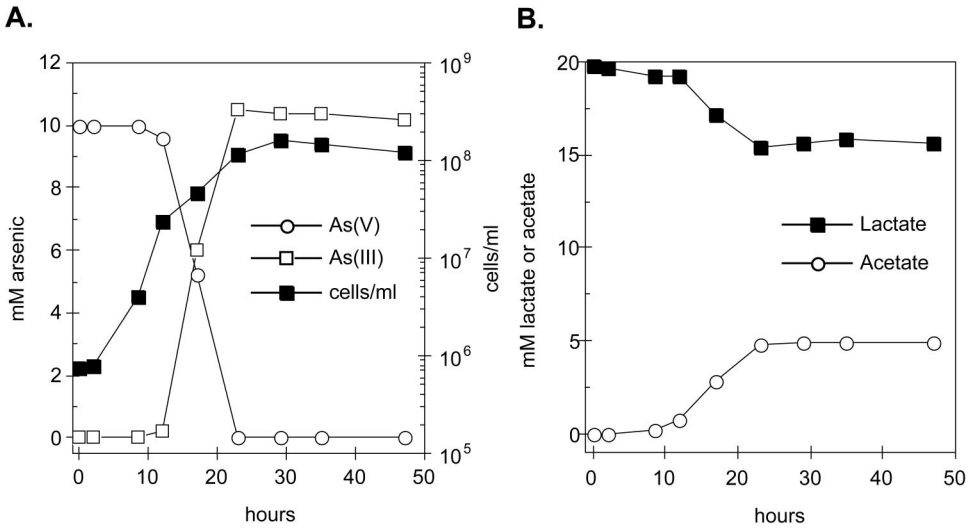


FIG. 2. (A) Respiratory arsenate reduction and growth of *Shewanella* sp. strain ANA-3 on lactate as the electron donor. (B) Oxidation of lactate and accumulation of acetate during respiration on arsenate. Data are representative of triplicate cultures.

(9.9 g of cells/mol) and generation time (2.8 h), which are similar to the values obtained when ANA-3 is grown in higher-phosphate medium (Fig. 2). When ARSB1 and ANA-3 were grown in higher-phosphate medium (~3 mM) and 5 mM As(V), no differences in As(V) respiration rates or growth rates were observed (data not shown).

When ARSB1 and wild-type ANA-3 were grown anaerobically on lactate and fumarate in the presence of increasing As(III) (Fig. 5B and C), the growth of ARSB1 was completely inhibited in 5 mM As(III) (Fig. 5C). However, ARSB1 could grow in 1 mM As(III) similar to the wild type (Fig. 5B). At 2.5 mM As(III) concentrations, the growth of the other *Shewanella* species listed in Table 1 was also inhibited.

DISCUSSION

The primary objective of the present study was to isolate and characterize a bacterium that would be useful for dissecting the molecular basis of respiratory As(V) reduction. Many As(V) reducers have been described physiologically, yet little progress has been made in identifying the gene(s) involved in As(V) respiration and the biochemical details of their protein products. This stems in part from the fact that the previous isolates are all strict anaerobes that have short lifetimes on the bench. Although it is possible to successfully establish genetic systems and do biochemical work in strict anaerobes (e.g., *Geobacter metallireducans* and *Desulfovibrio desulfuricans*) (11,

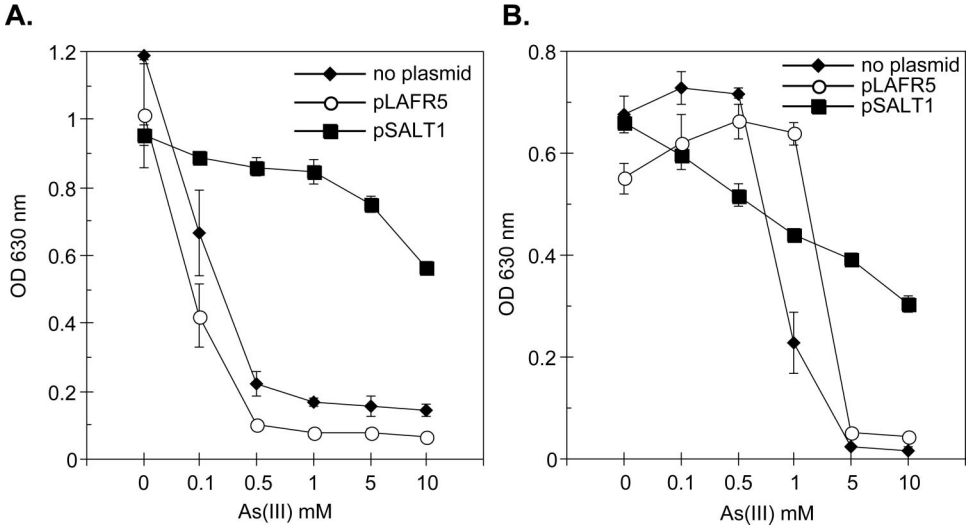


FIG. 3. The As^r cosmid pSALT1 confers As(III) resistance to *E. coli* AW3110 (A) and *S. oneidensis* MR-1 (B). Strains were grown aerobically in LB medium with the specified As(III) concentrations. Tetracycline was added at 15 μg/ml to strains harboring pSALT1 or the cosmid vector pLAFR5. The initial OD_{630} was <0.05 on average in all experiments. Values and error bars represent the averages and standard deviations of quadruplicate samples, respectively, after 24 h of incubation.

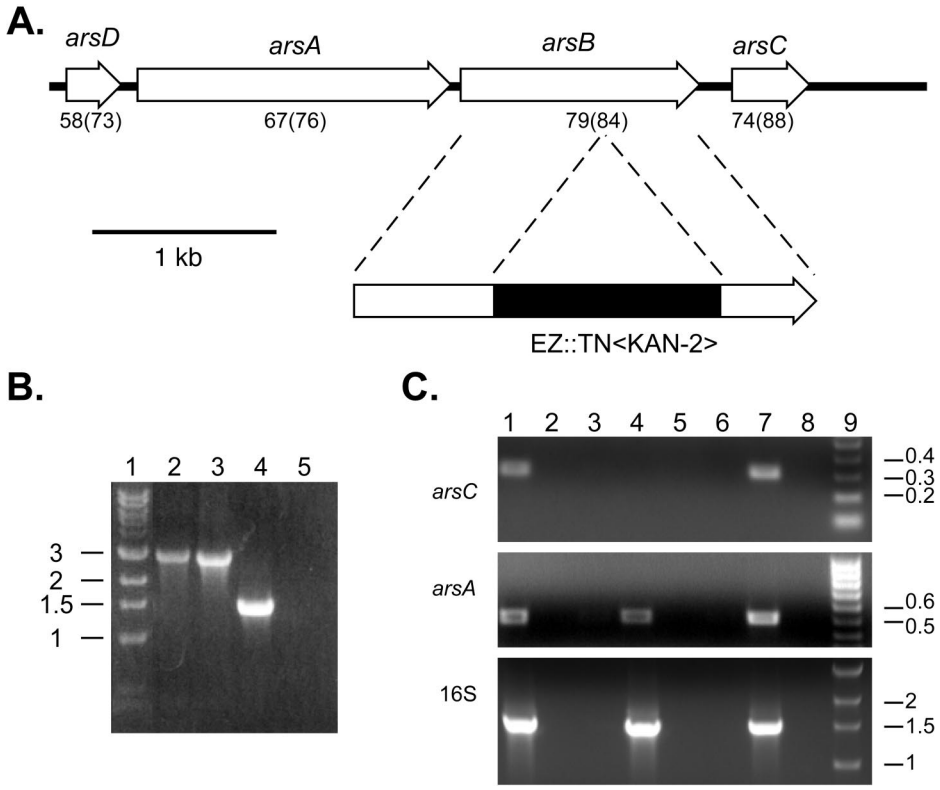


FIG. 4. (A) Map of the ANA-3 *ars* operon and location of the *arsB* mutation in ARSB1. The mutation in *arsB* with the EZ::TN<KAN-2> transposon is indicated by the black 1.2-kb size box. Numbers under the genes indicate the percent identity (similarity) to the corresponding R773 *ars* operon homolog. (B) Gel picture showing the results of PCR with primers TNARSBF and TNARSBR with genomic DNA of the ARSB1 strain (lane 2), pSALT1-B10 cosmid (lane 3), wild-type ANA-3 (lane 4), and a reagent negative control (lane 5). Lane 1 contains a 1-kb ladder. (C) RT-PCR analysis for the expression of *arsC*, *arsA*, and 16S rDNA genes after ANA-3 and ARSB1 were grown for 4 h in the presence of 1 mM As(V). Lanes 1 to 3 and lanes 4 to 6 correspond to ANA-3 and ARSB1, respectively. Lanes also correspond to RT with primer (lanes 1 and 4), no RT added (lanes 2 and 5), RT without primer (lanes 3 and 6), ANA-3 genomic DNA (PCR-positive control) (lane 7), water (PCR-negative control) (lane 8), and DNA ladders in kilobases (lane 9).

57), because ANA-3 can grow overnight aerobically on LB medium, exhibits robust anaerobic growth in minimal medium, and plates easily, it provides an attractive model system for molecular studies. Moreover, many of the existing genetic tools developed for *S. oneidensis* strain MR-1 can be adapted for use in ANA-3, and because ANA-3 is so closely related to *S. oneidensis* MR-1 the recent completion of the genome sequence for *S. oneidensis* MR-1 (16) aids its genetic analysis. To demonstrate the utility of ANA-3 as a model genetic system for studying As(V)-respiration, we selectively disrupted genes involved in arsenic resistance (the *ars* genes) and studied their impact on As(V) respiration.

An initial hint that ANA-3 contained an *ars* operon came from its robust growth on high concentrations of As(V). When grown on 5 mM As(V) with lactate in excess, ANA-3's molar growth yield on lactate (~10 g of cells/mol of lactate) was twice that previously reported for the As(V)-respiring strains *Sulfurospirillum barnesii* SES-3 (5.3 g of cells/mol of lactate) and *Desulfotomaculum auripigmentum* OREX-4 (5.6 g of cells/mol of lactate) (24, 37). At 10 mM As(V), ANA-3 grew as well as it did at 5 mM As(V), whereas the growth of SES-3 and OREX-4 was significantly impaired. This suggested that ANA-3's resistance to high concentrations of arsenic might be due to the presence of a high-level *ars* detoxification system, including an *arsA* gene, since expression of this system confers resistance to high concentrations of As (31). Additional physiological evidence in support of this was provided by the fact that ANA-3 could reduce As(V) when grown aerobically and exhibited resistance to 10 mM As(III).

Several new *ars* operons have recently been identified by using the As(III)-sensitive *E. coli ars* deletion strain AW3110. For example, the *ars* operons of *Pseudomonas fluorescens* MSP3 and *Thiobacillus ferrooxidans* were shown to confer resistance up to 2 mM As(III) when expressed in AW3110 (8, 43). To determine whether ANA-3 contained an *ars* operon as predicted, we tested whether DNA from ANA-3 could confer As(III) resistance to AW3110 and to the As(III)-sensitive *S. oneidensis* strain MR-1. Positive identification and sequencing of a cosmid that functionally rescued strains AW3110 and MR-1 in the presence of high As(III) concentrations confirmed the presence of four open reading frames with striking homology to *arsD*, *arsA*, *arsB*, and *arsC* of the *E. coli* R773 (Table 3). Although many genomes of sequenced microorganisms possess an As(III) efflux pump [namely, a homolog of the ArsB or ACR3, both encoding membrane-bound As(III) efflux channels], high-level resistance normally requires the addition of a large 63-kDa ATPase subunit, ArsA (44). The presence of

TABLE 3. Percent amino acid identity and similarity between the *ars* homologs of *Shewanella* sp. ANA-3 and those of *S. oneidensis* MR-1 and *E. coli* R773 arsenic resistance plasmid

Gene[a]	% Amino acid identity (% similarity)	
	S. oneidensis MR-1	*E. coli* R773 plasmid
arsD	NA[e]	58 (73)
arsA	NA	67 (76)
arsB	15 (32)[b]	79 (84)
arsC	14 (27)[c], 30 (46)[d]	74 (88)

[a] Putative homolog of *Shewanella* sp. strain ANA-3.
[b] TIGR accession no. SO0534.
[c] TIGR accession no. SO0533.
[d] TIGR accession no. SO2871.
[e] NA, not applicable.

an *arsA* homolog in the ANA-3 *ars* operon thus may explain ANA-3's resistance to 10 mM As(III).

ANA-3 is the first respiratory As(V) reducer shown to have an *ars* gene cluster, but identification of the *ars* genes in an As(V)-respiring organism is not surprising. Although Macy et al. (28) could not detect *ars* genes in the As(V) reducer *Desulfomicrobium* strain Ben-RB when using an *E. coli* R773 *arsC* probe, this could be due to sequence differences between the Ben-RB *ars* operon and the *E. coli* R773 operon. Indeed, this would be expected given the sequence diversity of the *ars* operon among different genera of As-resistant bacteria (47) and the fact that Ben-RB is phylogenetically distant from *E. coli*, whereas ANA-3 is more closely related. As more genetic work on different As(V)-respiring strains is performed, it will be interesting to see whether ANA-3 is exceptional with respect to its possession of an *arsA*-containing *ars* operon or representative of many As(V) respirers. How ANA-3 acquired this operon is an intriguing open question.

Knowing that ANA-3 possessed an *ars* operon, our next question became whether the *ars* detoxification system was required by ANA-3 to respire As(V). This was interesting for two reasons. First, we were curious as to whether the *ars* detoxification system conferred an advantage to cells respiring As(V). Second, we wanted to determine whether ArsC could account for As(V) reduction under conditions of As(V) respiration. The construction of an *arsB* mutant that was polar onto *arsC* (strain ARSB1) enabled us to consider both of these issues. ARSB1 did not grow in medium amended with >5 mM As(III), nor did ARSB1 reduce As(V) in LB medium-As(V) cultures. Interestingly, when respiring on As(V) in low-phosphate medium, ARSB1 reduced 7 to 8 mM As(V) at a slower rate compared to the wild type, achieving 25% lower cell density. These observations, along with the fact that no *arsC*-specific mRNA was detected by RT-PCR in ARSB1, suggest that there is an additional As(V) reductase that is used for As(V) respiration and that ArsC is not used for respiratory As(V) reduction.

There are several possible explanations for the As(III) resistance phenotype in ARSB1: (i) the presence of a duplicate *ars* operon compensates for the loss of this copy of *arsB* and *arsC*; (ii) the expression of an additional arsenic detoxification pathway during As(V) respiration compensates for the loss of this copy of *arsB* and *arsC*; and/or (iii) the enzyme used to reduce As(V) to As(III) during respiration resides in the periplasm and has a high affinity for As(V)—thus, the loss of this copy of *arsB* and *arsC* does not seriously affect the cell. Because Southern blot analysis with an *arsB* gene probe revealed only one hybridizing band within the genomic DNA of ANA-3, we believe the first explanation is unlikely. Although we do not yet have any direct evidence that either supports or rejects the second explanation, it seems possible, based on positive identification of homologs in the *S. oneidensis* MR-1 genome to genes involved in other As resistance systems (Table 4), that ANA-3 may also possess an additional As resistance system. However, because the *arsB* mutant could not reduce As(V) aerobically, it appears that a functional ArsC homolog is not made.

Regardless of whether an additional As resistance pathway exists and is expressed when ANA-3 is respiring As(V), we favor the third explanation. We base this position upon the fact that we recently identified a respiratory As(V) reductase whose coding sequence motifs suggest that it resides in the periplasm of ANA-3 (C. W. Saltikov and D. K. Newman, unpublished data). If this enzyme is able to scavenge As(V) faster than As(V) can enter the cytoplasm through inorganic phosphate (P_i) transporters (45) when the As(V)/P_i ratio is low, we would expect the need for a cytosolic As efflux system to be minimal. Conversely, when the As(V)/P_i ratio is high, we would expect more As(V) to enter the cell. Preliminary evidence in support of this interpretation is that the ARSB1 mutant exhibits a growth defect relative to the wild type when the As(V)/P_i ratio is high. Although more work is required to confirm this interpretation, including kinetic analyses of As(V) binding and/or turnover rates in the presence of various P_i, we favor it as a working hypothesis.

In summary, our results show that the ArsB efflux system is not required for ANA-3 to respire on As(V) and that ArsC is not required for ANA-3 to reduce As(V) to As(III) when respiring As(V). Nevertheless, the presence of the *ars* operon does appear to provide ANA-3 with additional protection against the toxicity of As(III) when respiring high concentrations of As(V). Whether or not As(V)-respiring microorganisms inhabiting natural systems require high-level As detoxification systems remains to be determined.

TABLE 4. Percent amino acid identity and similarity of the predicted arsenic resistance proteins of *S. oneidensis* MR-1 to the closest known proteins in the GenBank database

Putative protein	TIGR accession no.	Closest match in GenBank
ArsR	SO0532	58% identity and 70% similarity to *Pseudomonas aeruginosa* PAO1 ArsR; NP_250967
ACR3	SO0534	49% identity and 62% similarity to *Pyrococcus furiosus* DSM 3638 ACR3; NP_578281
ArsC	SO0533	50% identity and 62% similarity to *Pseudomonas aeruginosa* PAO1 ArsC; NP_250969
ArsC	SO2871	58% identity and 74% similarity to *Pseudomonas aeruginosa* PAO1 ArsC; NP_249641

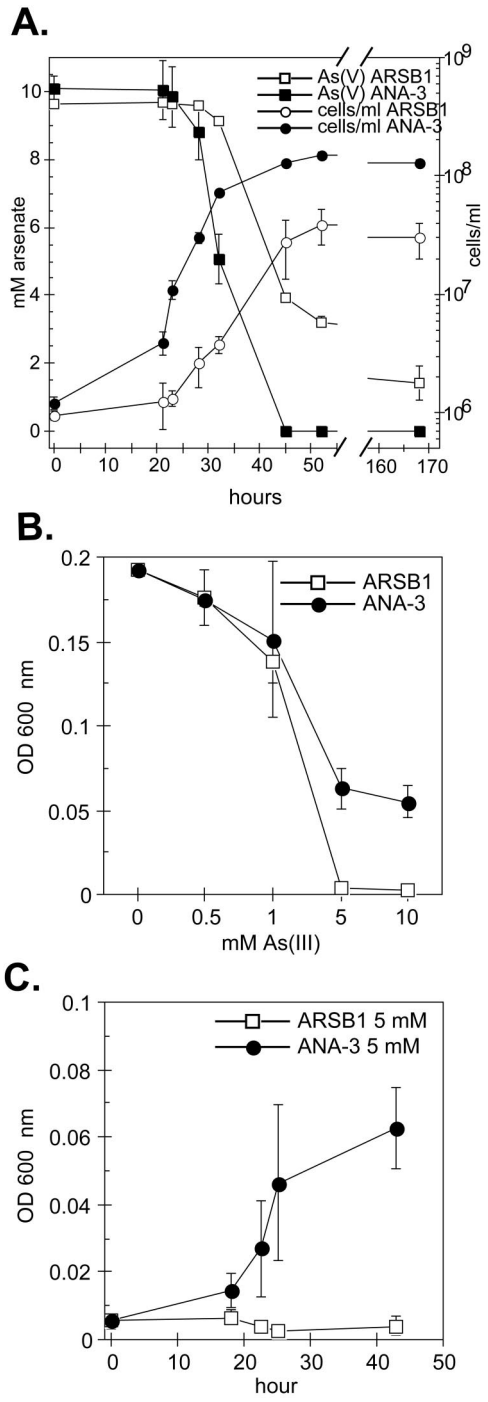


FIG. 5. Anaerobic As(V) respiration with lactate as the sole carbon source and electron donor (A), resistance profile to increasing As(III) concentrations (B), and time course for growth in 5 mM As(III) (C) for ANA-3 and ARSB1. In panels B and C, both strains were grown anaerobically on lactate and fumarate with As(III) added at the specified concentrations, and the initial OD_{600} values were similar to that of the blank medium (i.e., 0.005). The values and error bars in all three panels represent the averages and ranges of duplicate samples, respectively.

ACKNOWLEDGMENTS

We thank Joanna Levitt and Anabel Anton for help with the isolation of ANA-3 during the 1998 MBL Microbial Diversity Course, Doug Lies and members of the Newman lab for valuable discussions, Angela Snow for laboratory assistance, and B. P. Rosen for providing *E. coli* strain AW3110.

Funding was provided by grants from the Luce Foundation and the Packard Foundation to D.K.N. and by a National Science Foundation Postdoctoral Fellowship in Microbial Biology to C.W.S.

REFERENCES

1. **Ahmann, D., L. R. Krumholz, H. F. Hemond, D. R. Lovley, and F. M. Morel.** 1997. Microbial mobilization of arsenic from sediments of the Aberjona watershed. Environ. Sci. Technol. **31:**2923–2930.
2. **Armienta, M. A., G. Villasenor, R. Rodriguez, L. K. Ongley, and H. Mango.** 2001. The role of arsenic-bearing rocks in groundwater pollution at Zimapan Valley, Mexico. Environ. Geol. **40:**571–581.
3. **Ausubel, F. M.** 1992. Short protocols in molecular biology, 2nd ed.: a compendium of methods from current protocols in molecular biology. John Wiley & Sons, Inc./Green Publishing Associates, New York, N.Y.
4. **Beliaev, A. S., and D. A. Saffarini.** 1998. *Shewanella Putrefaciens mtrB* encodes an outer membrane protein required for Fe(III) and Mn(IV) reduction. J. Bacteriol. **180:**6292–6297.
5. **Bentley, R., and T. G. Chasteen.** 2002. Microbial methylation of metalloids: arsenic, antimony, and bismuth. Microbiol. Mol. Biol. Rev. **66:**250–274.
6. **Blum, J. S., A. B. Bindi, J. Buzzelli, J. F. Stolz, and R. S. Oremland.** 1998. *Bacillus arsenicoselenatis*, sp. nov., and *Bacillus selenitireducens*, sp. nov.: two haloalkaliphiles from Mono Lake, California, that respire oxyanions of selenium and arsenic. Arch. Microbiol. **171:**19–30.
7. **Bowman, J. P., S. A. Mccammon, D. S. Nichols, J. H. Skerratt, S. M. Rea, P. D. Nichols, and T. A. Mcmeekin.** 1997. *Shewanella gelidimarina* sp. nov. and *Shewanella frigidimarina* sp. nov., novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5 Omega 3) and grow anaerobically by dissimilatory Fe(III) reduction. Int. J. Syst. Bacteriol. **47:**1040–1047.
8. **Butcher, B. G., S. M. Deane, and D. E. Rawlings.** 2000. The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. Appl. Environ. Microbiol. **66:**1826–1833.
9. **Carlin, A., W. Shi, S. Dey, and B. P. Rosen.** 1995. The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. J. Bacteriol. **177:**981–986.
10. **Cline, E.** 1969. Spectrophotometric determination of hydrogen sulfide in natural waters. Limnol. Oceanogr. **14:**454–458.
11. **Coppi, M. V., C. Leang, S. J. Sandler, and D. R. Lovley.** 2001. Development of a genetic system for *Geobacter sulfurreducens*. Appl. Environ. Microbiol. **67:**3180–3187.
12. **Dehio, C., and M. Meyer.** 1997. Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn*5* in *Bartonella henselae* following conjugal plasmid transfer from *Escherichia coli*. J. Bacteriol. **179:**538–540.
13. **Dowdle, P. R., A. M. Laverman, and R. S. Oremland.** 1996. Bacterial dissimilatory reduction of arsenic(V) to arsenic(III) in anoxic sediments. Appl. Environ. Microbiol. **62:**1664–1669.
14. **Fredrickson, J. K., J. M. Zachara, D. W. Kennedy, H. L. Dong, T. C. Onstott, N. W. Hinman, and S. M. Li.** 1998. Biogenic iron mineralization accompanying the dissimilatory reduction of hydrous ferric oxide by a groundwater bacterium. Geochim. Cosmochim. Acta **62:**3239–3257.
15. **Gihring, T. M., and J. F. Banfield.** 2001. Arsenite oxidation and arsenate respiration by a new *Thermus* isolate. FEMS Microbiol. Lett. **204:**335–340.
16. **Heidelberg, J. F., I. T. Paulsen, K. E. Nelson, E. J. Gaidos, W. C. Nelson, T. D. Read, J. A. Eisen, R. Seshadri, N. Ward, B. Methe, R. A. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. T. DeBoy, R. J. Dodson, A. S. Durkin, D. H. Haft, J. F. Kolonay, R. Madupu, Peterson, J. D., L. A. Umayam, O. White, A. M. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. R. Utterback, L. A. McDonald, T. V. Feldblyum, H. O. Smith, J. C. Venter, K. H. Nealson, and C. M. Fraser.** 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. Nature **20:**1118–1123.
17. **Herbel, M. J., J. S. Blum, S. E. Hoeft, S. M. Cohen, L. L. Arnold, J. Lisak, J. F. Stolz, and R. S. Oremland.** 2002. Dissimilatory arsenate reductase activity and arsenate-respiring bacteria in bovine rumen fluid, hamster feces, and the termite hindgut. FEMS Microbiol. Ecol. **41:**59–67.
18. **Holmes, B., S. P. Lapage, and H. Malnick.** 1975. Strains of *Pseudomonas-Putrefaciens* from clinical material. J. Clin. Pathol. **28:**149–155.
19. **Huber, R., M. Sacher, A. Vollmann, H. Huber, and D. Rose.** 2000. Respiration of arsenate and selenate by hyperthermophilic *Archaea*. Syst. Appl. Microbiol. **23:**305–314.
20. **Johnson, D. L., and M. E. Q. Pilson.** 1972. Spectrophotometric determination of arsenite, arsenate, and phosphate in natural waters. Anal. Chim. Acta **58:**289–299.
21. **Johnson, J. L.** 1981. Genetic characterization, p. 450–472. *In* P. Gerhardt, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg, and G. B. Phillips (ed.), Manual of methods for general bacteriology. American Society for Microbiology, Washington, D.C.

22. **Keen, N. T., S. Tamaki, D. Kobayashi, and D. Trollinger.** 1988. Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria. Gene **70:**191–197.
23. **Krafft, T., and J. M. Macy.** 1998. Purification and characterization of the respiratory arsenate reductase of *Chrysiogenes arsenatis*. Eur. J. Biochem. **255:**647–653.
24. **Laverman, A. M., J. S. Blum, J. K. Schaefer, E. J. P. Phillips, D. R. Lovely, and R. S. Oremland.** 1995. Growth of strain SES-3 with arsenate and other diverse electron acceptors. Appl. Environ. Microbiol. **61:**3556–3561.
25. **Leonardo, M. R., D. P. Moser, E. Barbieri, C. A. Brantner, B. J. Macgregor, B. J. Paster, E. Stackebrandt, and K. H. Nealson.** 1999. *Shewanella pealeana* sp. nov., a member of the microbial community associated with the accessory nidamental gland of the squid *Loligo pealei*. Int. J. Syst. Bacteriol. **49:**1341–1351.
26. **Lovley, D. R., and E. J. P. Phillips.** 1988. Novel mode of microbial energy-metabolism–organic-carbon oxidation coupled to dissimilatory reduction of iron or manganese. Appl. Environ. Microbiol. **54:**1472–1480.
27. **Macy, J. M., K. Nunan, K. D. Hagen, D. R. Dixon, P. J. Harbour, M. Cahill, and L. I. Sly.** 1996. *Chrysiogenes arsenatis* gen. nov., sp. nov., a new arsenate-respiring bacterium isolated from gold mine wastewater. Int. J. Syst. Bacteriol. **46:**1153–1157.
28. **Macy, J. M., J. M. Santini, B. V. Pauling, A. H. O'Neill, and L. I. Sly.** 2000. Two new arsenate/sulfate-reducing bacteria: mechanisms of arsenate reduction. Arch. Microbiol. **173:**49–57.
29. **Makemson, J. C., N. R. Fulayfil, W. Landry, L. M. Vanert, C. F. Wimpee, E. A. Widder, and J. F. Case.** 1997. *Shewanella woodyi* sp. nov., an exclusively respiratory luminous bacterium isolated from the Alboran Sea. Int. J. Syst. Bacteriol. **47:**1034–1039.
30. **Miller, T. L., and M. J. Wolin.** 1974. A serum bottle modification of the Hungate technique for cultivating obligate anaerobes. Appl. Environ. Microbiol. **27:**985–987.
31. **Mukhopadhyay, R., B. P. Rosen, L. Phung, and S. Silver.** 2002. Microbial arsenic: from geocycles to genes and enzymes. FEMS Microbiol. Rev. **26:**311–321.
32. **Myers, C. R., and K. H. Nealson.** 1988. Bacterial manganese reduction and growth with manganese oxide as the sole electron-acceptor. Science **240:**1319–1321.
33. **Nealson, K. H., C. R. Myers, and B. B. Wimpee.** 1991. Isolation and identification of manganese-reducing bacteria and estimates of microbial Mn(IV)-reducing potential in the Black Sea. Deep-Sea Res. **38**(Suppl. 2)**:**S907–S920.
34. **Neidhardt, F. C., and H. E. Umbarger.** 1991. Chemical composition of *Escherichia coli*, p. 13–16. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
35. **Newman, D. K., D. Ahmann, and F. M. M. Morel.** 1998. A brief review of microbial arsenate respiration. Geomicrobiol. J. **15:**255–268.
36. **Newman, D. K., T. J. Beveridge, and F. M. M. Morel.** 1997. Precipitation of arsenic trisulfide by *Desulfotomaculum auripigmentum*. Appl. Environ. Microbiol. **63:**2022–2028.
37. **Newman, D. K., E. K. Kennedy, J. D. Coates, D. Ahmann, D. J. Ellis, D. R. Lovley, and F. M. Morel.** 1997. Dissimilatory arsenate and sulfate reduction in *Desulfotomaculum auripigmentum* sp. nov. Arch. Microbiol. **168:**380–388.
38. **Newman, D. K., and R. Kolter.** 2000. A role for excreted quinones in extracellular electron transfer. Nature **405:**94–97.
39. **Niggemyer, A., S. Spring, E. Stackebrandt, and R. F. Rosenzweig.** 2001. Isolation and characterization of a novel As(V)-reducing bacterium: implications for arsenic mobilization and the genus *Desulfitobacterium*. Appl. Environ. Microbiol. **67:**5568–5580.
40. **Nriagu, J.** 2002. Arsenic poisoning through the ages, p. 21–22. *In* W. T. Frankenberger (ed.), Environmental chemistry of arsenic. Marcel Dekker, Inc., New York, N.Y.
41. **Oremland, R. S., J. S. Blum, C. W. Culbertson, P. T. Visscher, L. G. Miller, P. R. Dowdle, and F. E. Strohmaier.** 1994. Isolation, growth, and metabolism of an obligately anaerobic, selenate-respiring bacterium, strain SES-3. Appl. Environ. Microbiol. **60:**3011–3019.
42. **Oremland, R. S., P. R. Dowdle, S. Hoeft, J. O. Sharp, J. K. Schaefer, L. G. Miller, J. S. Blum, R. L. Smith, N. S. Bloom, and D. Wallschlaeger.** 2000. Bacterial dissimilatory reduction of arsenate and sulfate in meromictic Mono Lake, California. Geochim. Cosmochim. Acta **64:**3073–3084.
43. **Prithivirajsingh, S., S. K. Mishra, and A. Mahadevan.** 2001. Detection and analysis of chromosomal arsenic resistance in *Pseudomonas fluorescens* strain MSP3. Biochem. Biophys. Res. Commun. **280:**1393–1401.
44. **Rosen, B. P.** 1999. Families of arsenic transporters. Trends. Microbiol. **7:**207–212.
45. **Rosenberg, H., L. M. Russell, P. A. Jacomb, and K. Chegwidden.** 1982. Phosphate exchange in the *pit* transport system in *Escherichia coli*. J. Bacteriol. **149:**123–130.
46. **Ruimy, R., V. Breittmayer, P. Elbaze, B. Lafay, O. Boussemart, M. Gauthier, and R. Christen.** 1994. Phylogenetic analysis and assessment of the genera *Vibrio*, *Photobacterium*, *Aeromonas*, and *Plesiomonas* deduced from small-subunit ribosomal-RNA sequences. Int. J. Syst. Bacteriol. **44:**416–426.
47. **Saltikov, C. W., and B. H. Olson.** 2002. Homology of *Escherichia coli* R773 *arsA*, *arsB*, and *arsC* genes in arsenic-resistant bacteria isolated from raw sewage and arsenic-enriched creek waters. Appl. Environ. Microbiol. **68:**280–288.
48. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
49. **Santini, J. M., L. I. Sly, A. M. Wen, D. Comrie, P. De Wulf-Durand, and J. M. Macy.** 2002. New arsenite-oxidizing bacteria isolated from Australian gold mining environments: phylogenetic relationships. Geomicrobiol. J. **19:**67–76.
50. **Silver, S., L. T. Phung, and B. P. Rosen.** 2002. Arsenic metabolism: resistance, reduction, and oxidation, p. 254. In W. T. Frankenberger (ed.), Environmental chemistry of arsenic. Marcel Dekker, Inc., New York, N.Y.
51. **Simidu, U., K. Kitatsukamoto, T. Yasumoto, and M. Yotsu.** 1990. Taxonomy of four marine bacterial strains that produce tetrodotoxin. Int. J. Syst. Bacteriol. **40:**331–336.
52. **Smibert, R. M., and N. Krieg.** 1994. Phenotypic characterization, p. 607–654. *In* P. Gerhardt, R. G. E. Murray, W. A. Wood, and N. Krieg (ed.), Methods for general and molecular bacteriology. American Society for Microbiology, Washington, D.C.
53. **Stolz, J. F., and R. S. Oremland.** 1999. Bacterial respiration of arsenic and selenium. FEMS Microbiol. Rev. **23:**615–627.
54. **Swofford, D. L.** 1999. PAUP: phylogenetic analysis using parsimony (and other methods), version 4.0.b10. Sinauer Associates, Sunderland, Mass.
55. **Venkateswaran, K., M. E. Dollhopf, R. Aller, E. Stackebrandt, and K. H. Nealson.** 1998. *Shewanella amazonensis* sp. nov., a novel metal-reducing facultative anaerobe from Amazonian shelf muds. Int. J. Syst. Bacteriol. **48:**965–972.
56. **Venkateswaran, K., D. P. Moser, M. E. Dollhopf, D. P. Lies, D. A. Saffarini, B. J. MacGregor, D. B. Ringelberg, D. C. White, M. Nishijima, H. Sano, J. Burghardt, E. Stackebrandt, and K. H. Nealson.** 1999. Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. Int. J. Syst. Bacteriol. **49:**705–724.
57. **Wall, J. D., T. Murnan, J. Argyle, R. S. English, and B. J. Rapp-Giles.** 1996. Transposon mutagenesis in *Desulfovibrio desulfuricans*: development of a random mutagenesis tool from Tn*7*. Appl. Environ. Microbiol. **62:**3762–3767.
58. **Wilkie, J. A., and J. G. Hering.** 1998. Rapid oxidation of geothermal arsenic(III) in streamwaters of the Eastern Sierra Nevada. Environ. Sci. Technol. **32:**657–662.
59. **Zachara, J. M., J. K. Fredrickson, S. M. Li, D. W. Kennedy, S. C. Smith, and P. L. Gassman.** 1998. Bacterial reduction of crystalline Fe^{3+} oxides in single phase suspensions and subsurface materials. Am. Mineral. **83:**1426–1443.
60. **Ziemke, F., M. G. Hofle, J. Lalucat, and R. Rossello-Mora.** 1998. Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. Int. J. Syst. Bacteriol. **48:**179–186.

Genetic identification of a respiratory arsenate reductase

Chad W. Saltikov and Dianne K. Newman†

Department of Geological and Planetary Sciences, California Institute of Technology, Mailstop 100-23, Pasadena, CA 91125

Communicated by Douglas C. Rees, California Institute of Technology, Pasadena, CA, July 14, 2003 (received for review May 1, 2003)

For more than a decade, it has been recognized that arsenate [$H_2AsO_4^{1-}$; As(V)] can be used by microorganisms as a terminal electron acceptor in anaerobic respiration. Given the toxicity of arsenic, the mechanistic basis of this process is intriguing, as is its evolutionary origin. Here we show that a two-gene cluster (*arrAB*; *a*rsenate *r*espiratory *r*eduction) in the bacterium *Shewanella* sp. strain ANA-3 specifically confers respiratory As(V) reductase activity. Mutants with in-frame deletions of either *arrA* or *arrB* are incapable of growing on As(V), yet both are able to grow on a wide variety of other electron acceptors as efficiently as the wild-type. Complementation by the wild-type sequence rescues the mutants' ability to respire As(V). *arrA* is predicted to encode a 95.2-kDa protein with sequence motifs similar to the molybdenum containing enzymes of the dimethyl sulfoxide reductase family. *arrB* is predicted to encode a 25.7-kDa iron–sulfur protein. *arrA* and *arrB* comprise an operon that contains a twin arginine translocation (Tat) motif in ArrA (but not in ArrB) as well as a putative anaerobic transcription factor binding site upstream of *arrA*, suggesting that the respiratory As(V) reductase is exported to the periplasm via the Tat pathway and under anaerobic transcriptional control. These genes appear to define a new class of reductases that are specific for respiratory As(V) reduction.

The consumption of arsenic (As)-tainted surface waters and ground waters has created a public health crisis in many countries (1, 2). Although much of the As contamination derives from natural weathering and dissolution of As-bearing minerals, recognition that microorganisms can alter the mobility of As in natural waters through redox transformations (3) drove the discovery of the first arsenate [As(V)]-respiring bacterium nearly a decade ago (4). Since then, many more microorganisms that can reduce As(V) to arsenite [H_3AsO_3, As(III)] have been discovered (5–8), but a mechanistic understanding of this metabolism has lagged. To date, only three studies have described the biochemistry of arsenate respiration (5, 9, 10), and detailed biochemical analyses have not been performed. In part, the limitations of these studies can be attributed to the fact that the As(V)-respiring organisms being studied were not genetically tractable. To quantify the geochemical impact of As(V)-respiring microorganisms in a given locale, we must be able to predict when these organisms will be active, and how rapidly they will transform As(V). Identification of the gene(s) that control this process, elucidation of their regulation, and determination of the kinetics of their protein products, are necessary steps toward understanding the specific contribution of As(V)-respiring bacteria to As-cycling in the environment.

In response to this need, a new As(V)-respiring species, *Shewanella* strain ANA-3 that is amenable to genetic analysis, was recently isolated (8). This organism contains two systems for reducing As(V). One is similar to the well conserved *ars* detoxification system from *Escherichia coli* plasmid R773 (11) and is advantageous, but not required, for respiratory As(V) reduction (8); a separate system appears to be required for As(V) respiration. In this report, we use genetic analysis to identify and describe the operon that encodes that latter system. Here we identify specific genes that are required for respiratory As(V) reduction, which opens the door for detailed biochemical study of their gene products. Moreover, the gene sequences we report represent a previously unrecognized class of respiratory reductases.

Materials and Methods

Strains and Plasmids. Bacterial strains of *E. coli* and *Shewanella* used in this study are listed in Table 1. Plasmids that were used or constructed are described in Table 1.

Growth Conditions. LB medium (12) was used for routine culturing of ANA-3 and *E. coli* strains. Anaerobic growth medium for strains of ANA-3 consisted of the following minimal medium: 0.225 g/liter K_2HPO_4/0.225 g/liter KH_2PO_4/0.46 g/liter NaCl/0.225 g/liter $(NH_4)_2SO_4$/0.117 g/liter $MgSO_4 \cdot 7H_2O$/4.2 g/liter $NaHCO_3$. The medium was supplemented with 20 mM sodium lactate, 10 mM Na_2HAsO_4, and 5 ml/liter each of trace element and vitamin solutions (8).

Identification of the Respiratory As(V) Reductase System. A partial *arrA* gene (1.2 kb) was identified on a previously isolated cosmid clone generated from genomic DNA of ANA-3 (8). Inverse PCR was used to isolate and sequence the downstream region of the partial *arrA* sequence. Genomic DNA (1 μg) was digested in a 50-μl volume with *Hin*dIII or *Bam*HI. After heat inactivation of the enzymes at 65°C for 20 min, intramolecular ligation reactions were performed: ≈400 ng of DNA was diluted to a concentration of ≈2 ng/μl and ligated with 2 μl of T4 DNA ligase (New England Biolabs) at 16°C overnight. After heat inactivation at 65°C for 15 min, reactions were adjusted to 500 μl with sterile water, applied to a Nanosep 30K centrifugal device (Pall Gelman Laboratory), and concentrated to ≈40 μl according to the manufacturer's instructions. PCR was performed with 2 μl of concentrated ligation reactions as follows: 200 nM of each inverse primer ARR-R1 (5′-CACCATTCGACTAAACCGTAGG-3′) and ARR-L1 (5′-GCGAAAGCCTATATGGATGAGA-3′), 200 μM dNTP mix, 2.5 units of *Pfu*Turbo DNA polymerase (Stratagene), 1× *Pfu*Turbo reaction buffer in a 50-μl reaction volume. The PCR products were purified by using a PCR cleanup kit (Qiagen, Valencia, CA) and sequenced directly by primer walking. Sequence data were assembled by using TIGR ASSEMBLER (www.tigr.org/software/assembler), and ORFs were detected by using ORFINDER (www.ncbi.nlm.nih.gov/gorf/gorf.html). The predicted translation of each ORF was compared with the sequences within GenBank by using BLAST (www.ncbi.nlm.nih.gov/blast).

Mutagenesis. To determine whether *arrA* and *arrB* were required for As(V) respiration, two mutant strains of ANA-3 were made (ARRA3 and ARRB1) that contained in-frame deletions of the respective genes. Vectors pΔ*arrA*3 and pΔ*arrB*1 were con-

Abbreviations: As, arsenic; As(V), arsenate; As(III), arsenite; TMAO, trimethylamine-*N*-oxide.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AY271310).

†To whom correspondence should be addressed. E-mail: dkn@caltech.edu.

MICROBIOLOGY

Table 1. Strains and plasmids used in this study

Strain or plasmid	Genotype or markers; characteristics and uses	Source or ref.
E. coli strains		
DH10β	Host for *E. coli* cloning; F-*mcrA* Δ*(mrr-hsdRMS-mcrBC)* Φ*80dlacZ*Δ*M15* Δ*(codB-lacI)3 deoR recA1 endA1 araD139* Δ*(ara-leu)7697 galU galK* λ*-rpsL(*StrR*)*	Life Technology
UQ950	*E. coli* DH5α λ(pir) host for cloning; F-Δ*(argF-lac)169* Φ*80dlacZ58(*ΔM15) *glnV44*(AS) *rfbD1 gyrA96*(NalR) *recA1 endA1 spoT1 thi-1 hsdR17 deoR* λ*pir+*	D. Lies, Caltech
WM3064	Donor strain for conjugation: *thrB*1004 *pro thi rpsL hsdS lacZ*ΔM15 RP4–1360 Δ(*araBAD)567* Δ*dapA*1341::[*erm pir*(wt)]	W. Metcalf, Univ. of Illinois, Urbana
Shewanella strains		
ANA-3	Isolated from an As-treated wooden pier pilling in a brackish estuary (Eel Pond, Woods Hole, MA); Contains *arsDABC*; respires on As(V); resistant to > 5 mM arsenite.	8
ARRA3	*Shewanella* sp. str. ANA-3, Δ*arrA*3; does not respire As(V)	This study
ARRB1	*Shewanella* sp. str. ANA-3, Δ*arrB*1; does not respire As(V)	This study
Plasmids/vectors		
pSALT1	pLAFR5-based 45 kb As(III)^r cosmid from ANA-3 genomic DNA library; contains *arsDABC*, Tc^r, confers resistance to As(III) and As(V).	8
pSMV10	9.1-kb mobilizable suicide vector; *oriR6K*, *mobRP4*, *sacB*, Km^r Gm^r	D. Lies, Caltech
pΔ*arrA3*	2-kb fusion PCR fragment containing Δ*arrA* cloned into the *Spe*I site of pSMV10; used to make the ARRA3 Δ*arrA* strain.	This study
pΔ*arrBI*	2-kb fusion PCR fragment containing Δ*arrB* cloned into the *Spe*I site of pSMV10; used to make the ARRBI Δ*arrB* strain.	This study
pBBR1MCS-2	5.1-kb broad-host range plasmid: Km^r, *lacZ*	15
*parrA*4	*arrA4* PCR fragment, including the promoter region, cloned into the *Spe*I site of pBBR1MCS-2	This study
*parrB*1	*arrB1* PCR fragment cloned into the *Xho*I/*Spe*I sites of pBBR1MCS-2	This study

structed by a cross-over PCR technique (13). Fusion PCR products (with *Spe*I ends, underlined), generated by using *Pfu*-Turbo (Stratagene), were prepared in two sequential PCRs that involved amplification of 1-kb regions up- and downstream of *arrA* with primers: XARR-A-A (5′-GGACTAGTGTGAGTCCAGCAACGCTAT), XARR-A-B (5′-CCCATCCAGCATGCTTAAACAGACTTGATTCTCTTTCTTCATTTTC), XARR-A-C (5′-TGTTTAAGCATGCTGGATGGGCGTGTTGAGAAAGTGTGAGGT), and XARR-A-D (5′-GGACTAGTATGACTTGATCCCTGAAATTG). For *arrB* the primers were: XARR-B-A (5′-GGACTAGTATTCTATCGGTAATGGTGTCG), XARR-B-B (5′-CCCATCCAGCATGCTTAAACACATTCCTAATCTCATAGTGGG), XARR-B-C (5′-TGTTTAAGCATGCTGGATGGGAAAACCGCTTATTAATCATAGAGG), and XARR-B-D (5′-GGACTAGTGCATAGCAACAGCAACCTTT). The two 1-kb flanking PCR products for each gene deletion were purified by using a Qiagen PCR clean-up column and mixed together to perform a second PCR with primers XARR-A-A and XARR-A-D (for *arrA*), or XARR-B-A and XARR-B-D (for *arrB*). The resulting 2-kb fusion products were digested with *Spe*I and ligated into the *Spe*I site of pSMV10, and transformed into a DH5α *λpir* strain (UQ950). The resulting mutagenesis vectors were transformed into the plasmid donor strain, *E. coli* WM3064 (a *dap* auxotroph, derivative of strain B2155; ref. 14), grown on LB + 0.3 mM diaminopimelic acid (DAP), and transferred to ANA-3 by conjugation (2:1 donor/recipient ratio; 5 h mating on a LB/DAP plate at 30°C). ANA-3 transconjugants containing the integrated mutagenesis vector were selected on LB agar plates supplemented with 50 μg/ml kanamycin (8). Kanamycin-resistant colonies occurred at frequencies ranging from 10^{-8} to 10^{-9} per recipient. These colonies were grown in the absence of antibiotic and subcultured once into LB. After overnight growth at 30°C, dilutions were plated on LB/10% sucrose (to select for loss of the *sacB* gene on the integrated mutagenesis vector), dried in a laminar flow hood, and incubated at 30°C overnight. Plates were replica printed onto LB and LB/kanamycin agar. Kanamycin-sensitive colonies were screened by PCR for the deletion of *arrA* or *arrB*.

Plasmids p*arrA*4 and p*arrB*1 were constructed to complement the ARRA3 and ARRB1 *arr* deletion mutants. PCR was used to generate *arrA*4 [*Spe*I ArrA-F1 (5′-GGACTAGTGAATAGGAGGCGATAAATGGAG), *Spe*I ArrA-R1 (5′-GGACTAGTCCTAATCTCATAGTGGGTTACCTC)] and *arrB*1 [*Xho*I ArrB-F1 (5′-GGCTCGAGGAAAGTGTGAGGTAACCCA), *Spe*I ArrB-R1 (5′-GGACTAGTCTCTATGATTAATAAGCGGT)] gene fragments, which were then ligated to the broad-host range plasmid pBBR1MCS-2 (15) to generate the complementation vectors. The appropriate vector was mated into the ANA-3 *arr* deletion strains by using *E. coli* WM3064 as the donor; transconjugants were selected by plating on LB agar supplemented with 50 μg/ml kanamycin without diaminopimelic acid. Kanamycin-resistant colonies were streak-purified several times on LB agar plates containing kanamycin.

A microtiter dish assay was used to test whether the *arr* deletion strains could grow with other terminal electron acceptors including fumarate (20 mM), nitrate (20 mM), thiosulfate (20 mM), trimethylamine-*N*-oxide (TMAO) (10 mM), poorly crystalline manganese oxides and iron hydroxides (20 mM) (8), and 2,6-anthraquinonedisulfonate (5 mM). Approximately 10^8 cells of overnight LB-grown cultures were washed two times in 1× PBS (12). A total of 200 μl of anaerobic media for each electron donor were inoculated in quadruplicate wells of a microtiter plate at 10^6 cells per ml. Microtiter plates were incubated at 30°C in an anaerobic chamber (15% CO_2/80% N_2/5% H_2). Growth was monitored over several days by observing the increase in the optical density at 600 nm. For metal oxides, utilization of the electron acceptors was inferred by observing the darkening (from red to black) of the iron oxide or the clearing (from black to clear) of the manganese oxide in the plate. Controls without electron donors and/or cells were also included.

Time Course of Respiratory Growth on Arsenate. Strains of ANA-3 were grown in LB or LB/kanamycin overnight at 30°C. Cells were washed two times in 1× PBS and resuspended to a final density of ≈10^8 cells per ml. Washed cells were inoculated into 10 ml of As(V)/lactate minimal salts medium at a cell density of

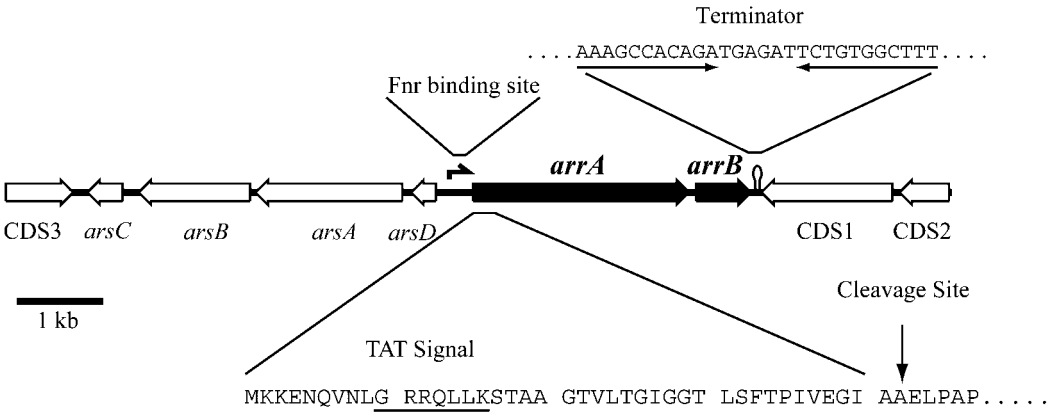


Fig. 1. Molecular organization of the *arrAB* gene cluster. The *arrA* gene is located upstream in the opposite orientation to an *ars* operon. CDS1 and CDS2 are similar to glutathione synthetases and a *S. oneidensis* MR-1 conserved hypothetical protein, respectively. The arrow between *arsD* and *arrA* indicates the location of potential Fnr-like binding sites. The loop after *arrB* is a putative transcriptional terminator region. Analysis of the putative Tat signal sequence within the first 42 amino acid residues of ArrA reveals a potential cleavage site (arrow).

$\approx 10^6$ per ml and incubated at 30°C. Cultures were sampled periodically and analyzed for cell growth, As(V) reduction, and lactate oxidation.

Phylogenetic Analysis. Sequences were aligned with CLUSTALW 1.82 (16). Phylogenetic analyses were performed by using the program PAUP* 4.0b10 (17). The distance criterion was used to construct unrooted neighbor-joining trees. After ignoring gaps in the multisequence alignments for ArrA and ArrB, the final number of amino acids included in the phylogenies were 510 and 144, respectively.

Analytical Techniques. As(V), As(III), lactate, and acetate were quantified by high-performance liquid chromatography (HPLC) as described (8). Changes in culture optical density were monitored at 600 nm.

Results

Identification of the *arrAB* Gene Cluster. Previously, we identified a region of DNA that encoded an arsenic resistance operon, *arsDABC*, from a genomic library of ANA-3 (8). Sequencing upstream of this cosmid clone (pSALT1) revealed a partial gene fragment (*arrA*) whose predicted amino acid sequence indicated that it had a molybdopterin cofactor binding site. This sequence homology suggested that the *arrA* gene might encode a component of an anaerobic reductase, and that this reductase might be the respiratory As(V)-reductase. We therefore sequenced the DNA downstream of the partial *arrA* fragment. Fig. 1 shows a map of the predicted ORFs. The complete *arrA* ORF is followed by another ORF, *arrB,* immediately downstream of *arrA*. Two other predicted coding sequences with opposite orientations to *arrAB* also occur downstream (CDS1 and CDS2). BLAST analyses of CDS1 and CDS2 showed that they were similar to glutathione synthases and a conserved hypothetical protein found in *Shewanella oneidensis* strain MR-1, respectively. Two putative Fnr-like binding motifs TTGAT-(N_{14})-ATAAA and TTGAT-(N_{12})-AGCAA (18) exist between *arsD* and *arrA*. A potential transcriptional terminator 32 bp downstream of *arrB* was also identified (AAAGCCACAGATGAGATTCTGTGGCTTT; inverted repeats underlined). Ribosome binding sites occur 13 bp upstream of *arrA* (AGGAG) and 13 bp upstream of *arrB* (GAGG). These sequence features suggest that *arrA* and *arrB* comprise an operon.

Physiology of the Mutants. To test the prediction that the *arrAB* gene cluster encodes the respiratory As(V) reductase, two strains of ANA-3 with nonpolar deletions in either *arrA* or *arrB* were constructed. Fig. 2 shows the respiratory growth response of strain ARRA3 (Δ*arrA*) and strain ARRB1 (Δ*arrB*) in comparison to the wild type when grown anaerobically with As(V) as the terminal electron acceptor. Whereas the wild type completely reduced As(V) to As(III) and reached stationary phase within 24 h, the deletion strains were unable to grow (Fig. 2*A*) and unable to reduce As(V) to As(III) (Fig. 2*B*) under these conditions. When ARRA3 and ARRB1 were complemented by either p*arrA4* or p*arrB1*, respectively, they regained their ability to grow and reduce As(V) under anaerobic conditions (Fig. 2*C*). The cell density for wild-type ANA-3 generally remains at $\approx 10^8$ cells per ml for several days (8); however, microscopic observation reveals that the cells shrink in size after reaching stationary phase. Consequently, the decline in optical density observed over time (see Fig. 2) most likely reflects this phenomenon.

To determine whether the mutants were specifically defective in respiratory electron transfer to As(V), we checked their ability to reduce other electron acceptors, including oxygen, fumarate, nitrate, poorly crystalline manganese oxide, poorly crystalline iron oxide, TMAO, thiosulfate, and the humic substance analog, anthraquinone disulfonate (AQDS). In all cases except for As(V), strain ARRA3 and strain ARRB1 were positive for growth and/or reduction similar to wild-type ANA-3 (8).

Sequence Analysis of *arrA* and *arrB*. The predicted molecular mass of the *arrA* gene product is 95.2 kDa and its predicted pI is 9. Closer inspection of the predicted amino acid sequence of ArrA suggests that it shares several common features of bacterial molybdopterin oxidoreductases. It contains a cysteine-rich motif (C-X_2-C-X_3-C-X_{27}-C) beginning at amino acid residue 61 and extending to position 96 that is predicted to coordinate an iron–sulfur cluster, a molybdopterin dinucleotide-binding domain (residues 715–837), and a twin arginine signal sequence (S/T-R-R-X-F-L-K) (residues 10–16) (19). Twin arginine signal sequences are commonly found at the N-terminal region of periplasmic oxidoreductases (e.g., the *Shewanella massilia* TMAO reductase (20) and the *E. coli* periplasmic nitrate reductase, NapA (21, 22)). When the signal peptide prediction program SIGNALP (www.cbs.dtu.dk/services/SignalP/) was used, a Tat recognition region, followed by a hydrophobic and a weakly charged region, were found in the first 42 aa of ArrA. These residues might comprise the part of the protein that crosses the cytoplasmic membrane, consistent with the identification of a potential cleavage site at Ala-42 detected by SIGNALP (indicated by an arrow in Fig. 1). In contrast, the *arrB* gene is directly

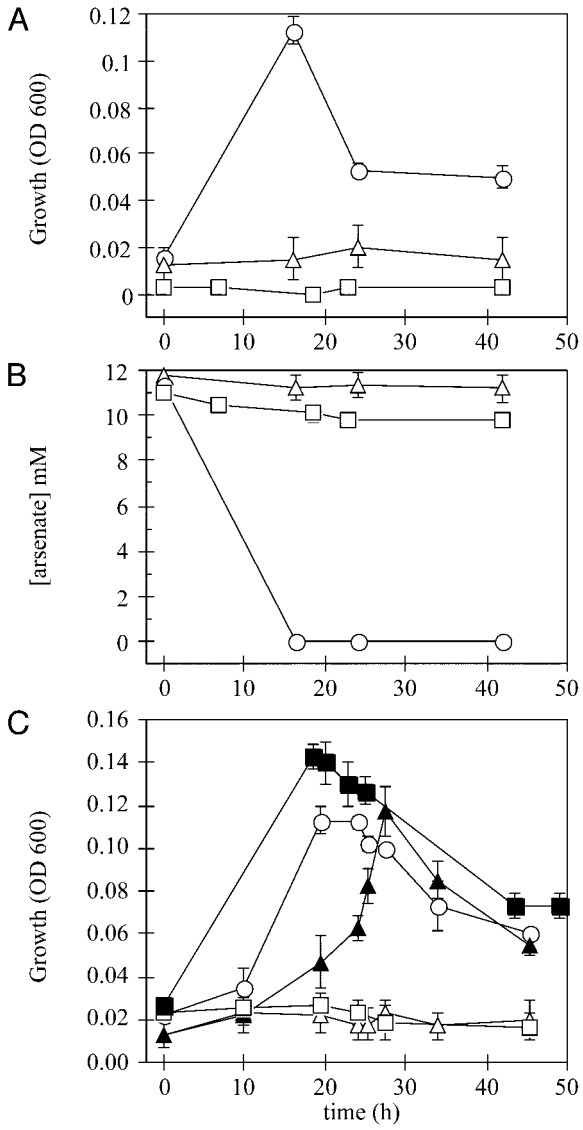


Fig. 2. As(V) respiration by wild-type ANA-3 (circles) and mutant strains ARRA3 (Δ*arrA*) (squares) and ARRB1 (Δ*arrB*) (triangles). (*A*) The time course for growth inferred by optical density at 600 nm. (*B*) The As(V) concentration in the medium at the various time points. (*C*) Anaerobic growth on As(V) by strain ARRA3 (Δ*arrA*) and strain ARRB1 (Δ*arrB*) is restored by providing a wild-type copy of the *arrA* or *arrB* gene on a complementation vector (filled symbols). Open symbols in *C* represent strains containing the complementation vector alone. Points and error bars in each panel represent the averages and standard deviations of triplicate samples.

downstream of *arrA* and is predicted to encode a protein that has a molecular mass of 25.7 kDa and a pI of 9. Sequence analysis of ArrB indicates that it could contain four iron–sulfur clusters at the following cysteine residues: (*i*) cys12, 15, 18, 69; (*ii*) cys 22, 60, 65, 57; (*iii*) cys 89, 92, 95, 183; and (*iv*) cys 99, 164, 167, 179. The arrangement of the four putative iron–sulfur clusters are predicted to be similar to the model proposed for the *E. coli* DMSO reductase iron–sulfur subunit, DmsB (23). ArrB does not contain a Tat signal sequence.

Phylogenetic Analysis of ArrA and ArrB. To gain insight into the evolutionary history of *arrA* and *arrB*, phylogenetic trees of their predicted protein products were constructed. Fig. 3*A* shows the phylogenetic relationship of ArrA to representative members of the dimethyl sulfoxide reductase family of molybdenum containing enzymes (19), including periplasmic and membrane-associated nitrate reductases (NapA and NarG), TMAO and DMSO reductases (TorA/DorA and DmsA), biotinsulfoxide reductases (BisC), formate dehydrogenases (FdhF/A/G), selenate reductase (SerA), arsenite oxidase (AsoA), and polysulfide reductase (PsrA/PhsA). ArrA clusters most closely with a hypothetical protein from the As(V)-respiring bacterium *Desulfitobacterium hafniense* (Desu0744; 51% amino acid identity, 68% amino acid similarity) and a putative arsenate reductase from *Bacillus selenitireducens* (ref. 24; 46% amino acid identity, 63% amino acid similarity).

Fig. 3*B* shows the phylogenetic relationship of ArrB to respiratory proteins containing iron–sulfur subunits including SdhB (succinate dehydrogenase), FrdB (fumarate reductase), NrfC (nitrite reductase), PsrB/PhsB, DmsB, and NarH. ArrB forms a cluster with NrfC and PrsB/PhsB. ArrB is most similar to a hypothetical iron–sulfur protein from *D. hafniense* (Desu0743; 53% amino acid identity, 61% amino acid similarity). As in strain ANA-3, the ORF encoding the ArrB homolog in *D. hafniense* is located immediately downstream of the ORF that encodes the ArrA homolog.

Discussion

When genetic analysis was used, a new gene cluster that encodes a respiratory As(V) reductase in the *Shewanella* species strain ANA-3 was identified. The two genes in this cluster, *arrA* and *arrB*, encode proteins that are predicted to be similar to those involved in the respiratory reduction of DMSO, polysulfide, TMAO, selenate, and nitrate, as well as the oxidation of formate and arsenite. Specifically, ArrA is predicted to bind a molybdenum cofactor, and ArrB is predicted to contain iron–sulfur cultures. A Tat motif at the N terminus of ArrA suggests that the As(V) reductase complex is exported to the periplasm, consistent with the fact that molybdopterin-containing proteins often are exported by the Tat pathway in their folded state (25). A Tat motif is also present in the sequence of the ArrA homolog from *D. hafniense.*

Detailed biochemical studies will be needed to describe the exact mechanism of respiratory electron transfer to As(V), but it is clear from our sequence data that the mechanism will be unlike that used to reduce As(V) by the detoxifying ArsC enzymes (11). Although the detoxifying and respiratory As(V) reductases share a common reaction substrate and product, they are completely different in structure and function. In the case of the ArsC enzymes, specific thiol groups are known to transfer electrons from intracellular reductants like glutathione or thioredoxin to As(V) in a redox cascade and no energy is gained from this process (11). In the case of the ArrA/ArrB respiratory reductase from ANA-3, a periplasmic protein complex containing a molybdopterin subunit and iron–sulfur clusters would appear to accept electrons delivered by c-type cytochromes in the cytoplasmic membrane. This is reminiscent of the type of electron transfer system thought to be involved in the oxidation of As(III) (11, 26), but with reverse electron flow. The apparent electron transfer scheme of the respiratory As(V) reductase is basically that of a classic bioenergetic chain for anaerobic respiration.

The discovery that the genes encoding the second As(V) reductase in ANA-3 are similar to those that encode enzymes involved in other types of anaerobic respiration is not surprising. What is interesting, however, is that these genes appear to define a new class of reductases that are specific for respiration on As(V); moreover, these genes appear to have a broad phylogenetic distribution. A previous report described the purification of a respiratory As(V) reductase from *Chrysiogenes arsenatis* that couples the oxidation of methyl viologen to the reduction of As(V) to As(III) (9). This enzyme has two subunits: an 87-kDa molybdenum-containing subunit (ArrA), and a 29-kDa iron–

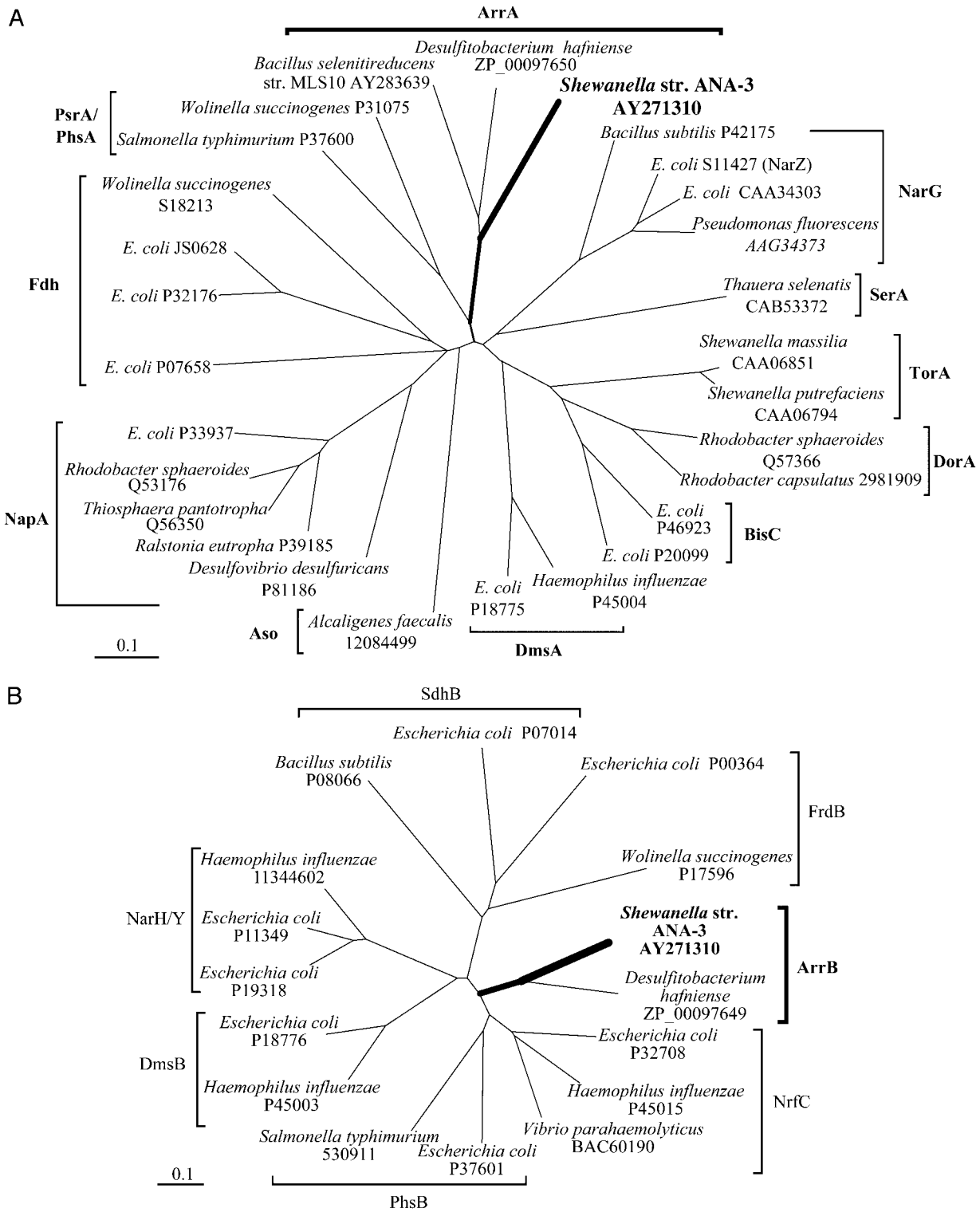


Fig. 3. Unrooted neighbor-joining trees for representative sequences from the DMSO reductase family of molybdoenzymes (19) (*A*) and iron–sulfur proteins (*B*). Aso, arsenite oxidase; Fdh, formate dehydrogenase; Nap, periplasmic nitrate reductase; Dor/Dms, DMSO reductase; Bis, biotinsulfoxide reductase; Tor, trimethylamineoxide reductase; Ser, selenate reductase; Nar, membrane-associated nitrate reductase; Arr, arsenate respiratory reductase; Psr/Phs, polysulfide reductase; Sdh, succinate dehydrogenase; Frd, fumarate reductase; Nrf, nitrite reductase. GenBank accession numbers are included next to the corresponding species. Scale bar represents the number of amino acid changes per site.

sulfur-containing subunit (ArrB). When we compare the N-terminal sequences of the *C. arsenatis* ArrA and ArrB proteins to the predicted N-terminal sequences of ArrA and ArrB from strain ANA-3, we find that they are more similar to each other than they are to any other known proteins. Furthermore, two hypothetical proteins in the genome sequence of the As(V) respiring bacterium, *D. hafniense* (7), are strikingly similar to ArrA and ArrB from ANA-3, suggesting that they too might

encode components of a respiratory As(V) reductase. Biochemical and molecular data emerging from studies of the haloalkaliphilic As(V) respiring bacterium *B. selenitireducens* strain MLS10, also suggest that ArrA and ArrB homologs are present in this strain (24). These organisms are only very distantly related to each other, representing three distinct phyla within the *Bacteria* (27): *Shewanella* ANA-3 (8) is a member of the γ-subgroup of the *Proteobacteria*, *C. arsenatis* (28) is considered the type species for the *Chrysiogenetes*, whereas *D. hafniense* (29) and *B. selenitireducens* (30) both represent the phylum *Firmicutes*. It is possible, therefore, that the *arrA* and *arrB* genes might have been distributed through horizontal gene transfer; alternatively, they may be evolutionarily deep. Before drawing conclusions about the evolution of these genes and/or how they may be used to design molecular probes to detect As(V) respiratory activity in the environment, more genetic studies describing functional arsenate-reductases from other arsenate-respiring organisms will be necessary.

We thank John Stolz for helpful discussions and access to unpublished sequence data for *Bacillus selenitrireducans*. Jared Leadbetter and members of the Newman laboratory provided constructive criticism of the manuscript. This study was made possible by grants from the Luce Foundation and the Packard Foundation (to D.K.N.). This material is based on work supported by the National Science Foundation under a grant awarded in 2002 (to C.W.S.).

1. Nriagu, J. (2002) in *Environmental Chemistry of Arsenic*, ed. Frankenberger, W. T. (Dekker, Basel), pp. 21–22.
2. Wilkie, J. A. & Hering, J. G. (1998) *Environ. Sci. Technol.* **32,** 657–662.
3. Cullen, W. R. & Reimer, K. J. (1989) *Chem. Rev.* **89,** 713–764.
4. Ahmann, D., Roberts, A. L., Krumholz, L. R. & Morel, F. M. (1994) *Nature* **371,** 750 (lett.).
5. Oremland, R. S., Newman, D. K., Kail, B. W. & Stolz, J. F. (2002) in *Environmental Chemistry of Arsenic*, ed. Frankenberger, W. T. (Dekker, Basel), pp. 273–295.
6. Herbel, M. J., Blum, J. S., Hoeft, S. E., Cohen, S. M., Arnold, L. L., Lisak, J., Stolz, J. F. & Oremland, R. S. (2002) *FEMS Microbiol. Ecol.* **41,** 59–67.
7. Niggemyer, A., Spring, S., Stackebrandt, E. & Rosenzweig, R. F. (2001) *Appl. Environ. Microbiol.* **67,** 5568–5580.
8. Saltikov, C. W., Cifuentes, A., Venkateswaran, K. & Newman, D. K. (2003) *Appl. Environ. Microbiol.* **69,** 2800–2809.
9. Krafft, T. & Macy, J. M. (1998) *Eur. J. Biochem.* **255,** 647–653.
10. Macy, J. M., Santini, J. M., Pauling, B. V., O'neill, A. H. & Sly, L. I. (2000) *Arch. Microbiol.* **173,** 49–57.
11. Mukhopadhyay, R., Rosen, B. P., Phung, L. & Silver, S. (2002) *FEMS Microbiol. Rev.* **26,** 311–325.
12. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
13. Link, A. J., Phillips, D. & Church, G. M. (1997) *J. Bacteriol.* **179,** 6228–6237.
14. Dehio, C. & Meyer, M. (1997) *J. Bacteriol.* **179,** 538–540.
15. Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M. & Peterson, K. M. (1994) *BioTechniques* **16,** 800–802.
16. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
17. Swofford, D. L. (2003) PAUP*: *Phylogenic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA).
18. Zumft, W. G. (1997) *Microbiol. Mol. Biol. Rev.* **61,** 533–616.
19. McEwan, A. G., Ridge, J. P., Mcdevitt, C. A. & Hugenholtz, P. (2002) *Geomicrobiol. J.* **19,** 3–21.
20. Dos Santos, J. P., Iobbi-Nivol, C., Couillault, C., Giordano, G. & Mejean, V. (1998) *J. Mol. Biol.* **284,** 421–433.
21. Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., Cole, J. A. & Turner, R. J. (1998) *Cell* **93,** 93–101.
22. Thomas, G., Potter, L. & Cole, J. A. (1999) *FEMS Microbiol. Lett.* **174,** 167–171.
23. Weiner, J. H., Rothery, R. A., Sambasivarao, D. & Trieber, C. A. (1992) *Biochim. Biophys. Acta* **1102,** 1–18.
24. Afkar, E., Lisak, J., Saltikov, C. W., Basu, P., Oremland, R. S. & Stolz, J. F. (2003) *FEMS Microbiol. Lett.*, in press.
25. Berks, B. C. (1996) *Mol. Microbiol.* **22,** 393–404.
26. Anderson, G. L., Williams, J. & Hille, R. (1992) *J. Biol. Chem.* **267,** 23674–23682.
27. Garrity, G., Winters, M. & Searles, D. (2001) in *Bergey's Manual of Systematic Bacteriology* (Springer, New York).
28. Macy, J. M., Nunan, K., Hagen, K. D., Dixon, D. R., Harbour, P. J., Cahill, M. & Sly, L. I. (1996) *I Int. J. Syst. Bacteriol.* **46,** 1153–1157.
29. Christiansen, N. & Ahring, B. K. (1996) *Int. J. Syst. Bacteriol.* **46,** 442–448.
30. Blum, J. S., Bindi, A. B., Buzzelli, J., Stolz, J. F. & Oremland, R. S. (1998) *Arch. Microbiol.* **171,** 19–30.

arrA Is a Reliable Marker for As(V) Respiration

D. Malasarn,[1] C. W. Saltikov,[2] K. M. Campbell,[3] J. M. Santini,[4] J. G. Hering,[3] D. K. Newman[2,3]*

Bacteria play an important role in controlling the geochemistry of arsenic. A tragic example is the case of Bangladesh, where microorganisms have been implicated in the release of arsenic into drinking water supplies and the exposure of millions of people to chronic arsenic poisoning (*1*–*3*). Arsenate [As(V)] respiration [i.e., the oxidation of organic carbon, hydrogen, or sulfide coupled with As(V) reduction to arsenite] is one of the microbial processes that contributes to arsenic mobilization (*4*). It has been difficult to monitor the activity of As(V)-respiring bacteria because they are phylogenetically diverse, and this metabolic capability is not consistently present within any given clade. Here we report that a conserved functional gene, *arrA*, can be used to detect As(V)-respiratory activity in the environment.

The *arrA* gene from the Gram-negative γ-Proteobacterium, *Shewanella* species strain ANA-3, encodes for a reductase that catalyzes respiratory As(V) reduction (*5*). *arrA* is a well-conserved gene, having 61 to 100% similarity at the amino acid level when compared with seven phylogenetically diverse As(V)-respiring bacteria (supporting online text). The ArrA proteins form a unique group within the dimethyl sulfoxide (DMSO) reductase family of molybdenum-containing enzymes, which includes other terminal reductases used in microbial respiration (fig. S1).

Because of the high degree of conservation within the ArrA protein subfamily, we designed degenerate polymerase chain reaction primers, ArrAfwd (5′-AAGGTGTATGGAATAAAGCGTTTgtbgghgaytt-3′) and ArrArev (5′-CCTGTGATTTCAGGTGCCcaytyvggngt-3′), to amplify a diagnostic region of *arrA*. These primers were tested on 13 phylogenetically diverse As(V)-respiring bacteria, one As(V)-respiring archaeon, and five negative-control strains that cannot respire As(V) but that possess other genes within the DMSO reductase family (*6*). Twelve of the 13 As(V)-respiring bacteria tested positive for *arrA*; no fragments were amplified from the As(V)-respiring archaeon or negative-control strains (fig. S2). ArrAfwd and ArrArev thus appear to be reliable markers for *arrA* for the majority of As(V)-respiring bacteria.

Poorly crystalline ferric (hydr)oxide [$Fe(OH)_3$] has been shown to be the most critical sedimentary phase in controlling arsenic mobility in a variety of locales, including anaerobic sediments of the Haiwee Reservoir in Olancha, California (*7*), and Bengal delta aquifers (*8*). Accordingly, we prepared As(V)-saturated $Fe(OH)_3$ for experiments with strain ANA-3 and the mutant strain ANA-3Δ*arrA* (*6*). Both strain ANA-3 and strain ANA-3Δ*arrA* are capable of respiratory Fe(III) reduction and As(V) reduction for the purpose of detoxification using the ArsC As(V) reductase, but only ANA-3 is capable of respiratory As(V) reduction using ArrA.

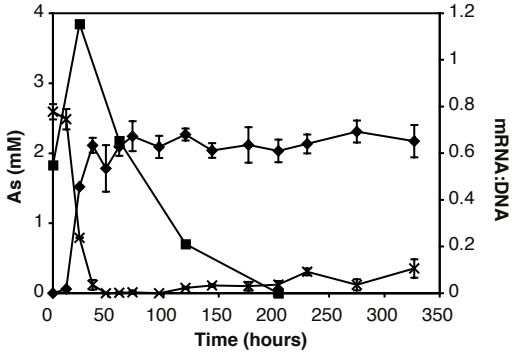


Fig. 1. *arrA* is required for As(V) reduction under iron-rich conditions. Concentrations of total As(V) (crosses) and As(III) (diamonds) are shown for samples containing *Shewanella* sp. strain ANA-3 in the presence of As(V)-saturated $Fe(OH)_3$. *arrA* expression (i.e., the mRNA:DNA ratio) is also indicated (squares). Data represent the average and standard deviation of triplicate samples, except in the case of the mRNA:DNA ratio, for which a representative data set is shown.

Reduction of As(V) occurred in samples incubated with strain ANA-3, but not in the uninoculated samples or in samples inoculated with strain ANA-3Δ*arrA*, showing that As(V) reduction is not mediated abiotically or in the absence of *arrA*, even when *arsC* is present. With strain ANA-3, the maximal expression of *arrA* (the ratio of *arrA* mRNA transcript per *arrA* gene copy number) corresponded to the fastest rate of As(V) reduction (Fig. 1). This shows that *arrA* is required to catalyze the conversion of As(V) to arsenite [As(III)] in iron-rich systems and that ArrAfwd and ArrArev can be used to track *arrA* expression.

Geochemical studies of Haiwee sediments have shown that arsenic is sorbed to $Fe(OH)_3$ and that As(III) predominates below a few centimeters (*7*). Using ArrAfwd and ArrArev, *arrA* fragments were amplified from DNA and total RNA extracted from these sediments, revealing that *arrA* is present and expressed at this site (*6*). Seven representative DNA fragments and 14 mRNA fragments were sequenced to confirm that these products were bona fide *arrA* gene fragments. These fragments ranged from ~62 to 97% identity to *arrA* sequences from *Bacillus selenitireducens*, *Chrysiogenes arsenatis*, and strain ANA-3 (fig. S3).

It is intriguing from an evolutionary perspective that *arrA* is so well conserved. On a practical level, it enables a simple molecular assay to be used to determine whether respiratory As(V) reduction is contributing to the speciation and mobilization of arsenic in a variety of environments.

References and Notes

1. C. F. Harvey *et al.*, *Science* **298**, 1602 (2002).
2. F. S. Islam *et al*, *Nature* **430**, 68 (2004).
3. A. H. Smith, E. O. Lingas, M. Rahman, *Bull. World Health Organ.* **78**, 1093 (2000).
4. R. S. Oremland, J. F. Stolz, *Science* **300**, 393 (2003).
5. C. W. Saltikov, D. K. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10983 (2003).
6. Materials and methods are available as supporting material on *Science* Online.
7. P. E. Kneebone, P. A. O'Day, N. Jones, J. G. Hering, *Environ. Sci. Technol.* **36**, 381 (2002).
8. J. Akai *et al.*, *Appl. Geochim.* **19**, 215 (2004).
9. We thank R. Oremland, F. Rosenzweig, and C. House for sending us strains. All new *arrA* gene sequences have been deposited in GenBank. Accession numbers are available in supporting online material. Supported by the Luce and Packard Foundations (D.K.N.) and NSF award nos. BES-0201888 (J.G.H.) and DBI-0200145 (C.W.S.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5695/455/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References and Notes

6 July 2004; accepted 16 August 2004

[1]Division of Biology, [2]Division of Geological and Planetary Sciences, [3]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA. [4]Department of Microbiology, La Trobe University, 3086 Victoria, Australia.

*To whom correspondence should be addressed. E-mail: dkn@caltech.edu

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 2005, p. 8228–8235 Vol. 71, No. 12
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.12.8228–8235.2005

UniFrac: a New Phylogenetic Method for Comparing Microbial Communities

Catherine Lozupone[1] and Rob Knight[2]*

Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309,[1] and Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309[2]

Received 3 May 2005/Accepted 26 August 2005

We introduce here a new method for computing differences between microbial communities based on phylogenetic information. This method, UniFrac, measures the phylogenetic distance between sets of taxa in a phylogenetic tree as the fraction of the branch length of the tree that leads to descendants from either one environment or the other, but not both. UniFrac can be used to determine whether communities are significantly different, to compare many communities simultaneously using clustering and ordination techniques, and to measure the relative contributions of different factors, such as chemistry and geography, to similarities between samples. We demonstrate the utility of UniFrac by applying it to published 16S rRNA gene libraries from cultured isolates and environmental clones of bacteria in marine sediment, water, and ice. Our results reveal that (i) cultured isolates from ice, water, and sediment resemble each other and environmental clone sequences from sea ice, but not environmental clone sequences from sediment and water; (ii) the geographical location does not correlate strongly with bacterial community differences in ice and sediment from the Arctic and Antarctic; and (iii) bacterial communities differ between terrestrially impacted seawater (whether polar or temperate) and warm oligotrophic seawater, whereas those in individual seawater samples are not more similar to each other than to those in sediment or ice samples. These results illustrate that UniFrac provides a new way of characterizing microbial communities, using the wealth of environmental rRNA sequences, and allows quantitative insight into the factors that underlie the distribution of lineages among environments.

Sequencing of 16S rRNA genes from environmental samples has revealed that microbial diversity is far more extensive than had ever been imagined from studies of cultured microorganisms alone and that microorganisms represent the majority of the phylogenetic diversity of life on earth (34). Culture-independent studies of microbial populations were pioneered in the Pace lab in 1985 (35, 36); the technique is now so prevalent that an estimated 151,339 sequences from small-subunit-rRNA environmental clones had been deposited in GenBank as of 1 August 2005. [We estimated the total number of environmental clone small-subunit-rRNA gene sequences published in GenBank with an Entrez search with the string "(SSU OR 16S OR 18S OR small subunit) AND (rRNA OR rDNA OR ribosomal RNA) AND (uncult* OR unidentified OR unknown)" (modified from reference 37).] Only half of the 52 major bacterial lineages described in the last comprehensive review have cultivated representatives, and widespread, numerically dominant phylotypes are often only distantly related to culturable strains (37). Thus, our sole source of information about the biology of much of the diversity of life is the environmental distribution of sequences.

Several statistical techniques have been developed to use environmental 16S rRNA clone sequences to compare microbial communities between samples. Unfortunately, many of these techniques are limited because they do not account for the different degrees of similarity between sequences. Sequences are usually grouped if their 16S rRNA genes are 95 to 99% identical (16, 30); with a cutoff of 98%, such techniques would treat sequences with 3% and 40% sequence divergence equally. This results in a substantial loss of information since 16S rRNA and phenotypic variances are positively correlated (33). Techniques with this limitation include the Sørenson and Jaccard indices of group overlap (28), the LibShuff (40; http://www.arches.uga.edu/~whitman/libshuff.html) and ∫-LibShuff (39; http://www.plantpath.wisc.edu/fac/joh/S-LibShuff.html) methods, and hierarchical clustering and ordination of samples based on the distribution of sequences belonging to different groups (17).

Phylogenetic distance measures can provide far more power because they exploit the degree of divergence between different sequences. Two phylogenetic approaches that assess whether communities differ significantly in composition, the *P* and F_{ST} tests, have recently been developed (30). The *P* test uses parsimony to determine whether the distribution of modern sequences in different environments reflects a history of fewer changes between environments than would be expected by chance. The F_{ST} test identifies cases where more sequence variation exists between two communities than within a single community. Although these techniques greatly increase our ability to test for differences between pairs of communities, they have only been applied to determining whether samples are significantly different and have not been used to compare many samples simultaneously with clustering or ordination techniques. In addition, neither measure accounts for branch length information when comparing samples.

Here we introduce a new phylogenetic method, called UniFrac, that measures the distance between communities based on the lineages they contain. UniFrac can be used to compare many samples simultaneously because it satisfies the technical

* Corresponding author. Mailing address: Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309. Phone: (303) 492-1984. Fax: (303) 492-7744. E-mail: rob@spot.colorado.edu.

requirements for a distance metric (it is always positive, is transitive, and satisfies the triangle inequality) and can thus be used with standard multivariate statistics such as unweighted-pair group method using average linkages (UPGMA) clustering (9) and principal coordinate analysis (23). Similarly, UniFrac is more powerful than nonphylogenetic distance measures because it exploits the different degrees of similarity between sequences. To demonstrate the utility of the UniFrac metric for comparing multiple community samples and determining the factors that explain the most variation, we compared bacterial populations in different types of geographically dispersed marine environments.

Small-subunit-rRNA gene surveys have been performed in many marine environments, including oligotrophic open-ocean (11), coastal temperate (1, 22) and polar (2, 8) seawater, polar sea ice (3, 6, 7), and marine sediments (4, 5, 25, 38). Comparing this range of samples using the UniFrac technique provides a coherent picture of the distribution of bacterial lineages that provides a context for many individual published observations while allowing us to test specific ideas about the distribution of bacterial lineages. In particular, we asked the following questions.

How does culturing affect similarities between samples? Few organisms in environmental samples are culturable, but it is unknown whether the cultured isolates from an environment yield communities that resemble the communities from the original habitat. We addressed this issue by comparing gene libraries from both cultured bacteria and uncultured environmental samples of marine ice, water, and sediment to test whether the cultured samples appeared more similar to uncultured samples from the same environment or to each other.

How cosmopolitan are bacterial lineages? Although many studies have suggested that bacteria are mostly cosmopolitan (11, 13, 32, 45), others have suggested that for certain habitats, such as the Arctic and Antarctic, geographical separation plays a major role in structuring communities because of difficulties in dispersal (in this case, of transferring psychrophilic bacteria across the warm equatorial region) (2, 42). We addressed this controversy by comparing marine ice and sediment from the Arctic and Antarctic.

Are marine ice, sediment, and seawater three distinct, homogeneous habitats? Marine ice, sediment, and water are generally treated in the literature as distinct habitat types with distinct challenges. We compared 16S rRNA libraries from geographically diverse marine water, sediment, and ice samples to test whether these habitat types harbor consistent bacterial communities that differ from one another.

UniFrac metric. The unique fraction metric, or UniFrac, measures the phylogenetic distance between sets of taxa in a phylogenetic tree as the fraction of the branch length of the tree that leads to descendants from either one environment or the other, but not both (Fig. 1). This measure thus captures the total amount of evolution that is unique to each state, presumably reflecting adaptation to one environment that would be deleterious in the other. rRNA is used purely as a phylogenetic marker, indicating the relative amount of sequence evolution that has occurred in each environment. Intuitively, if two environments are similar, few adaptations would be needed to transfer from one community to the other. Consequently, most nodes in a phylogenetic tree would have descendants from both communities, and much of the branch length in the tree would be shared (Fig. 1A). In contrast, if two communities are so distinct that an organism adapted to one could not survive in the other, then the lineages in each community would be distinct, and most of the branch length in the tree would lead to descendants from only one of the two communities (Fig. 1B).

Like the *P* test and the F_{ST} test, UniFrac can be used to determine whether two communities differ significantly by using Monte Carlo simulations. Two communities are considered different if the fraction of the tree unique to one environment is greater than would be expected by chance. We performed randomizations by keeping the tree constant and randomizing the environment that was assigned to each sequence in the tree (Fig. 1C).

UniFrac can also be used to produce a distance matrix describing the pairwise phylogenetic distances between the sets of sequences collected from many different microbial communities (Fig. 1D). We compared two samples by removing from the tree all sequences that were not in either sample and computing the UniFrac for each reduced tree. Standard multivariate statistics, such as UPGMA clustering (9) and principal coordinate analysis (23), can then be applied to the distance matrix to allow comparisons between the biotas in different environments (Fig. 1D).

MATERIALS AND METHODS

Environmental samples. We analyzed 20 small-subunit-rRNA sequence libraries generated in 12 different studies of marine environments (Table 1). For studies reporting both cultured and uncultured sequences or sampling from multiple environment types, we used sequence annotations to distinguish the different sampling methods and to assign sequences to specific environmental samples.

Three of the studies evaluated bacterial communities in Arctic and/or Antarctic sea ice based on cultured isolates (3), environmental clones (7), or both (6). Five of the studies derived sequences from the water columns of marine environments, including pelagic bacteria from the North Sea (8), bacterioplankton assemblages from the Arctic Ocean (2), subsurface subtropical waters of the Atlantic and Pacific oceans (11), and temperate coastal water in the Great South Bay in Long Island (22) and from the marine end of the Plum Island Sound estuary in northeastern Massachusetts (1). The remaining four studies examined marine sediment, including sediments from off the coast of Spitzbergen in the Arctic Ocean (38), associated with *Calyptogena* communities in the deepest cold-seep area in the Japan Trench (25), and from the Antarctic continental shelf (4, 5). While three of the sediment papers reported sequences from multiple sediment cores in the same region (4, 25, 38), one reported sequences from three different depths within a single sediment core (5).

Sequences from the 12 studies were initially assigned to 23 samples. After the removal of sequences with many sequencing errors and nonbacterial sequences, the samples contained between 9 and 544 sequences. Small samples could produce misleading results because of stochastic variation in the subset of the lineages sampled. To avoid these effects, we excluded from the analysis three samples represented by 12 or fewer sequences. These included samples containing 9 sequences from uncultured clones in the North Sea (8), 10 sequences from Arctic sea ice (7), and 12 sequences from cultured isolates in Arctic seawater underlying sea ice (3). After the removal of these sequences, each sample was represented by at least 17 sequences (Table 1).

Data analysis. We implemented UniFrac and associated analyses in Python 2.3.4 and ran all calculations on a Macintosh G4 computer running OSX 10.3.8. All code is available at http://bayes.colorado.edu/unifrac.zip. We implemented UPGMA clustering (9) and principal coordinate analysis (23) as described previously.

We downloaded small-subunit-rRNA sequences generated in the 12 different studies of marine environments (Table 1) from GenBank, imported them into the Arb package (26), and aligned them using a combination of the Arb autoaligner and manual curation. Because several studies used bacterium-specific

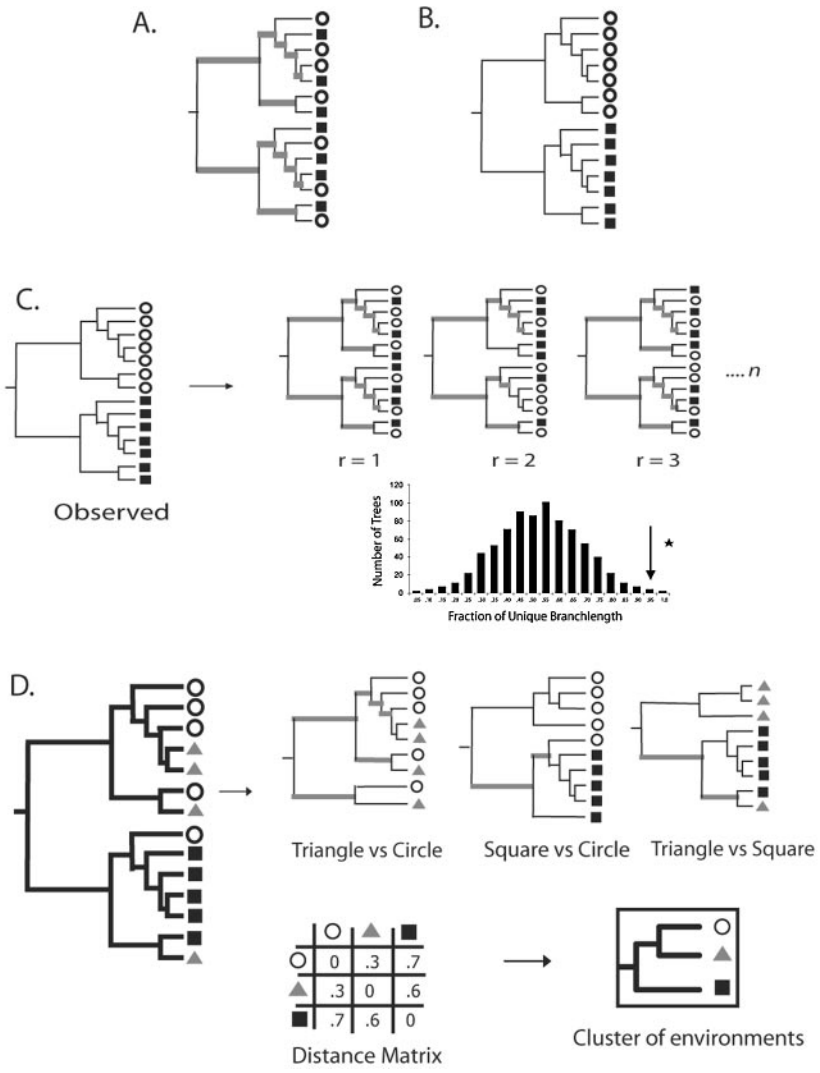


FIG. 1. Calculation of the UniFrac distance metric. Squares, triangles, and circles denote sequences derived from different communities. Branches attached to nodes are colored black if they are unique to a particular environment and gray if they are shared. (A) Tree representing phylogenetically similar communities, where a significant fraction of the branch length in the tree is shared (gray). (B) Tree representing two communities that are maximally different so that 100% of the branch length is unique to either the circle or square environment. (C) Using the UniFrac metric to determine if the circle and square communities are significantly different. For *n* replicates (*r*), the environment assignments of the sequences were randomized, and the fraction of unique (black) branch lengths was calculated. The reported *P* value is the fraction of random trees that have at least as much unique branch length as the true tree (arrow). If this *P* value is below a defined threshold, the samples are considered to be significantly different. (D) The UniFrac metric can be calculated for all pairwise combinations of environments in a tree to make a distance matrix. This matrix can be used with standard multivariate statistical techniques such as UPGMA and principal coordinate analysis to compare the biotas in the environments.

primers, we excluded all nonbacterial sequences from the analysis. We added the aligned sequences to a tree representing a range of phylogenetic groups from the Ribosomal Database Project II (29) by Phil Hugenholtz (15). This sequence addition used the parsimony insertion tool and a lane mask (lanemaskPH) supplied in the same database so that only phylogenetically conserved regions were considered. We exported the tree from Arb and annotated each sequence with 1 of 20 sample designations (Table 1). We then performed significance tests, UPGMA clustering, and principal coordinate analysis using UniFrac.

Jackknifing. We used jackknifing to determine how the number and evenness of sequences in the different environments affected the UPGMA clustering results. Specifically, we repeated the UniFrac analysis with trees that contained only a subset of the sequences and measured the number of times we recovered each node that occurred in the UPGMA tree from the full data set. In each simulation, we evaluated 100 reduced trees in which all of the environments were represented by the same specified number of sequences, using sample sizes of 17, 20, 31, 36, 40, and 58 sequences. These thresholds reflect the sample sizes from different environments in our original data set. If an environment had more than the specified number of sequences, we removed sequences at random; environments with fewer sequences were removed from the tree entirely.

RESULTS

We used UniFrac to determine which of the microbial communities represented by the 20 different samples were significantly different (Table 2) and as the basis for a distance matrix to cluster the samples using UPGMA (Fig. 2) and to perform principal coordinate analysis (Fig. 3). We used jackknifing to assess confidence in the nodes of the UPGMA tree (Table 3).

TABLE 1. Gene library information

Sample[a]	Reference	No. of sequences	Water column depth (m)	Sediment depth (cm)	Latitude, longitude	Temp (°C)
SRU1	38	79	155	0–1.1	76°58′N, 15°34′E	2.6
STU2	25	33	6,400		40°06′N, 144°11′E	
SNU3	4	36	709–940	1–2	66°S, 143°E	
SNC4	4	31	709–940		66°S, 143°E	
SNU5	5	101	761	0–0.4	66°32′S 143°38′E	
SNU6	5	146	761	1.5–2.5	66°32′S 143°38′E	
SNU7	5	231	761	20–21	66°32′S 143°38′E	
WRU8	2	87	55, 131		72–88°N, 51–356°E	
WTC9	8	36	1		54°09′N, 7°52′E	
WTU10	22	75	1–2		40°N, 73°E	23.8–29.2
WTC11	22	21	1–2		40°N, 72°E	23.8–29.2
WTU12	1	544			42°N, 71°E	16
WPU13	11	17	10		32°37′N 64°57′W	
WPU14	11	40	100, 500		31°49′N 64°57′W	
INC15	3	58			68°S, 78°E	
IRU16	6	62			80°N, 0°E	
IRC17	6	109			80°N, 0°E	
INU18	6	20			70°S, 15°E	
INC19	6	87			70°S, 15°E	
INU20	7	75			62–77°S, 74–165°E	

[a] The first character in the sample name designates the environment type (S, marine sediment; W, water; and I, ice). The second character indicates the geographic location (R, Arctic; N, Antarctic; T, temperate; and P, tropical). The third character indicates whether the sequences were derived from cultured isolates (C) or environmental clones (U).

The results show biologically meaningful patterns that unite many individual observations in the literature and reveal several striking features of microbial communities in marine environments.

Samples from cultured isolates resemble each other rather than uncultured samples from the same environment. Although most bacteria in seawater and sediment cannot be cultivated with standard techniques (8, 10, 18, 38), most bacteria in sea ice are thought to be culturable, as sea ice samples have a high viable/total count ratio (6, 14, 19) and considerable overlap in phylotypes between cultured and uncultured samples (6, 7). To test this hypothesis, we examined the relation-

TABLE 2. UniFrac *P* values[a]

Sample	Compared sample(s) (*P* value)
SRU1	SNU3 (0.118), STU2 (0.111)
STU2	SNU3 (0.201), SRU1 (0.111), SNU5 (0.066), SNU6 (0.107)
SNU3	SNU6 (0.802), SNU7 (0.070), SRU1 (0.118), STU2 (0.201)
SNC4	WTC11 (0.105), SNU5 (0.053)
SNU5	SNC4 (0.053), STU2 (0.066)
SNU6	SNU7 (0.394), SNU3 (0.802), STU2 (0.107)
SNU7	SNU6 (0.394), SNU3 (0.070)
WRU8	
WTC9	WTC11 (0.639), INU20 (0.076), INC19 (0.155), INU18 (0.097)
WTU10	
WTC11	SNC4 (0.105), WTC9 (0.639), INC19 (0.055)
WTU12	
WPU13	WPU14 (0.238)
WPU14	WPU13 (0.238)
INC15	
IRU16	INU18 (0.257)
IRC17	
INU18	WTC9 (0.097), INU20 (0.055), INC19 (0.233), IRC17 (0.257)
INC19	WTC11 (0.055), WTC9 (0.155), INU18 (0.233)
INU20	WTC9 (0.076), INU18 (0.055)

[a] UniFrac *P* values were based on comparisons to 1,000 randomized trees. Results are listed only if the *P* value (listed in parentheses) is ≥0.05. All other pairwise comparisons indicated that the communities were significantly different.

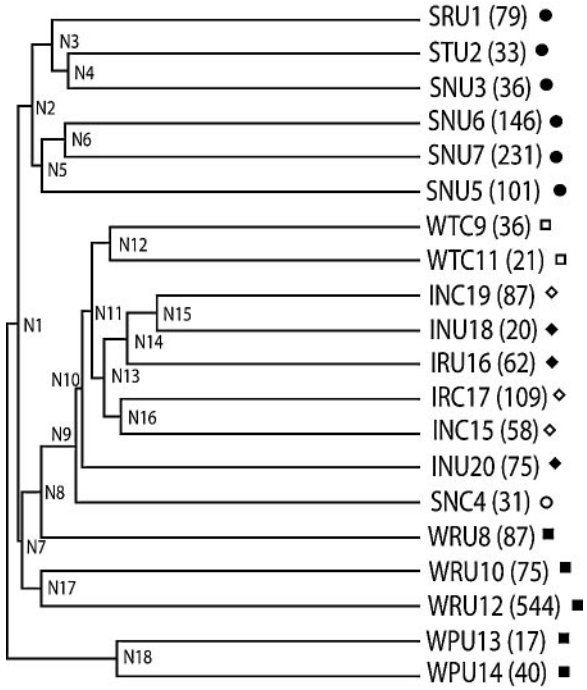


FIG. 2. UPGMA cluster of marine samples. The number of sequences that represent each environment is indicated next to the sample name, as well as the symbol with which the sample is represented in Fig. 3.

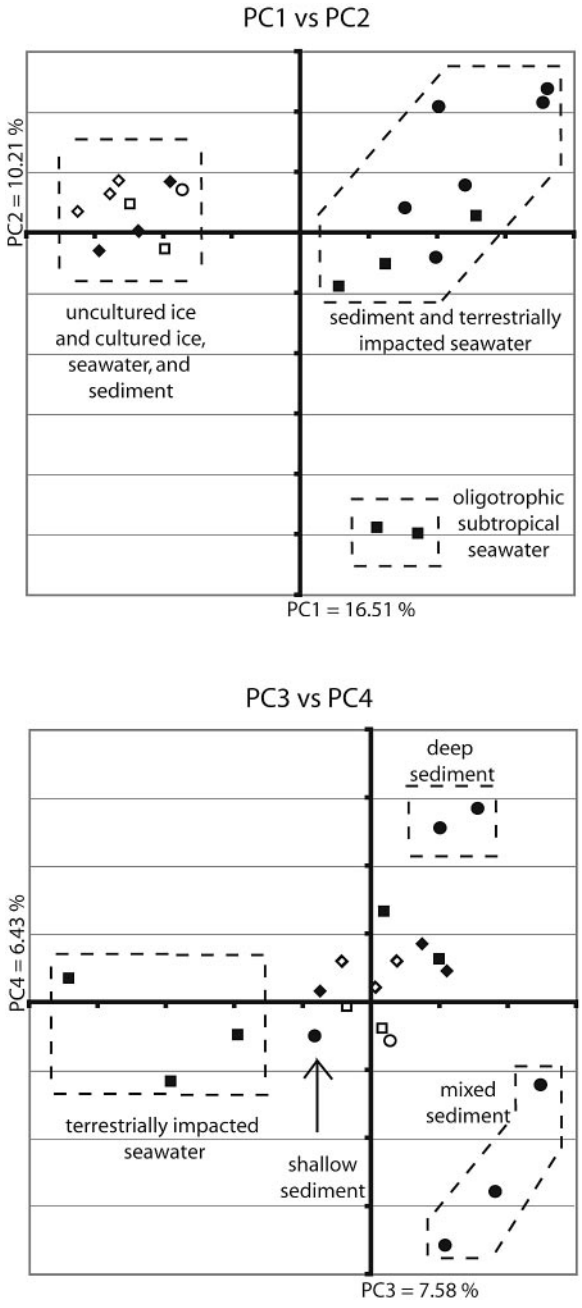


FIG. 3. First four principal coordinates from a principal coordinate analysis of marine samples. Samples from marine ice are represented by diamonds, sediment samples are represented by circles, and water samples are represented by squares. Shapes representing samples derived from cultured isolates are open, and those representing samples from environmental clones are filled. The percentages in the axis labels represent the percentages of variation explained by the principal coordinates.

TABLE 3. UPGMA jackknifing results

Node	% of trials with node[a]					
	17	20	31	36	40	58
N1	3	14	31	27	12	NA
N2	8	1	29	33	48	63
N3	1	8	7	11	NA	NA
N4	14	16	11	NA	NA	NA
N5	1	0	0	1	27	37
N6	27	36	57	67	53	63
N7	23	23	36	44	52	66
N8	22	17	17	39	31	37
N9	52	58	64	NA	NA	NA
N10	8	16	79	96	94	100
N11	6	12	40	46	NA	NA
N12	13	31	NA	NA	NA	NA
N13	16	38	41	38	64	79
N14	34	50	29	23	12	6
N15	69	77	NA	NA	NA	NA
N16	18	40	27	28	28	21
N17	24	35	43	46	37	50
N18	97	NA	NA	NA	NA	NA

[a] For each node in the UPGMA tree (Fig. 2) (rows), the numbers show the percentages of trials ($n = 100$) that the node occurred in when each environment was represented by only 17, 20, 31, 36, 40, or 58 sequences (columns). The node names correspond to the node labels in Figure 2. NA, not available.

ship between cultured isolates and environmental clone sequences derived from the same locations for sediment (SNC4 and SNU3), seawater (WTC11 and WTU10), and ice from both the Arctic (IRC17 and IRU16) and Antarctic (INC19 and INU18) (see Table 1 for an explanation of sample abbreviations). We also included additional cultured samples from seawater (WTC9) and cultured and uncultured samples from ice (INC15 and INU20).

Cultured and uncultured sea ice bacteria cluster with each other and with the other cultured isolates (Fig. 2). This association is well supported by jackknife values (Table 3). The node that groups the cultured and uncultured ice samples together (Fig. 2, N10) is recovered 100% of the time, with 58 sequences per sample (note that at this point only five of the six ice samples are still in the tree because one sample has only 20 sequences). Pairwise significance tests for differences between environments further support this observation (Table 2). The cultured component of the Antarctic ice sample (INC19) does not differ significantly from environmental clones from the same sample (INU18).

In contrast, bacteria cultured from sediment (SNC4) and seawater (WTC11 and WTC9) cluster with other cultured samples rather than with environmental clones from the same studies (SNU3 and WTU10) in the UPGMA tree. This observation is again supported by jackknife values (Table 3). With 31 sequences, SNC4 clusters with the other cultured sequences 64% of the time (Table 3, N9) but never clusters with SNU3 or exclusively with the sediment samples (data not shown). Likewise, with 36 sequences per sample, WTC9 clusters with other cultured sequences 96% of the time (Table 3, N10). In addition, pairwise significance tests (Table 2) show that the culturable components of a seawater sample (WTC11) and a sediment sample (SNC4) differ significantly from the environmental clone sequences from the same environment (WTU10 and SNU3, respectively) but not from cultured samples from different environments.

The sequences of the culturable components of the seawater and sediment samples most resemble the environmental clone sequences from sea ice. This observation is best illustrated by principal coordinate analysis (Fig. 3). In principal coordinate analysis, a distance matrix is used to plot n samples in n-

dimensional space. The vector through the space that describes as much variation as possible is principal coordinate 1. Orthogonal axes are subsequently assigned to explain as much of the variation not yet explained by previously assigned axes as possible. When few independent factors cause most of the variation, the first two or three principal coordinates often explain most of the variation in the data. In this case, the first four principal coordinates describe 41% of the variation, suggesting that many independent factors cause variation between samples (as might be expected for such diverse environments). Strikingly, plots of the principal components produce biologically meaningful clusters of samples, even though the individual components account for little of the variation (Fig. 3).

The first principal coordinate, which explains 17% of the variation in the data, clearly separates all samples of cultured isolates and uncultured ice from all samples of uncultured sediment and seawater (Fig. 3). This result suggests that bacteria capable of growing either in sea ice or in pure culture share some property, such as the ability to grow rapidly in the absence of symbionts, that is the largest factor contributing to the variation between these samples.

Geography plays a minor role in structuring communities compared to the environment type. Our analyses support the hypothesis that geography plays a minimal role in structuring bacterial communities and that bacterial types are dispersed widely in similar habitat types across the globe (11, 13, 32). Sea ice samples from the Arctic (IRU16 and IRC17) did not separate from those from the Antarctic (INC15, INU18, INC19, and INU20) in either the UPGMA or principal coordinate clusters (Fig. 2 and 3). Similarly, bacterial community samples from Antarctic sediment (SNU3 and -5 to -7) cluster with those from sediments from the Arctic (SRU1) and Japan (STU2) (Fig. 2). This result shows that, as expected, the environment type (ice, seawater, or sediment) dominates the differences between communities. Within an environment type, we found no support for the hypothesis that samples from each pole would form a discrete cluster. This result may indicate that other differences between the samples had a greater impact on bacterial composition than being located on opposite sides of the earth and contradicts the prediction that the communities in the Arctic and Antarctic would differ because of difficulties in dispersing psychrophilic bacteria across the warm equatorial region (42). However, the poor jackknife values for resolving these nodes (Table 2, nodes N3, N14, and N16) may indicate that more sequences and more samples are needed to resolve this issue definitively.

Uncultured bacterial communities in sediment and ice form distinct clusters, but communities in seawater samples do not. Finally, our analyses support the hypothesis that each marine sediment and ice community analyzed here forms a distinct group. Environmental clone libraries from each of these types of environment cluster together by UPGMA (Fig. 2), even though they were retrieved from very different locations (Table 1). In addition, uncultured sediment samples always differ significantly from seawater and ice samples but differ little from each other (Table 2). For instance, bacteria in sediment cores from the Antarctic continental shelf (SNU3), the Arctic coastal region (SRU1), and the deepest cold-seep area of the Japan Trench (STU2) did not significantly differ, despite large differences in depth, proximity to land, and geographical location. With 58 sequences, the four remaining sediment samples grouped together 63% of the time (Table 3, N2).

In contrast, seawater samples do not all cluster together but are grouped in biologically meaningful ways. For instance, an Arctic seawater sample (WRU8) clusters with an Arctic sea ice sample in the UPGMA cluster (Fig. 2). Nutrient-rich coastal seawater communities (WTU10 and WTU12) cluster together, as do oligotrophic open ocean communities (WPU13 and WPU14) (Fig. 2 and 3). These associations are often supported by jackknife values. For instance, with only 17 sequences, WPU12 and WPU14 group together 97% of the time (Table 2, N18), and with 58 sequences, WTU12 groups with WTU10 50% of the time (Table 2, N17) and with WRU8 49% of the time (data not shown). In contrast, nodes that group all of the seawater samples together are only rarely observed (5% of the time for groups with 17 sequences and 4% of the time for groups with 40 sequences). Thus, there are major differences in bacterial communities between different types of seawater, suggesting that, unlike marine sediment and ice, seawater should not be considered a distinct, homogeneous environment.

The bacterial community in the Arctic seawater sample (WRU8) appears to be more similar to those in coastal water (WTU10 and WTU12) than to those in open ocean seawater (WPU13 and WPU14). This community clusters closer to the coastal communities in both UPGMA and principal coordinate analyses (Fig. 2 and 3). One possible explanation is that like the coastal communities, the Arctic Ocean has high inputs of terrigenous matter: it is estimated that 25% of the dissolved organic carbon in the Arctic Ocean is derived from river runoff (44). The terrestrially impacted seawater communities (WTU10, WTU12, and WRU8) also resemble the communities in sediment samples. This is most clearly shown in principal coordinate analyses, where they cluster near each other in PC1 and PC2 but clearly separate along PC3 (Fig. 3).

DISCUSSION

The detection of biologically meaningful patterns of variation between marine samples illustrates the utility of UniFrac for explaining the distribution of bacterial lineages in the environment. The ability of UniFrac to integrate sequence data from many diverse studies makes it suitable for large-scale comparisons between environments. The ability of UniFrac to integrate sequence data from many diverse studies makes it suitable for large-scale comparisons, between environments despite variability in data collection techniques. For instance, different studies used different sequencing primers, and thus little of the 16S rRNA molecule was present in all sequences in the alignment. This made it impossible to use algorithms that require the same sequence region to create the phylogenetic tree for analysis. Potential imperfections in the Arb parsimony insertion tool for creating the phylogeny, however, were not great enough to confound the detection of biologically meaningful patterns of variation.

Those who performed the previous studies also chose clones for sequencing using different methods: some screened clones with restriction enzyme-based techniques (6–8, 22, 25, 38) or denaturing gradient gel electrophoresis (2, 4) prior to sequencing, while some sequenced samples directly (1, 5, 11). Since

UniFrac does not count the number of times each sequence is observed, these data can still be compared, although the "evenness" component that is a standard measure of diversity (27) is not currently represented in UniFrac. Since we compared environments on a large scale, the ability of particular lineages of organisms to survive in each environment is more likely to represent the relevant aspects of similarity between environments than the relative abundance of each surviving lineage. However, although the practice of predicting abundance from environmental clone data is sometimes questioned because of PCR bias and differences in genomic DNA extraction methods and rRNA copy numbers (21, 43), such data can be useful, especially on smaller spatial and temporal scales (20, 24, 31, 41). We have thus also developed a variant of the algorithm that weights the phylogenetic differences according to the abundance of each lineage, which will allow questions about evenness to be addressed.

Jackknifing the UPGMA tree revealed that surprisingly small sample sizes can be sufficient to detect associations between groups of samples. For example, the oligotrophic seawater samples WPU13 and WPU14 cluster together stably with only 17 sequences. However, samples that are more diverse, such as those from sediments, or less distinct, such as the Antarctic and Arctic ice samples, require more sequences for robust conclusions to be drawn. We recently demonstrated that UniFrac is robust even for very similar samples when the sample size is large. We were able to detect an association between kinship and gut microbial community structure in related mice, using sequence sets of 200 to 500 per mouse (24). We thus expect that the utility of UniFrac will increase as larger environmental samples become available.

Our analysis provides a unified framework for explaining previous observations in the literature, such as the observation that culturing affects the observed diversity in seawater and sediment but not that in ice. It also allows broader conclusions, such as the observation that terrestrially impacted seawater samples from polar and temperate climates resemble each other and sediment samples but differ greatly from tropical oligotrophic seawater samples. Terrestrially impacted seawater probably resembles sediment more than oligotrophic seawater for reasons other than relative nutrient availability, since one sediment sample (STU2) was obtained from 6,400 m below sea level and received low inputs of organic carbon (25). The resemblance may instead arise because terrestrially impacted seawater has a higher concentration of particles, and particle-associated and freely suspended marine bacteria are known to differ (12).

The large differences between different seawater communities are surprising, since the ubiquity of certain bacterial lineages in pelagic systems, such as SAR11 and SAR86, has been taken as evidence that much of the ocean harbors similar bacteria (11, 22). In contrast, the different sediment samples are remarkably similar. This supports the hypothesis that large portions of the sea floor have similar biotas because of similar environmental conditions such as nutrient availability and temperature (e.g., 90% of the sea floor has temperatures below 4°C) and similar processes such as sulfate reduction (7, 38).

Conclusion. The utility of UniFrac for making broad comparisons between the biotas of different environments based on 16S rRNA sequences has enormous potential to shed light on biological factors that structure microbial communities. The vast wealth of 16S rRNA sequences in GenBank and of environmental information about these sequences in the literature, combined with powerful phylogenetic tools, will greatly enhance our understanding of how microbial communities adapt to unique environmental challenges.

ACKNOWLEDGMENTS

Catherine Lozupone was supported in part by NIH predoctoral training grant T32 GM08759.

We thank Mike Yarus, Norm Pace, Scott Kelley, Ruth Ley, Shelley Copley, and Corrella Detweiler for their comments on the manuscript.

REFERENCES

1. **Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz.** 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature **430:**551–554.
2. **Bano, N., and J. T. Hollibaugh.** 2002. Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. Appl. Environ. Microbiol. **68:**505–518.
3. **Bowman, J. P., S. A. McCammon, M. V. Brown, D. S. Nichols, and T. A. McMeekin.** 1997. Diversity and association of psychrophilic bacteria in Antarctic sea ice. Appl. Environ. Microbiol. **63:**3068–3078.
4. **Bowman, J. P., S. A. McCammon, J. A. Gibson, L. Robertson, and P. D. Nichols.** 2003. Prokaryotic metabolic activity and community structure in Antarctic continental shelf sediments. Appl. Environ. Microbiol. **69:**2448–2462.
5. **Bowman, J. P., and R. D. McCuaig.** 2003. Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment. Appl. Environ. Microbiol. **69:**2463–2483.
6. **Brinkmeyer, R., K. Knittel, J. Jurgens, H. Weyland, R. Amann, and E. Helmke.** 2003. Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. Appl. Environ. Microbiol. **69:**6610–6619.
7. **Brown, M. V., and J. P. Bowman.** 2001. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). FEMS Microbiol. Ecol. **35:**267–275.
8. **Eilers, H., J. Pernthaler, F. O. Glöckner, and R. Amann.** 2000. Culturability and in situ abundance of pelagic bacteria from the North Sea. Appl. Environ. Microbiol. **66:**3044–3051.
9. **Felsenstein, J.** 2004. Inferring phylogenies. Sinauer Associates, Inc., Sunderland, Mass.
10. **Ferguson, R. L., E. N. Buckley, and A. V. Palumbo.** 1984. Response of marine bacterioplankton to differential filtration and confinement. Appl. Environ. Microbiol. **47:**49–55.
11. **Fuhrman, J. A., K. McCallum, and A. A. Davis.** 1993. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. Appl. Environ. Microbiol. **59:**1294–1302.
12. **Giovannoni, S. J., and M. Rappé.** 2000. Evolution, diversity, and molecular ecology of marine prokaryotes, p. 47–84. *In* D. L. Kirchman (ed.), Microbial ecology of the oceans. John Wiley & Sons, Inc., New York, N.Y.
13. **Glöckner, F. O., E. Zaichikov, N. Belkova, L. Denissova, J. Pernthaler, A. Pernthaler, and R. Amann.** 2000. Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. Appl. Environ. Microbiol. **66:**5053–5065.
14. **Helmke, E., and H. Weyland.** 1995. Bacteria in sea ice and underlying water of the eastern Weddell Sea in midwinter. Mar. Ecol. Prog. Ser. **117:**269–287.
15. **Hugenholtz, P.** 2002. Exploring prokaryotic diversity in the genomic era. Genome Biol. **3:**reviews0003.
16. **Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. Bohannan.** 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Appl. Environ. Microbiol. **67:**4399–4406.
17. **Hur, I., and J. Chun.** 2004. A method for comparing multiple bacterial community structures from 16S rDNA clone library sequences. J. Microbiol. **42:**9–13.
18. **Jannasch, H. W., and G. E. Jones.** 1959. Bacterial populations in sea water as determined by different methods of enumeration. Limnol. Oceanogr. **4:**128–139.
19. **Junge, K., F. Imhoff, T. Staley, and J. W. Deming.** 2002. Phylogenetic diversity of numerically important Arctic sea-ice bacteria cultured at subzero temperature. Microb. Ecol. **43:**315–328.
20. **Juretschko, S., A. Loy, A. Lehner, and M. Wagner.** 2002. The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. Syst. Appl. Microbiol. **25:**84–99.
21. **Kanawaga, T.** 2003. Bias and artifacts in multitemplate polymerase chain reaction. J. Biosci. Bioeng. **96:**317–323.
22. **Kelly, K. M., and A. Y. Chistoserdov.** 2001. Phylogenetic analysis of the

succession of bacterial communities in the Great South Bay (Long Island). FEMS Microbiol. Ecol. **35:**85–95.
23. **Krzanowski, W. J.** 2000. Principles of multivariate analysis. A user's perspective. Oxford University Press, Oxford, United Kingdom.
24. **Ley, R. E., F. Backhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, and J. I. Gordon.** 2005. Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. USA **102:**11070–11075.
25. **Li, L., C. Kato, and K. Horikoshi.** 1999. Microbial diversity in sediments collected from the deepest cold-seep area, the Japan Trench. Mar. Biotechnol. (New York) **1:**391–400.
26. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer.** 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32:**1363–1371.
27. **Magurran, A. E.** 1988. Ecological diversity and its measurement. Princeton University Press, Princeton, N.J.
28. **Magurran, A. E.** 2004. Measuring biological diversity. Blackwell, Oxford, United Kingdom.
29. **Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, and J. M. Tiedje.** 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. **29:**173–174.
30. **Martin, A. P.** 2002. Phylogenetic approaches for describing and comparing the diversity of microbial communities. Appl. Environ. Microbiol. **68:**3673–3682.
31. **Massana, R., A. E. Murray, C. M. Preston, and E. F. DeLong.** 1997. Vertical distribution and phylogenetic characterization of marine planktonic archaea in the Santa Barbara Channel. Appl. Environ. Microbiol. **63:**50–56.
32. **Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni.** 1995. Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnol. Oceanogr. **40:**148–158.
33. **Nübel, U., F. Garcia-Pichel, M. Kuhl, and G. Muyzer.** 1999. Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. Appl. Environ. Microbiol. **65:**422–430.
34. **Pace, N. R.** 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734–740.
35. **Pace, N. R., D. A. Stahl, D. J. Lane, and G. J. Olsen.** 1986. The analysis of natural microbial populations by ribosomal RNA sequences. Adv. Microb. Ecol. **9:**1–55.
36. **Pace, N. R., D. A. Stahl, D. J. Lane, and G. J. Olsen.** 1985. Analyzing natural microbial populations by rRNA sequences. ASM News **51:**4–12.
37. **Rappé, M. S., and S. J. Giovannoni.** 2003. The uncultured microbial majority. Annu. Rev. Microbiol. **57:**369–394.
38. **Ravenschlag, K., K. Sahm, J. Pernthaler, and R. Amann.** 1999. High bacterial diversity in permanently cold marine sediments. Appl. Environ. Microbiol. **65:**3982–3989.
39. **Schloss, P. D., B. R. Larget, and J. Handelsman.** 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. Appl. Environ. Microbiol. **70:**5485–5492.
40. **Singleton, D. R., M. A. Furlong, S. L. Rathbun, and W. B. Whitman.** 2001. Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. Appl. Environ. Microbiol. **67:**4374–4376.
41. **Spear, J. R., J. J. Walker, T. M. McCollom, and N. R. Pace.** 2005. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. Proc. Natl. Acad. Sci. USA **102:**2555–2560.
42. **Staley, J. T., and J. J. Gosink.** 1999. Poles apart: biodiversity and biogeography of sea ice bacteria. Annu. Rev. Microbiol. **53:**189–215.
43. **von Wintzingerode, F., U. B. Göbel, and E. Stackebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. **21:**213–229.
44. **Wheeler, P. A., M. Gosselin, E. Sherr, D. Thibault, D. L. Kirchman, R. Benner, and T. E. Whitledge.** 1996. Active cycling of organic carbon in the central Arctic Ocean. Nature **380:**697–699.
45. **Zwart, G., W. D. Hiorns, B. A. Methe, M. P. Van Agterveld, R. Huismans, S. C. Nold, J. P. Zehr, and H. J. Laanbroek.** 1998. Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. Syst. Appl. Microbiol. **21:**546–556.

QIIME allows analysis of high-throughput community sequencing data

To the Editor: High-throughput sequencing is revolutionizing microbial ecology studies. Efforts like the Human Microbiome Projects[1] and the US National Ecological Observatory Network[2] are helping us to understand the role of microbial diversity in habitats within our own bodies and throughout the planet.

Pyrosequencing using error-correcting, sample-specific barcodes allows hundreds of communities to be analyzed simultaneously in multiplex[3]. Integrating information from thousands of samples, including those obtained from time series, can reveal large-scale patterns that were inaccessible with lower-throughput sequencing methods. However, a major barrier to achieving such insights has been the lack of software that can handle these increasingly massive datasets. Although tools exist to perform library demultiplexing and taxonomy assignment[4,5], tools for downstream analyses are scarce.

Here we describe 'quantitative insights into microbial ecology' (QIIME; prounounced 'chime'), an open-source software pipeline built using the PyCogent toolkit[6], to address the problem of taking sequencing data from raw sequences to interpretation and database deposition. QIIME, available at http://qiime.sourceforge.net/, supports a wide range of microbial community analyses and visualizations that have been central to several recent high-profile studies, including network analysis, histograms of within- or between-sample diversity and analysis of whether 'core' sets of organisms are consistently represented in certain habitats. QIIME also provides graphical displays that allow users to interact with the data. Our implementation is highly modular and makes extensive use of unit testing to ensure the accuracy of results. This modularity allows alternative components for functionalities such as choosing operational taxonomic units (OTUs), sequence alignment, inferring phylogenetic trees and phylogenetic and taxon-based analysis of diversity within and between samples (including incorporation of third-party applications for many steps) to be easily integrated and benchmarked against one another (**Supplementary Fig. 1**).

We applied the QIIME workflow to a combined analysis of previously collected data (see **Supplementary Discussion**) for distal gut bacterial communities from conventionally raised mice, adult

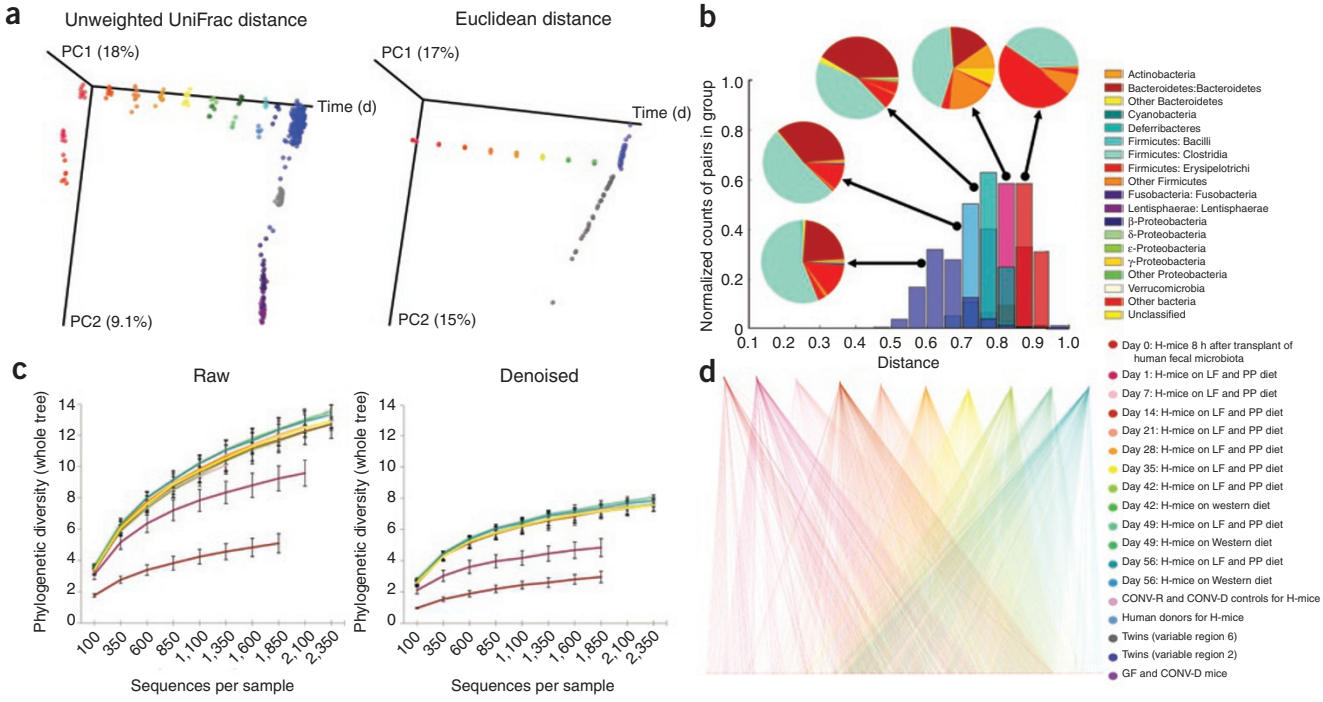


Figure 1 | QIIME analyses of the distal gut microbiotas of conventionally raised and conventionalized mice, gnotobiotic mice colonized with a human fecal gut microbiota (H-mice), and human adult mono- and dizygotic twins. (**a**) Principal coordinates analysis plots for mice, H-mice and twins. Colors correspond to separate samples by species and time point, and are consistent throughout the panels. (**b**) Unweighted UniFrac distance histograms between the data for fecal microbiota of human twins; human donors for the H-mice study; day 56 post-transplant H-mice on a low-fat (LF) and plant polysaccharide–rich (PP) diet; day 1 H-mice (LF and PP diet); and day 0 H-mice. Taxonomic classifications are presented at the class level. (**c**) Alpha diversity rarefaction plots of phylogenetic diversity for the H-mice samples. (**d**) OTU network connectivity of H-mice time series data. CONV-D, conventionalized mice; CONV-R, conventionally raised mice; and GF, germ-free mice.

human monozygotic and dizygotic twins and their mothers, and a time series study of adult germ-free mice after they received human fecal microbiota (**Fig. 1**, **Supplementary Table 1** and **Supplementary Discussion**). This analysis combined ten full 454 FLX runs and one partial run, totalling 3.8 million bacterial 16S rRNA sequences from previously published studies, including reads from different regions of the 16S rRNA gene.

QIIME is thus a robust platform for combining heterogeneous experimental datasets and for rapidly obtaining new insights about various microbial communities. Because QIIME scales to millions of sequences and can be used on platforms from laptops to high-performance computing clusters, we expect it to keep pace with advances in sequencing technology and to facilitate characterization of microbial community patterns ranging from normal variations to pathological disturbances in many human, animal and other environmental ecosystems.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank our collaborators for their helpful suggestions on features, documentation and the manuscript, and our funding agencies for their commitment to open-source software. This work was supported in part by Howard Hughes Medical Institute and grants from the Crohn's and Colitis Foundation of America, the German Academic Exchange Service, the Bill and Melinda Gates Foundation, the Colorado Center for Biofuels and Biorefining and the US National Institutes of Health (DK78669, GM65103, GM8759, HG4872 and its ARRA supplement, HG4866, DK83981 and LM9451).

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

J Gregory Caporaso[1,12], Justin Kuczynski[2,12], Jesse Stombaugh[1,12], Kyle Bittinger[3], Frederic D Bushman[3], Elizabeth K Costello[1], Noah Fierer[4], Antonio Gonzalez Peña[5], Julia K Goodrich[5], Jeffrey I Gordon[6], Gavin A Huttley[7], Scott T Kelley[8], Dan Knights[5], Jeremy E Koenig[9], Ruth E Ley[9], Catherine A Lozupone[1], Daniel McDonald[1], Brian D Muegge[6], Meg Pirrung[1], Jens Reeder[1], Joel R Sevinsky[10], Peter J Turnbaugh[6], William A Walters[2], Jeremy Widmann[1], Tanya Yatsunenko[6], Jesse Zaneveld[2] & Rob Knight[1,11]

[1]Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado, USA. [2]Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado, USA. [3]Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [4]Cooperative Institute for Research in Environmental Sciences and Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado, USA. [5]Department of Computer Science, University of Colorado, Boulder, Colorado, USA. [6]Center for Genome Sciences, Washington University School of Medicine, St. Louis, Missouri, USA. [7]Computational Genomics Laboratory, John Curtin School of Medical Research, The Australian National University, Canberra, Australian Capital Territory, Australia. [8]Department of Biology, San Diego State University, San Diego, California, USA. [9]Department of Microbiology, Cornell University, Ithaca, New York, USA. [10]Luca Technologies, Golden, Colorado, USA. [11]Howard Hughes Medical Institute, Boulder, Colorado, USA. [12]These authors contributed equally to this work.
e-mail: rob.knight@colorado.edu

PUBLISHED ONLINE 11 APRIL 2010; DOI:10.1038/NMETH.F.303

1. National Institutes of Health Human Microbiome Project Working Group *et al. Genome Res.* **19**, 2317–2323 (2009).
2. Hopkin, M. *Nature* **444**, 420–421 (2006).
3. Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J. & Knight, R. *Nat. Methods* **5**, 235–237 (2008).
4. Cole, J.R. *et al. Nucleic Acids Res.* **37**, D141–D145 (2009).
5. Schloss, P.D. *et al. Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
6. Knight, R. *et al. Genome Biol.* **8**, R171 (2007).

Intensity normalization improves color calling in SOLiD sequencing

To the Editor: Applied Biosystems' SOLiD system[1] is a commonly used massively parallel DNA sequencing platform for applications from genotyping and structural variation analysis[1] to transcriptome quantification and reconstruction[2]. Like other sequencing technologies, it measures fluorescence intensities from dye-labeled molecules to determine the sequence of DNA fragments. Ultimately, sequences are determined by complicated statistical manipulations of noisy intensity measurements, and systematic biases may mislead downstream analysis[3]. Several proposed methods improve base calling and quality metrics for other sequencing technologies[3–5], and we now present Rsolid, software implementing an intensity normalization strategy for the SOLiD platform that substantially improves yield and accuracy at small computational costs (6% increase in total matches, 13% increase in perfect matches, 5% reduced error rate and a substantial reduction in false positive single-nucleotide polymorphism (SNP) calls in an *Escherichia coli* genomic DNA sample).

In the SOLiD system, the proportions of color calls across sequencing cycles are extremely variable (**Fig. 1a**), even though they should be equal across sequencing cycles and proportional to the dinucleotide content of the library (**Supplementary Methods**). This bias can be traced to the fluorescence intensity measurements used to make the color calls (**Supplementary Fig. 1**). The distributions of intensities are similar across channels in early sequencing cycles, but a color bias starts to appear in later cycles. The Rsolid method uses a simple and computationally efficient procedure to normalize the color-channel

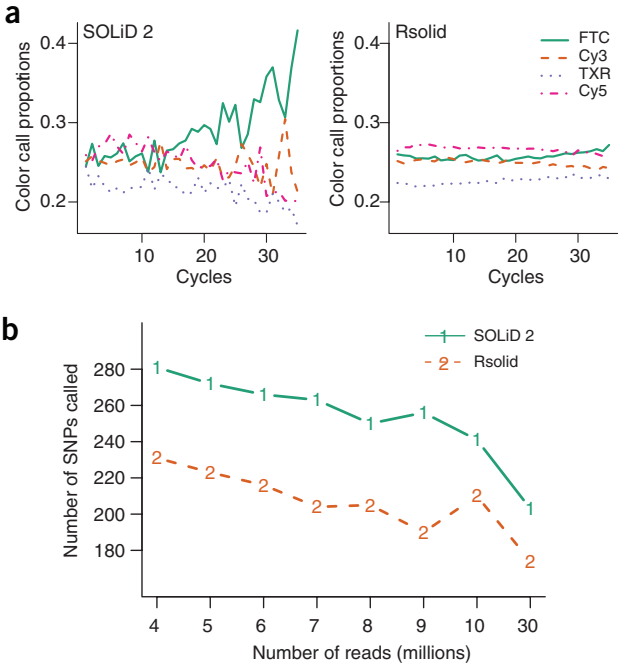


Figure 1 | Effect of normalization on color proportions and SNP calling. (**a**) Color proportions in sample of *E. coli* genomic DNA on each sequencing cycle. Color calls as reported by the SOLiD 2 system (left) and after normalization by Rsolid (right). FTX, TXR, Cy3 and Cy5 are dyes used by SOLiD. (**b**) Number of false positive SNPs called in *E. coli* at various coverage. After normalization, fewer SNPs were called even at high coverage (30 M reads correspond to ~100-fold coverage).

ARTICLES

A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin[1]*, Ruiqiang Li[1]*, Jeroen Raes[2,3], Manimozhiyan Arumugam[2], Kristoffer Solvsten Burgdorf[4], Chaysavanh Manichanh[5], Trine Nielsen[4], Nicolas Pons[6], Florence Levenez[6], Takuji Yamada[2], Daniel R. Mende[2], Junhua Li[1,7], Junming Xu[1], Shaochuan Li[1], Dongfang Li[1,8], Jianjun Cao[1], Bo Wang[1], Huiqing Liang[1], Huisong Zheng[1], Yinlong Xie[1,7], Julien Tap[6], Patricia Lepage[6], Marcelo Bertalan[9], Jean-Michel Batto[6], Torben Hansen[4], Denis Le Paslier[10], Allan Linneberg[11], H. Bjørn Nielsen[9], Eric Pelletier[10], Pierre Renault[6], Thomas Sicheritz-Ponten[9], Keith Turner[12], Hongmei Zhu[1], Chang Yu[1], Shengting Li[1], Min Jian[1], Yan Zhou[1], Yingrui Li[1], Xiuqing Zhang[1], Songgang Li[1], Nan Qin[1], Huanming Yang[1], Jian Wang[1], Søren Brunak[9], Joel Doré[6], Francisco Guarner[5], Karsten Kristiansen[13], Oluf Pedersen[4,14], Julian Parkhill[12], Jean Weissenbach[10], MetaHIT Consortium†, Peer Bork[2], S. Dusko Ehrlich[6] & Jun Wang[1,13]

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.

It has been estimated that the microbes in our bodies collectively make up to 100 trillion cells, tenfold the number of human cells, and suggested that they encode 100-fold more unique genes than our own genome[1]. The majority of microbes reside in the gut, have a profound influence on human physiology and nutrition, and are crucial for human life[2,3]. Furthermore, the gut microbes contribute to energy harvest from food, and changes of gut microbiome may be associated with bowel diseases or obesity[4–8].

To understand and exploit the impact of the gut microbes on human health and well-being it is necessary to decipher the content, diversity and functioning of the microbial gut community. 16S ribosomal RNA gene (rRNA) sequence-based methods[9] revealed that two bacterial divisions, the Bacteroidetes and the Firmicutes, constitute over 90% of the known phylogenetic categories and dominate the distal gut microbiota[10]. Studies also showed substantial diversity of the gut microbiome between healthy individuals[4,8,10,11]. Although this difference is especially marked among infants[12], later in life the gut microbiome converges to more similar phyla.

Metagenomic sequencing represents a powerful alternative to rRNA sequencing for analysing complex microbial communities[13–15]. Applied to the human gut, such studies have already generated some 3 gigabases (Gb) of microbial sequence from faecal samples of 33 individuals from the United States or Japan[8,16,17]. To get a broader overview of the human gut microbial genes we used the Illumina Genome Analyser (GA) technology to carry out deep sequencing of total DNA from faecal samples of 124 European adults. We generated 576.7 Gb of sequence, almost 200 times more than in all previous studies, assembled it into contigs and predicted 3.3 million unique open reading frames (ORFs). This gene catalogue contains virtually all of the prevalent gut microbial genes in our cohort, provides a broad view of the functions important for bacterial life in the gut and indicates that many bacterial species are shared by different individuals. Our results also show that short-read metagenomic sequencing can be used for global characterization of the genetic potential of ecologically complex environments.

Metagenomic sequencing of gut microbiomes

As part of the MetaHIT (Metagenomics of the Human Intestinal Tract) project, we collected faecal specimens from 124 healthy, overweight and obese individual human adults, as well as inflammatory bowel disease (IBD) patients, from Denmark and Spain (Supplementary Table 1). Total DNA was extracted from the faecal specimens[18] and an average of 4.5 Gb (ranging between 2 and 7.3 Gb) of sequence was generated for each sample, allowing us to capture most of the

[1]BGI-Shenzhen, Shenzhen 518083, China. [2]European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [3]VIB—Vrije Universiteit Brussel, 1050 Brussels, Belgium. [4]Hagedorn Research Institute, DK 2820 Copenhagen, Denmark. [5]Hospital Universitari Val d'Hebron, Ciberehd, 08035 Barcelona, Spain. [6]Institut National de la Recherche Agronomique, 78350 Jouy en Josas, France. [7]School of Software Engineering, South China University of Technology, Guangzhou 510641, China. [8]Genome Research Institute, Shenzhen University Medical School, Shenzhen 518000, China. [9]Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. [10]Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France. [11]Research Center for Prevention and Health, DK-2600 Glostrup, Denmark. [12]The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [13]Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. [14]Institute of Biomedical Sciences, University of Copenhagen & Faculty of Health Science, University of Aarhus, 8000 Aarhus, Denmark.
*These authors contributed equally to this work.
†Lists of authors and affiliations appear at the end of the paper.

novelty (see Methods and Supplementary Table 2). In total, we obtained 576.7 Gb of sequence (Supplementary Table 3).

Wanting to generate an extensive catalogue of microbial genes from the human gut, we first assembled the short Illumina reads into longer contigs, which could then be analysed and annotated by standard methods. Using SOAPdenovo[19], a de Bruijn graph-based tool specially designed for assembling very short reads, we performed *de novo* assembly for all of the Illumina GA sequence data. Because a high diversity between individuals is expected[8,16,17], we first assembled each sample independently (Supplementary Fig. 3). As much as 42.7% of the Illumina GA reads was assembled into a total of 6.58 million contigs of a length >500 bp, giving a total contig length of 10.3 Gb, with an N50 length of 2.2 kb (Supplementary Fig. 4) and the range of 12.3 to 237.6 Mb (Supplementary Table 4). Almost 35% of reads from any one sample could be mapped to contigs from other samples, indicating the existence of a common sequence core.

To assess the quality of the Illumina GA-based assembly we mapped the contigs of samples MH0006 and MH0012 to the Sanger reads from the same samples (Supplementary Table 2). A total of 98.7% of the contigs that map to at least one Sanger read were collinear over 99.6% of the mapped regions. This is comparable to the contigs that were generated by 454 sequencing for one of the two samples (MH0006) as a control, of which 97.9% were collinear over 99.5% of the mapped regions. We estimate assembly errors to be 14.2 and 20.7 per megabase (Mb) of Illumina- and 454-based contigs, respectively (see Methods and Supplementary Fig. 5), indicating that the short- and long-read-based assemblies have comparable accuracies.

To complete the contig set we pooled the unassembled reads from all 124 samples, and repeated the *de novo* assembly process. About 0.4 million additional contigs were thus generated, having a length of 370 Mb and an N50 length of 939 bp. The total length of our final contig set was thus 10.7 Gb. Some 80% of the 576.7 Gb of Illumina GA sequence could be aligned to the contigs at a threshold of 90% identity, allowing for accommodation of sequencing errors and strain variability in the gut (Fig. 1), almost twice the 42.7% of sequence that was assembled into contigs by SOAPdenovo, because assembly uses more stringent criteria. This indicates that a vast majority of the Illumina sequence is represented by our contigs.

To compare the representation of the human gut microbiome in our contigs with that from previous work, we aligned them to the reads from the two largest published gut metagenome studies (1.83 Gb of Roche/454 sequencing reads from 18 US adults[8], and 0.79 Gb of Sanger reads from 13 Japanese adults and infants[17]), using the 90% identity threshold. A total of 70.1% and 85.9% of the reads from the Japanese and US samples, respectively, could be aligned to our contigs (Fig. 1), showing that the contigs include a high fraction of sequences from previous studies. In contrast, 85.7% and 69.5% of our contigs were not covered by the reads from the Japanese and US samples, respectively, highlighting the novelty we captured.

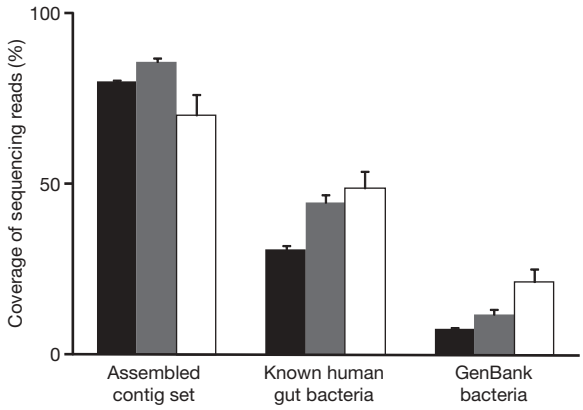


Figure 1 | Coverage of human gut microbiome. The three human microbial sequencing read sets—Illumina GA reads generated from 124 individuals in this study (black; $n = 124$), Roche/454 reads from 18 human twins and their mothers (grey; $n = 18$) and Sanger reads from 13 Japanese individuals (white; $n = 13$)—were aligned to each of the reference sequence sets. Mean values ± s.e.m. are plotted.

Only 31.0–48.8% of the reads from the two previous studies and the present study could be aligned to 194 public human gut bacterial genomes (Supplementary Table 5), and 7.6–21.2% to the bacterial genomes deposited in GenBank (Fig. 1). This indicates that the reference gene set obtained by sequencing genomes of isolated bacterial strains is still of a limited scale.

A gene catalogue of the human gut microbiome

To establish a non-redundant human gut microbiome gene set we first used the MetaGene[20] program to predict ORFs in our contigs and found 14,048,045 ORFs longer than 100 bp (Supplementary Table 6). They occupied 86.7% of the contigs, comparable to the value found for fully sequenced genomes (~86%). Two-thirds of the ORFs appeared incomplete, possibly due to the size of our contigs (N50 of 2.2 kb). We next removed the redundant ORFs, by pair-wise comparison, using a very stringent criterion of 95% identity over 90% of the shorter ORF length, which can fuse orthologues but avoids inflation of the data set due to possible sequencing errors (see Methods). Yet, the final non-redundant gene set contained as many as 3,299,822 ORFs with an average length of 704 bp (Supplementary Table 7).

We term the genes of the non-redundant set 'prevalent genes', as they are encoded on contigs assembled from the most abundant reads (see Methods). The minimal relative abundance of the prevalent genes was $\sim 6 \times 10^{-7}$, as estimated from the minimum sequence coverage of the unique genes (close to 3), and the total Illumina sequence length generated for each individual (on average, 4.5 Gb), assuming the average gene length of 0.85 kb (that is, $3 \times 0.85 \times 10^3 / 4.5 \times 10^9$).

We mapped the 3.3 million gut ORFs to the 319,812 genes (target genes) of the 89 frequent reference microbial genomes in the human gut. At a 90% identity threshold, 80% of the target genes had at least 80% of their length covered by a single gut ORF (Fig. 2b). This indicates that the gene set includes most of the known human gut bacterial genes.

We examined the number of prevalent genes identified across all individuals as a function of the extent of sequencing, demanding at least two supporting reads for a gene call (Fig. 2a). The incidence-based coverage richness estimator (ICE), determined at 100 individuals (the highest number the EstimateS[21] program could accommodate), indicates that our catalogue captures 85.3% of the prevalent genes. Although this is probably an underestimate, it nevertheless indicates that the catalogue contains an overwhelming majority of the prevalent genes of the cohort.

Each individual carried 536,112 ± 12,167 (mean ± s.e.m.) prevalent genes (Supplementary Fig. 6b), indicating that most of the 3.3 million gene pool must be shared. However, most of the prevalent genes were found in only a few individuals: 2,375,655 were present in less than 20%, whereas 294,110 were found in at least 50% of individuals (we term these 'common' genes). These values depend on the sampling depth; sequencing of MH0006 and MH0012 revealed more of the catalogue genes, present at a low abundance (Supplementary Fig. 7). Nevertheless, even at our routine sampling depth, each individual harboured 204,056 ± 3,603 (mean ± s.e.m.) common genes, indicating that about 38% of an individual's total gene pool is shared. Interestingly, the IBD patients harboured, on average, 25% fewer genes than the individuals not suffering from IBD (Supplementary Fig. 8), consistent with the observation that the former have lower bacterial diversity than the latter[22].

Common bacterial core

Deep metagenomic sequencing provides the opportunity to explore the existence of a common set of microbial species (common core) in

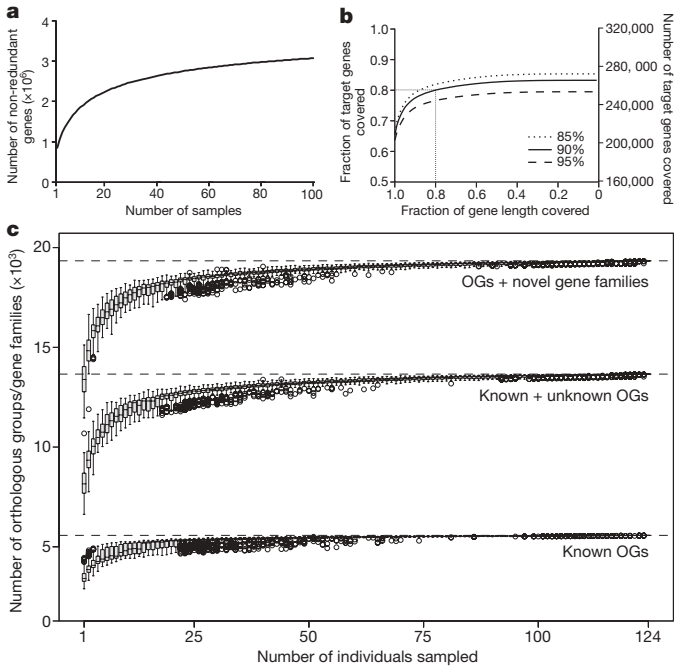


Figure 2 | Predicted ORFs in the human gut microbiome. **a**, Number of unique genes as a function of the extent of sequencing. The gene accumulation curve corresponds to the S_{obs} (Mao Tau) values (number of observed genes), calculated using EstimateS[21] (version 8.2.0) on randomly chosen 100 samples (due to memory limitation). **b**, Coverage of genes from 89 frequent gut microbial species (Supplementary Table 12). **c**, Number of functions captured by number of samples investigated, based on known (well characterized) orthologous groups (OGs; bottom), known plus unknown orthologous groups (including, for example, putative, predicted, conserved hypothetical functions; middle) and orthologous groups plus novel gene families (>20 proteins) recovered from the metagenome (top). Boxes denote the interquartile range (IQR) between the first and third quartiles (25th and 75th percentiles, respectively) and the line inside denotes the median. Whiskers denote the lowest and highest values within 1.5 times IQR from the first and third quartiles, respectively. Circles denote outliers beyond the whiskers.

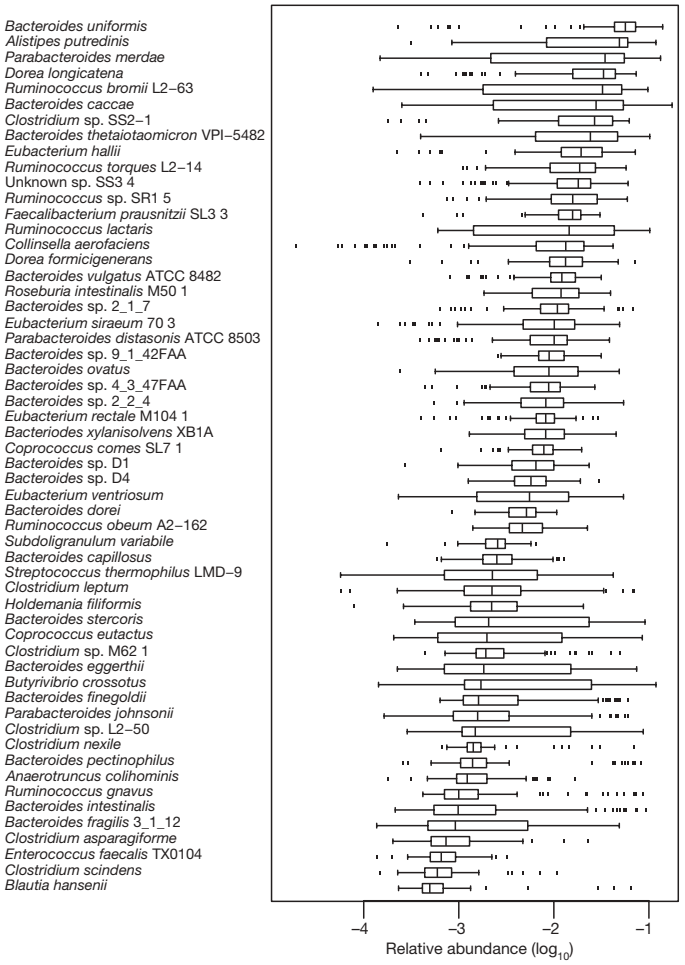


Figure 3 | Relative abundance of 57 frequent microbial genomes among individuals of the cohort. See Fig. 2c for definition of box and whisker plot. See Methods for computation.

the cohort. For this purpose, we used a non-redundant set of 650 sequenced bacterial and archaeal genomes (see Methods). We aligned the Illumina GA reads of each human gut microbial sample onto the genome set, using a 90% identity threshold, and determined the proportion of the genomes covered by the reads that aligned onto only a single position in the set. At a 1% coverage, which for a typical gut bacterial genome corresponds to an average length of about 40 kb, some 25-fold more than that of the 16S gene generally used for species identification, we detected 18 species in all individuals, 57 in ≥90% and 75 in ≥50% of individuals (Supplementary Table 8). At 10% coverage, requiring ~10-fold higher abundance in a sample, we still found 13 of the above species in ≥90% of individuals and 35 in ≥50%.

When the cumulated sequence length increased from 3.96 Gb to 8.74 Gb and from 4.41 Gb to 11.6 Gb, for samples MH0006 and MH0012, respectively, the number of strains common to the two at the 1% coverage threshold increased by 25%, from 135 to 169. This indicates the existence of a significantly larger common core than the one we could observe at the sequence depth routinely used for each individual.

The variability of abundance of microbial species in individuals can greatly affect identification of the common core. To visualize this variability, we compared the number of sequencing reads aligned to different genomes across the individuals of our cohort. Even for the most common 57 species present in ≥90% of individuals with genome coverage >1% (Supplementary Table 8), the inter-individual variability was between 12- and 2,187-fold (Fig. 3). As expected[10,23], Bacteroidetes and Firmicutes had the highest abundance.

A complex pattern of species relatedness, characterized by clusters at the genus and family levels, emerges from the analysis of the network based on the pair-wise Pearson correlation coefficients of 155 species present in at least one individual at ≥1% coverage (Supplementary Fig. 9). Prominent clusters include some of the most abundant gut species, such as members of the Bacteroidetes and *Dorea*/*Eubacterium*/*Ruminococcus* groups and also bifidobacteria, Proteobacteria and streptococci/lactobacilli groups. These observations indicate that similar constellations of bacteria may be present in different individuals of our cohort, for reasons that remain to be established.

The above result indicates that the Illumina-based bacterial profiling should reveal differences between the healthy individuals and patients. To test this hypothesis we compared the IBD patients and healthy controls (Supplementary Table 1), as it was previously reported that the two have different microbiota[22]. The principal component analysis, based on the same 155 species, clearly separates patients from healthy individuals and the ulcerative colitis from the Crohn's disease patients (Fig. 4), confirming our hypothesis.

Functions encoded by the prevalent gene set

We classified the predicted genes by aligning them to the integrated NCBI-NR database of non-redundant protein sequences, the genes in the KEGG (Kyoto Encyclopedia of Genes and Genomes)[24] pathways, and COG (Clusters of Orthologous Groups)[25] and eggNOG[26] databases. There were 77.1% genes classified into phylotypes, 57.5% to eggNOG clusters, 47.0% to KEGG orthology and 18.7% genes assigned to KEGG pathways, respectively (Supplementary Table 9).

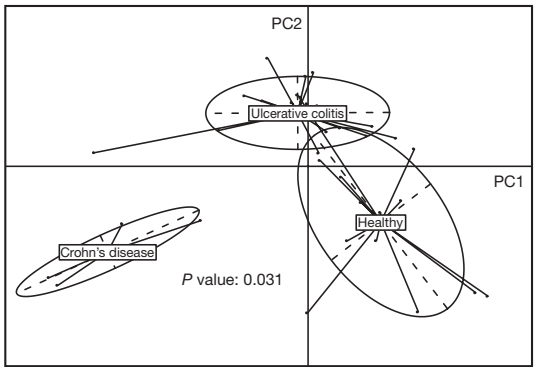


Figure 4 | Bacterial species abundance differentiates IBD patients and healthy individuals. Principal component analysis with health status as instrumental variables, based on the abundance of 155 species with ≥1% genome coverage by the Illumina reads in at least 1 individual of the cohort, was carried out with 14 healthy individuals and 25 IBD patients (21 ulcerative colitis and 4 Crohn's disease) from Spain (Supplementary Table 1). Two first components (PC1 and PC2) were plotted and represented 7.3% of whole inertia. Individuals (represented by points) were clustered and centre of gravity computed for each class; *P*-value of the link between health status and species abundance was assessed using a Monte-Carlo test (999 replicates).

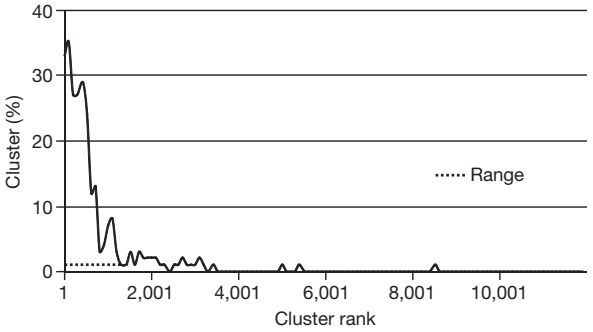


Figure 5 | Clusters that contain the *B. subtilis* essential genes. The clusters were ranked by the number of genes they contain, normalized by average length and copy number (see Supplementary Fig. 10), and the proportion of clusters with the essential *B. subtilis* genes was determined for successive groups of 100 clusters. Range indicates the part of the cluster distribution that contains 86% of the *B. subtilis* essential genes.

Almost all (99.96%) of the phylogenetically assigned genes belonged to the Bacteria and Archaea, reflecting their predominance in the gut. Genes that were not mapped to orthologous groups were clustered into gene families (see Methods). To investigate the functional content of the prevalent gene set we computed the total number of orthologous groups and/or gene families present in any combination of n individuals (with $n = 2$–124; see Fig. 2c). This rarefaction analysis shows that the 'known' functions (annotated in eggNOG or KEGG) quickly saturate (a value of 5,569 groups was observed): when sampling any subset of 50 individuals, most have been detected. However, three-quarters of the prevalent gut functionalities consists of uncharacterized orthologous groups and/or completely novel gene families (Fig. 2c). When including these groups, the rarefaction curve only starts to plateau at the very end, at a much higher level (19,338 groups were detected), confirming that the extensive sampling of a large number of individuals was necessary to capture this considerable amount of novel/unknown functionality.

Bacterial functions important for life in the gut

The extensive non-redundant catalogue of the bacterial genes from the human intestinal tract provides an opportunity to identify bacterial functions important for life in this environment. There are functions necessary for a bacterium to thrive in a gut context (that is, the 'minimal gut genome') and those involved in the homeostasis of the whole ecosystem, encoded across many species (the 'minimal gut metagenome'). The first set of functions is expected to be present in most or all gut bacterial species; the second set in most or all individuals' gut samples.

To identify the functions encoded by the minimal gut genome we use the fact that they should be present in most or all gut bacterial species and therefore appear in the gene catalogue at a frequency above that of the functions present in only some of the gut bacterial species. The relative frequency of different functions can be deduced from the number of genes recruited to different eggNOG clusters, after normalization for gene length and copy number (Supplementary Fig. 10a, b). We ranked all the clusters by gene frequencies and determined the range that included the clusters specifying well-known essential bacterial functions, such as those determined experimentally for a well-studied firmicute, *Bacillus subtilis*[27], hypothesizing that additional clusters in this range are equally important. As expected, the range that included most of *B. subtilis* essential clusters (86%) was at the very top of the ranking order (Fig. 5). Some 76% of the clusters with essential genes of *Escherichia coli*[28] were within this range, confirming the validity of our approach. This suggests that 1,244 metagenomic clusters found within the range (Supplementary Table 10; termed 'range clusters' hereafter) specify functions important for life in the gut.

We found two types of functions among the range clusters: those required in all bacteria (housekeeping) and those potentially specific for the gut. Among many examples of the first category are the functions that are part of main metabolic pathways (for example, central carbon metabolism, amino acid synthesis), and important protein complexes (RNA and DNA polymerase, ATP synthase, general secretory apparatus). Not surprisingly, projection of the range clusters on the KEGG metabolic pathways gives a highly integrated picture of the global gut cell metabolism (Fig. 6a).

The putative gut-specific functions include those involved in adhesion to the host proteins (collagen, fibrinogen, fibronectin) or in harvesting sugars of the globoseries glycolipids, which are carried on blood and epithelial cells. Furthermore, 15% of range clusters encode functions that are present in <10% of the eggNOG genomes (see Supplementary Fig. 11) and are largely (74.3%) not defined (Fig. 6b). Detailed studies of these should lead to a deeper comprehension of bacterial life in the gut.

To identify the functions encoded by the minimal gut metagenome, we computed the orthologous groups that are shared by individuals of our cohort. This minimal set, of 6,313 functions, is much larger than the one estimated in a previous study[8]. There are only 2,069 functionally annotated orthologous groups, showing that they gravely underestimate the true size of the common functional complement among individuals (Fig. 6c). The minimal gut metagenome includes a considerable fraction of functions (~45%) that are present in <10% of the sequenced bacterial genomes (Fig. 6c, inset). These otherwise rare functionalities that are found in each of the 124 individuals may be necessary for the gut ecosystem. Eighty per cent of these orthologous groups contain genes with at best poorly characterized function, underscoring our limited knowledge of gut functioning.

Of the known fraction, about 5% codes for (pro)phage-related proteins, implying a universal presence and possible important ecological role of bacteriophages in gut homeostasis. The most striking secondary metabolism that seems crucial for the minimal metagenome relates, not unexpectedly, to biodegradation of complex sugars and glycans harvested from the host diet and/or intestinal lining. Examples include degradation and uptake pathways for pectin (and its monomer, rhamnose) and sorbitol, sugars which are omnipresent in fruits and vegetables, but which are not or poorly absorbed by humans. As some gut microorganisms were found to degrade both of them[29,30], this capacity seems to be selected for by the gut ecosystem as a non-competitive source of energy. Besides these, capacity to ferment, for example, mannose, fructose, cellulose and sucrose is also part of the minimal metagenome. Together, these emphasize the

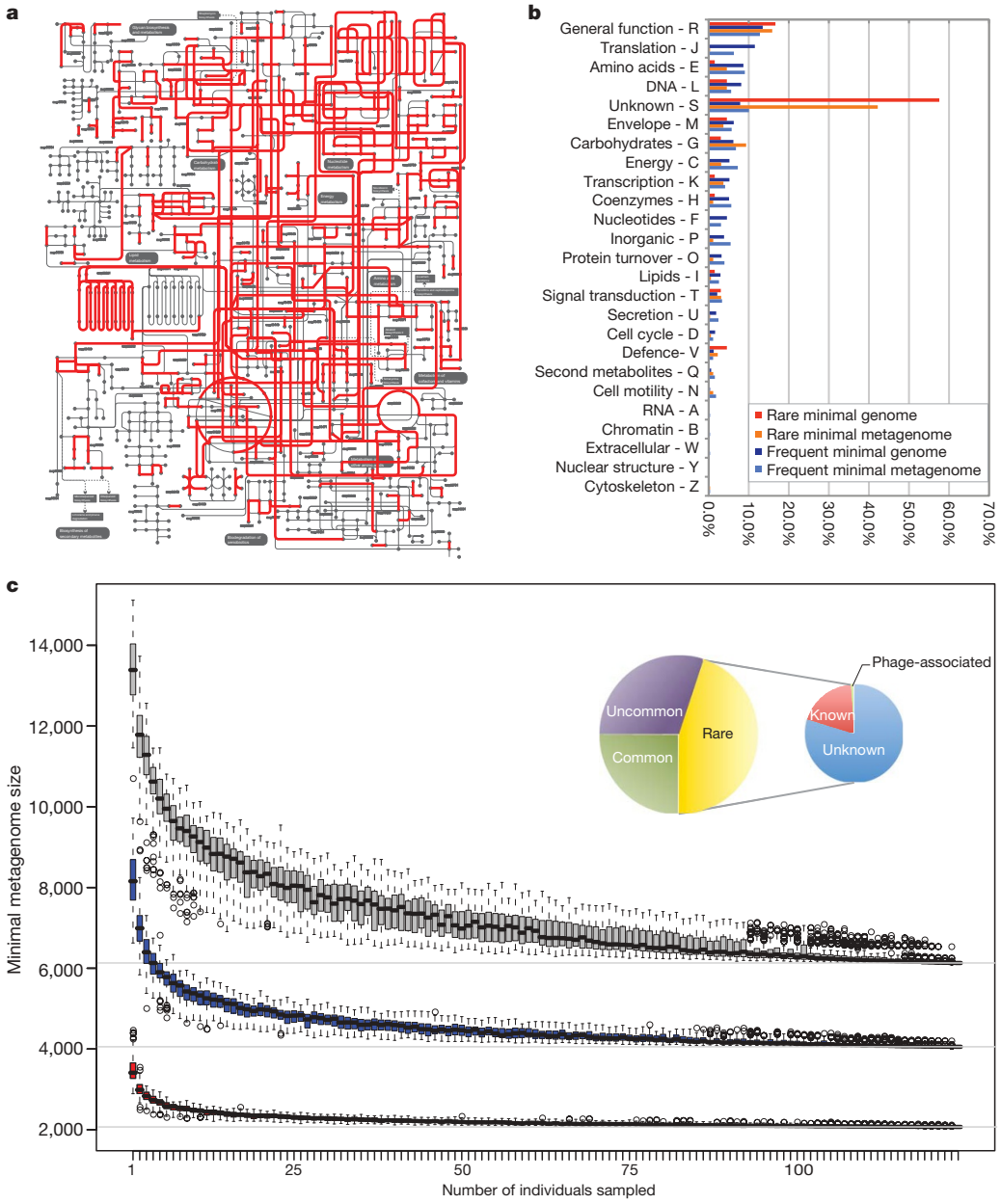


Figure 6 | Characterization of the minimal gut genome and metagenome. **a**, Projection of the minimal gut genome on the KEGG pathways using the iPath tool[38]. **b**, Functional composition of the minimal gut genome and metagenome. Rare and frequent refer to the presence in sequenced eggNOG genomes. **c**, Estimation of the minimal gut metagenome size. Known orthologous groups (red), known plus unknown orthologous groups (blue) and orthologous groups plus novel gene families (>20 proteins; grey) are shown (see Fig. 2c for definition of box and whisker plot). The inset shows composition of the gut minimal microbiome. Large circle: classification in the minimal metagenome according to orthologous group occurrence in STRING7[39] bacterial genomes. Common (25%), uncommon (35%) and rare (45%) refer to functions that are present in >50%, <50% but >10%, and <10% of STRING bacteria genomes, respectively. Small circle: composition of the rare orthologous groups. Unknown (80%) have no annotation or are poorly characterized, whereas known bacterial (19%) and phage-related (1%) orthologous groups have functional description.

strong dependence of the gut ecosystem on complex sugar degradation for its functioning.

Functional complementarities of the genome and metagenome

Detailed analysis of the complementarities between the gut metagenome and the human genome is beyond the scope of the present work. To provide an overview, we considered two factors: conservation of the functions in the minimal metagenome and presence/absence of functions in one or the other (Supplementary Table 11). Gut bacteria use mostly fermentation to generate energy, converting sugars, in part, to short-chain fatty acid, that are used by the host as energy source. Acetate is important for muscle, heart and brain cells[31], propionate is used in host hepatic neoglucogenic processes, whereas, in addition, butyrate is important for enterocytes[32]. Beyond short-chain fatty acid, a number of amino acids are indispensable to humans[33] and can be provided by bacteria[34]. Similarly, bacteria can contribute certain vitamins[3] (for example, biotin, phylloquinone) to the host. All of the steps of biosynthesis of these molecules are encoded by the minimal metagenome.

Gut bacteria seem to be able to degrade numerous xenobiotics, including non-modified and halogenated aromatic compounds (Supplementary Table 11), even if the steps of most pathways are not part of the minimal metagenome and are found in a fraction of individuals only. A particularly interesting example is that of benzoate, which is a common food supplement, known as E211. Its degradation by the coenzyme-A ligation pathway, encoded in the minimal metagenome, leads to pimeloyl-coenzyme-A, which is a precursor of biotin, indicating that this food supplement can have a potentially beneficial role for human health.

Discussion

We have used extensive Illumina GA short-read-based sequencing of total faecal DNA from a cohort of 124 individuals of European (Nordic and Mediterranean) origin to establish a catalogue of non-redundant human intestinal microbial genes. The catalogue contains 3.3 million microbial genes, 150-fold more than the human gene complement, and includes an overwhelming majority (>86%) of prevalent genes harboured by our cohort. The catalogue probably contains a large majority of prevalent intestinal microbial genes in the human population, for the following reasons: (1) over 70% of the metagenomic reads from three previous studies, including American and Japanese individuals[8,16,17], can be mapped on our contigs; (2) about 80% of the microbial genes from 89 frequent gut reference genomes are present in our set. This result represents a proof of principle that short-read sequencing can be used to characterize complex microbiomes.

The full bacterial gene complement of each individual was not sampled in our work. Nevertheless, we have detected some 536,000 prevalent unique genes in each, out of the total of 3.3 million carried by our cohort. Inevitably, the individuals largely share the genes of the common pool. At the present depth of sequencing, we found that almost 40% of the genes from each individual are shared with at least half of the individuals of the cohort. Future studies of world-wide span, envisaged within the International Human Microbiome Consortium, will complete, as necessary, our gene catalogue and establish boundaries to the proportion of shared genes.

Essentially all (99.1%) of the genes of our catalogue are of bacterial origin, the remainder being mostly archaeal, with only 0.1% of eukaryotic and viral origins. The gene catalogue is therefore equivalent to that of some 1,000 bacterial species with an average-sized genome, encoding about 3,364 non-redundant genes. We estimate that no more than 15% of prevalent genes of our cohort may be missing from the catalogue, and suggest that the cohort harbours no more than ~1,150 bacterial species abundant enough to be detected by our sampling. Given the large overlap between microbial sequences in this and previous studies we suggest that the number of abundant intestinal bacterial species may be not much higher than that observed in our cohort. Each individual of our cohort harbours at least 160 such bacterial species, as estimated by the average prevalent gene number, and many must thus be shared.

We assigned about 12% of the reference set genes (404,000) to the 194 sequenced intestinal bacterial genomes, and can thus associate them with bacterial species. Sequencing of at least 1,000 human-associated bacterial genomes is foreseen within the International Human Microbiome Consortium, via the Human Microbiome Project and MetaHIT. This is commensurate with the number of dominant species in our cohort and expected more broadly in human gut, and should enable a much more extensive gene to species assignment. Nevertheless, we used the presently available sequenced genomes to explore further the concept of largely shared species among our cohort and identified 75 species common to >50% of individuals and 57 species common to >90%. These numbers are likely to increase with the number of sequenced reference strains and a deeper sampling. Indeed, a 2–3-fold increase in sequencing depth raised by 25% the number of species that we could detect as shared between two individuals. A large number of shared species supports the view that the prevalent human microbiome is of a finite and not overly large size.

How can this view be reconciled with that of a considerable interpersonal diversity of innumerable bacterial species in the gut, arising from most previous studies using the 16S RNA marker gene[4,8,10,11]? Possibly the depth of sampling of these studies was insufficient to reveal common species when present at low abundance, and emphasized the difference in the composition of a relatively few dominant species. We found a very high variability of abundance (12- to 2,200-fold) for the 57 most common species across the individuals of our cohort. Nevertheless, a recent 16S rRNA-based study concluded that a common bacterial species 'core', shared among at least 50% of individuals under study, exists[35].

Detailed comparisons of bacterial genes across the individuals of our cohort will be carried out in the future, within the context of the ongoing MetaHIT clinical studies of which they are part. Nevertheless, clustering of the genes in families allowed us to capture a virtually full functional potential of the prevalent gene set and revealed a considerable novelty, extending the functional categories by some 30% in regard to previous work[8]. Similarly, this analysis has revealed a functional core, conserved in each individual of the cohort, which reflects the full minimal human gut metagenome, encoded across many species and probably required for the proper functioning of the gut ecosystem. The size of this minimal metagenome exceeds several-fold that of the core metagenome reported previously[8]. It includes functions known to be important to the host–bacterial interaction, such as degradation of complex polysaccharides, synthesis of short-chain fatty acids, indispensable amino acids and vitamins. Finally, we also identified functions that we attribute to a minimal gut bacterial genome, likely to be required by any bacterium to thrive in this ecosystem. Besides general housekeeping functions, the minimal genome encompasses many genes of unknown function, rare in sequenced genomes and possibly specifically required in the gut.

Beyond providing the global view of the human gut microbiome, the extensive gene catalogue we have established enables future studies of association of the microbial genes with human phenotypes and, even more broadly, human living habits, taking into account the environment, including diet, from birth to old age. We anticipate that these studies will lead to a much more complete understanding of human biology than the one we presently have.

METHODS SUMMARY

Human faecal samples were collected, frozen immediately and DNA was purified by standard methods[22]. For all 124 individuals, paired-end libraries were constructed with different clone insert sizes and subjected to Illumina GA sequencing. All reads were assembled using SOAPdenovo[19], with specific parameter '−M 3' for metagenomics data. MetaGene was used for gene prediction. A non-redundant gene set was constructed by pair-wise comparison of all genes, using BLAT[36] under the criteria of identity >95% and overlap >90%. Gene taxonomic assignments were made on the basis of BLASTP[37] search (*e*-value $<1\times10^{-5}$) of the NCBI-NR database and 126 known gut bacteria genomes. Gene functional annotations were made by BLASTP search (*e*-value $<1\times10^{-5}$) with eggNOG and KEGG (v48.2) databases. The total and shared number of orthologous groups and/or gene families were computed using a random combination of n individuals (with $n=2$ to 124, 100 replicates per bin).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 August; accepted 23 December 2009.

1. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124,** 837–848 (2006).
2. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307,** 1915–1920 (2005).
3. Hooper, L. V., Midtvedt, T. & Gordon, J. I. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* **22,** 283–307 (2002).
4. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
5. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
6. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102,** 11070–11075 (2005).
7. Zhang, H. *et al.* Human gut microbiota in obesity and after gastric bypass. *Proc. Natl Acad. Sci. USA* **106,** 2365–2370 (2009).
8. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
9. Zoetendal, E. G., Akkermans, A. D. & De Vos, W. M. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl. Environ. Microbiol.* **64,** 3854–3859 (1998).
10. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308,** 1635–1638 (2005).

11. Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Rev. Microbiol.* **6,** 776–788 (2008).
12. Palmer, C., Bik, E. M., Digiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5,** e177 (2007).
13. Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38,** 525–552 (2004).
14. von Mering, C. *et al.* Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315,** 1126–1130 (2007).
15. Tringe, S. G. & Rubin, E. M. Metagenomics: DNA sequencing of environmental samples. *Nature Rev. Genet.* **6,** 805–814 (2005).
16. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312,** 1355–1359 (2006).
17. Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14,** 169–181 (2007).
18. Suau, A. *et al.* Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl. Environ. Microbiol.* **65,** 4799–4807 (1999).
19. Li, R. & Zhu, H. *De novo* assembly of the human genomes with massively parallel short read sequencing. *Genome Res.* doi:10.1101/gr.097261.109 (17 December 2009).
20. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34,** 5623–5630 (2006).
21. Colwell, R. K. EstimateS: Statistical estimation of species richness and shared species from samples, version 8.2. ⟨http://viceroy.eeb.uconn.edu/estimates⟩ (1997).
22. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55,** 205–211 (2006).
23. Wang, X., Heazlewood, S. P., Krause, D. O. & Florin, T. H. Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J. Appl. Microbiol.* **95,** 508–520 (2003).
24. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32,** D277–D280 (2004).
25. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4,** 41 (2003).
26. Jensen, L. J. *et al.* eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* **36,** D250–D254 (2008).
27. Kobayashi, K. *et al.* Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA* **100,** 4678–4683 (2003).
28. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** doi: 10.1038/msb4100050 (2006).
29. Dongowski, G., Lorenz, A. & Anger, H. Degradation of pectins with different degrees of esterification by *Bacteroides thetaiotaomicron* isolated from human gut flora. *Appl. Environ. Microbiol.* **66,** 1321–1327 (2000).
30. Cummings, J. H. & Macfarlane, G. T. The control and consequences of bacterial fermentation in the human colon. *J. Appl. Bacteriol.* **70,** 443–459 (1991).
31. Wong, J. M., de Souza, R., Kendall, C. W., Emam, A. & Jenkins, D. J. Colonic health: fermentation and short chain fatty acids. *J. Clin. Gastroenterol.* **40,** 235–243 (2006).
32. Hamer, H. M. *et al.* The role of butyrate on colonic function. *Aliment. Pharmacol. Ther.* **27,** 104–119 (2008).
33. Elango, R., Ball, R. O. & Pencharz, P. B. Amino acid requirements in humans: with a special emphasis on the metabolic availability of amino acids. *Amino Acids* **37,** 19–27 (2009).
34. Metges, C. C. Contribution of microbial amino acids to amino acid homeostasis of the host. *J. Nutr.* **130,** 1857S–1864S (2000).
35. Tap, J. *et al.* Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11,** 2574–2584 (2009).
36. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12,** 656–664 (2002).
37. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
38. Letunic, I., Yamada, T., Kanehisa, M. & Bork, P. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci.* **33,** 101–103 (2008).
39. von Mering, C. *et al.* STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **35,** D358–D362 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to the faculty and staff of Beijing Genomics Institute at Shenzhen, whose names were not included in the author list, but who contributed to large-scale sequencing of this team work. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013): MetaHIT, grant agreement HEALTH-F4-2007-201052, the Ole Rømer grant from the Danish Natural Science Research Council, the Solexa project (272-07-0196), the Shenzhen Municipal Government of China, the National Natural Science Foundation of China (30725008), the International Science and Technology Cooperation Project (0806), China (CXB200903110066A; ZYC200903240076A), the Danish Strategic Research Council grant no 2106-07-0021 (Seqnet), and the Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care. Ciberehd is funded by Instituto de Salud Carlos III (Spain). We also thank X. Wang from the School of Biosciences and Bioengineering, South China University of Technology, for his coordination on the Innovative Program for Undergraduate Students in which J.L. and Y.X. joined.

Author Contributions All authors are members of the Metagenomics of the Human Intestinal Tract (MetaHIT) Consortium. So.L., H.Y., Je.W., J.D., F.G., K.K., O.P., S.B., J.P., Ji.W., S.D.E. and Ju.W. managed the project. T.N., T.H. and K.S.B. performed clinical analyses; F.L. and C.M. performed DNA extraction. X.Z., B.W., J.C., H.L., Hu.Z., K.T., D.L.P., E.P. and M.J. performed sequencing. Ju.W., S.D.E, P.B., R.L., J.R., M.A. and J.Q. designed the analyses. J.Q., Sha.L., D.L., J.L., J.X., Y.X., Ho.Z., M.B., H.B.N., T.S.-P., C.Y., She.L., T.Y., N.P., J.-M.B., P.L., D.R.M., S.D.E. and Y.Z. performed the data analyses. S.D.E., P.B., J.R., J.Q., R.L. and Ju.W. wrote the paper. J.T., A.L., P.R., Y.L. and N.Q. revised the paper. The MetaHIT Consortium members contributed to design and execution of the study.

Author Information The raw Illumina read data of all 124 samples has been deposited in the EBI, under the accession ERA000116. The contigs and gene set are available to download from the EMBL (http://www.bork.embl.de/~arumugam/Qin_et_al_2010/) and BGI (http://gutmeta.genomics.org.cn) websites. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share-Alike license, and is freely available to all readers at www.nature.com/nature. Correspondence and requests for materials should be addressed to Ju.W. (wangj@genomics.org.cn) or S.D.E. (dusko.ehrlich@jouy.inra.fr).

MetaHIT Consortium (additional members)

Maria Antolin[1], François Artiguenave[2], Hervé Blottiere[3], Natalia Borruel[1], Thomas Bruls[2], Francesc Casellas[1], Christian Chervaux[4], Antonella Cultrone[3], Christine Delorme[3], Gérard Denariaz[4], Rozenn Dervyn[3], Miguel Forte[5], Carsten Friss[6], Maarten van de Guchte[3], Eric Guedon[3], Florence Haimet[3], Alexandre Jamet[3], Catherine Juste[3], Ghalia Kaci[3], Michiel Kleerebezem[7], Jan Knol[4], Michel Kristensen[8], Severine Layec[3], Karine Le Roux[3], Marion Leclerc[3], Emmanuelle Maguin[3], Raquel Melo Minardi[2], Raish Oozeer[4], Maria Rescigno[9], Nicolas Sanchez[3], Sebastian Tims[7], Toni Torrejon[1], Encarna Varela[1], Willem de Vos[7], Yohanan Winogradsky[3] & Erwin Zoetendal[7]

[1]Hospital Universitari Val d'Hebron, Ciberehd, 08035 Barcelona, Spain. [2]Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France. [3]Institut National de la Recherche Agronomique, 78350 Jouy en Josas, France. [4]Danone Research, 91120 Palaiseau, France. [5]UCB Pharma SA, 28046 Madrid, Spain. [6]Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. [7]Wageningen Unviersiteit, 6710BA Ede, The Netherlands. [8]Hagedorn Research Institute, DK 2820 Copenhagen, Denmark. [9]Istituto Europeo di Oncologia, 20100 Mila, Italy.

METHODS

Human faecal sample collection. Danish individuals were from the Inter-99 cohort[40], varying in phenotypes according to BMI and status towards obesity/diabetes, whereas Spanish individuals were either healthy controls or patients with chronic inflammatory bowel diseases (Crohn's disease or ulcerative colitis) in clinical remission.

Patients and healthy controls were asked to provide a frozen stool sample. Fresh stool samples were obtained at home, and samples were immediately frozen by storing them in their home freezer. Frozen samples were delivered to the Hospital using insulating polystyrene foam containers, and then they were stored at −80 °C until analysis.

DNA extraction. A frozen aliquot (200 mg) of each faecal sample was suspended in 250 µl of guanidine thiocyanate, 0.1 M Tris (pH 7.5) and 40 µl of 10% N-lauroyl sarcosine. Then, DNA extraction was conducted as previously described[22]. The DNA concentration and its molecular size were estimated by nanodrop (Thermo Scientific) and agarose gel electrophoresis.

DNA library construction and sequencing. DNA library preparation followed the manufacturer's instruction (Illumina). We used the same workflow as described elsewhere to perform cluster generation, template hybridization, isothermal amplification, linearization, blocking and denaturization and hybridization of the sequencing primers. The base-calling pipeline (version IlluminaPipeline-0.3) was used to process the raw fluorescent images and call sequences.

We constructed one library (clone insert size 200 bp) for each of the first 15 samples, and two libraries with different clone insert sizes (135 bp and 400 bp) for each of the remaining 109 samples for validation of experimental reproducibility.

To estimate the optimal return between the generation of novel sequence and sequencing depth, we aligned the Illumina GA reads from samples MH0006 and MH0012 onto 468,335 Sanger reads totalling to 311.7 Mb generated from the same two samples (156.9 and 154.7 Mb, respectively, Supplementary Table 2), using the Short Oligonucleotide Alignment Program (SOAP)[41] and a match requirement of 95% sequence identity. With about 4 Gb of Illumina sequence, 94% and 89% of the Sanger reads (for MH0006 and MH0012, respectively) were covered. Further extensive sequencing, to 12.6 and 16.6 Gb for MH0006 and MH0012, respectively, brought only a moderate increase of coverage to about 95% (Supplementary Fig. 1). More than 90% of the Sanger reads were covered by the Illumina sequences to a very high and uniform level (Supplementary Fig. 2), indicating that there is little or no bias in the Illumina GA sequence. As expected, a large proportion of Illumina sequences (57% and 74% for M0006 and M0012, respectively) was novel and could not be mapped onto the Sanger reads. This fraction was similar at the 4 and 12–16 Gb sequencing levels, confirming that most of the novelty was captured already at 4 Gb.

We generated 35.4–97.6 million reads for the remaining 122 samples, with an average of 62.5 million reads. Sequencing read length of the first batch of 15 samples was 44 bp and the second batch was 75 bp.

Public data used. The sequenced bacteria genomes (totally 806 genomes) deposited in GenBank were downloaded from NCBI database (http://www.ncbi.nlm.nih.gov/) on 10 January 2009. The known human gut bacteria genome sequences were downloaded from HMP database (http://www.hmpdacc-resources.org/cgi-bin/hmp_catalog/main.cgi), GenBank (67 genomes), Washington University in St Louis (85 genomes, version April 2009, http://genome.wustl.edu/pub/organism/Microbes/Human_Gut_Microbiome/), and sequenced by the MetaHIT project (17 genomes, version September 2009, http://www.sanger.ac.uk/pathogens/metahit/). The other gut metagenome data used in this project include: (1) human gut metagenomic data sequenced from US individuals[8], which was downloaded from NCBI with the accession SRA002775; (2) human gut metagenomic data from Japanese individuals[17], which was downloaded from P. Bork's group at EMBL (http://www.bork.embl.de). The integrated NR database we constructed in this study included NCBI-NR database (version April 2009) and all genes from the known human gut bacteria genomes.

Illumina GA short reads *de novo* assembly. High-quality short reads of each DNA sample were assembled by the SOAPdenovo assembler[19]. In brief, we first filtered the low abundant sequences from the assembly according to 17-mer frequencies. The 17-mers with depth less than 5 were screened in front of assembly, for these low-frequency sequences were very unlikely to be assembled, whereas removing them would significantly reduce memory requirement and make assembly feasible in an ordinary supercomputer (512 GB memory in our institute).

Then the sequences were processed one by one and the de Bruijn graph data format was used to store the overlap information among the sequences. The overlap paths supported by a single read were unreliable and removed. Short low-depth tips and bubbles that were caused by sequencing errors or genetic variations between microbial strains were trimmed and merged, respectively. Read paths were used to solve the tiny repeats.

Finally, we broke the connections at repeat boundaries, and outputted the continuous sequences with unambiguous connections as contigs. The metagenomic special model was chosen, and parameters '−K 21' and '−K 23' were used for 44 bp and 75 bp reads, respectively, to indicate the minimal sequence overlap required.

After *de novo* assembly for each sample independently, we merged all the unassembled reads together and performed assembly for them, as to maximize the usage of data and assemble the microbial genomes that have low frequency in each read set, but have sufficient sequence depth for assembly by putting the data of all samples together.

Validating Illumina contigs using Sanger reads. We used BLASTN (WU-BLAST 2.0) to map Sanger reads from samples MH0006 and MH0012 (156.9 Mb and 154.7 Mb, respectively) to Illumina contigs (single best hit longer than 75 bp and over 95% identity) from the same samples. Each alignment was scanned for breakage of collinearity where both sequences have at least 50 bases left unaligned at one end of the alignment. Each such breakage was considered an assembly error in the Illumina contig at the location where collinearity breaks. Errors within 30 bp from each other were merged. An error was discarded if there exists a Sanger read that agrees with the contig structure for 60 bp on both sides of the error. For comparison, we repeated this on a Newbler2 assembly of 454 Titanium reads from MH0006 (550 Mb reads). Supplementary Fig. 5a shows the number of errors per Mb of assembled Illumina/454 contigs. We estimate 14.12 errors per Mb of contigs for the Illumina assembly, which is comparable to that of the 454 assembly (20.73 per Mb). 98.7% of Illumina contigs that map at least one Sanger read were collinear over 99.55% of the mapped regions, which is comparable to 97.86% of such 454 contigs being collinear over 99.48% of the mapped regions.

Evaluation of human gut microbiome coverage. The Illumina GA reads were aligned against the assembled contigs and known bacteria genomes using SOAP[41] by allowing at most two mismatches in the first 35-bp region and 90% identity over the read sequence. The Roche/454 and Sanger sequencing reads were aligned against the same reference using BLASTN with 1×10^{-8}, over 100 bp alignment length and minimal 90% identity cutoff. Two mismatches were allowed and identity was set 95% over the read sequence when aligned to the GA reads of MH0006 and MH0012 to Sanger reads from the same samples by SOAP.

Gene prediction and construction of the non-redundant gene set. We use MetaGene[20]—which uses di-codon frequencies estimated by the GC content of a given sequence, and predicts a whole range of ORFs based on the anonymous genomic sequences—to find ORFs from the contigs of each of the 124 samples as well as the contigs from the merged assembly.

The predicted ORFs were then aligned to each other using BLAT[36]. A pair of genes with greater than 95% identity and aligned length covered over 90% of the shorter gene was grouped together. The groups sharing genes were then merged, and the longest ORF in each merged group was used to represent the group, and the other members of the group were taken as redundancy. Therefore, we organized the non-redundant gene set from all the predicted genes by excluding the redundancy. Finally, the ORFs with length less than 100 bp were filtered. We translated the ORFs into protein sequences using the NCBI Genetic Codes11.

Identification of genes. To make a balance between identifying low-abundance genes and reducing the error-rate of identification, we explored the impact of the threshold set for read coverage required to identify a gene in individual microbiomes. The number of genes decreased about twice when the number of reads required for identification was increased from 2 to 6, and changed slowly thereafter (Supplementary Fig. 6a). Nevertheless, to include the rare genes into the analysis, we selected the threshold of 2 reads.

Gene taxonomic assignment. Taxonomic assignment of predicted genes was carried out using BLASTP alignment against the integrated NR database. BLASTP alignment hits with *e*-values larger than 1×10^{-5} were filtered, and for each gene the significant matches which were defined by *e*-values $\leq 10 \times$ *e*-value of the top hit were retained to distinguish taxonomic groups. Then we determined the taxonomical level of each gene by the lowest common ancestor (LCA)-based algorithm that was implemented in MEGAN[42]. The LCA-based algorithm assigns genes to taxa in the way that the taxonomical level of the assigned taxon reflects the level of conservation of the gene. For example, if a gene was conserved in many species, it was assigned to the LCA rather than to a species.

Gene functional classification. We used BLASTP to search the protein sequences of the predicted genes in the eggNOG database[26] and KEGG database[24] with *e*-value $\leq 1 \times 10^{-5}$. The genes were annotated as the function of the NOGs or KEGG homologues with lowest *e*-value. The eggNOG database is an integration of the COG and KOG databases. The genes annotated by COG were classified into the 25 COG categories, and genes that were annotated by KEGG were assigned into KEGG pathways.

Determination of minimal gut bacterial genome. The number of non-redundant genes assigned to the eggNOG clusters was normalized by gene length and cluster copy number (Supplementary Fig. 8). The clusters were ranked by normalized gene number and the range that included the clusters encoding essential *Bacillus subtilis* genes was determined, computing the proportion of these clusters among the successive groups of 100 clusters. Analysis of the range gene clusters involved, besides iPath projections, use of KEGG and manual verification of the completeness of the pathways and protein machineries they encode.

Determination of total functional complement and minimal metagenome. We computed the total and shared number of orthologous groups and/or gene families present in random combinations of n individuals (with $n = 2$ to 124, 100 replicates per bin). This analysis was performed on three groups of gene clusters: (1) known eggNOG orthologous groups (that is, those with functional annotation, excluding those in which the terms [Uu]ncharacteri[sz]ed, [Uu]nknown, [Pp]redicted or[Pp]utative occurred); (2) all eggNOG orthologous groups; (3) all orthologous groups plus gene families constructed from remaining genes not assigned to the two above categories. Families were clustered from all-against-all BLASTP results using MCL[43] with an inflation factor of 1.1 and a bit-score cutoff of 60.

Rarefaction analysis. Estimation of total gene richness was done using EstimateS on 100 randomly picked samples due to memory limitations. Because the CV value was >0.5, both chao2 (classic) and ICE richness estimators were calculated and the larger estimate of the two (ICE) was used. The estimate for this sample size was 3,621,646 genes (ICE) whereas S_{obs} (Mao Tau) was 3,090,575 genes, or 85.3%. The ICE estimator curve did not completely saturate, (data not shown) indicating that additional samples will need to be added to achieve a final, conclusive estimate.

Common bacterial core. To eliminate the influence of very similar strains and assess the presence of known microbial species among the individuals of the cohort, we used 650 sequenced bacterial and archaeal genomes as a reference set. The set was composed from 932 publicly available genomes, which were grouped by similarity, using a 90% identity cutoff and the similarity over at least 80% of the length. From each group only the largest genome was used. Illumina reads from 124 individuals were mapped to the set, for species profiling analysis and the genomes originating from the same species (by differing in size >20%) curated by manual inspection and by using the 16S-based clustering when the sequences were available.

Relative abundance of microbial genomes among individuals. We computed the genome coverage by uniquely mapping Illumina reads and normalized it to 1 Gb of sequence, to correct for different sequencing levels in different individuals. The coverage was summed over all species of the non-redundant bacterial genome set for each individual and the proportion of each species relative to the sum calculated.

Species co-existence network. For the 155 species that had genome coverage by the Illumina reads ≥1% in at least one individual we calculated the pair-wise inter-species Pearson correlations between sequencing depths (abundance) throughout the entire cohort of 124 individuals. From the resulting 11,175 inter-species correlations, correlations less than −0.4 or above 0.4 ($n = 342$) were visualized in a graph using Cytoscape[44] displaying the average genome coverage of each species as node size in the graph.

40. Toft, U. *et al.* The impact of a population-based multi-factorial lifestyle intervention on changes in long-term dietary habits: The Inter99 study. *Prev. Med.* **47**, 378–383 (2008).
41. Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967 (2009).
42. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
43. van Dongen, S. *Graph Clustering by Flow Simulation.* PhD thesis, Univ. Utrecht (2000).
44. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Week 10

Host-Microbe Symbiosis

Biotechnology

LETTER

doi:10.1038/nature12447

Bacterial colonization factors control specificity and stability of the gut microbiota

S. Melanie Lee[1], Gregory P. Donaldson[1]*, Zbigniew Mikulski[2]*, Silva Boyajian[1], Klaus Ley[2] & Sarkis K. Mazmanian[1]

Mammals harbour a complex gut microbiome, comprising bacteria that confer immunological, metabolic and neurological benefits[1]. Despite advances in sequence-based microbial profiling and myriad studies defining microbiome composition during health and disease, little is known about the molecular processes used by symbiotic bacteria to stably colonize the gastrointestinal tract. We sought to define how mammals assemble and maintain the *Bacteroides*, one of the most numerically prominent genera of the human microbiome. Here we find that, whereas the gut normally contains hundreds of bacterial species[2,3], germ-free mice mono-associated with a single *Bacteroides* species are resistant to colonization by the same, but not different, species. To identify bacterial mechanisms for species-specific saturable colonization, we devised an *in vivo* genetic screen and discovered a unique class of polysaccharide utilization loci that is conserved among intestinal *Bacteroides*. We named this genetic locus the commensal colonization factors (*ccf*). Deletion of the *ccf* genes in the model symbiont, *Bacteroides fragilis*, results in colonization defects in mice and reduced horizontal transmission. The *ccf* genes of *B. fragilis* are upregulated during gut colonization, preferentially at the colonic surface. When we visualize microbial biogeography within the colon, *B. fragilis* penetrates the colonic mucus and resides deep within crypt channels, whereas *ccf* mutants are defective in crypt association. Notably, the CCF system is required for *B. fragilis* colonization following microbiome disruption with *Citrobacter rodentium* infection or antibiotic treatment, suggesting that the niche within colonic crypts represents a reservoir for bacteria to maintain long-term colonization. These findings reveal that intestinal *Bacteroides* have evolved species-specific physical interactions with the host that mediate stable and resilient gut colonization, and the CCF system represents a novel molecular mechanism for symbiosis.

International microbiome sequencing initiatives are revealing detailed inventories of diverse bacterial communities across various body sites, diets and human populations[2–4]. Complex ecosystems have been forged by co-adaptation over millennia between animals and microbes to create stable and specific microbiomes[5,6], suggesting the evolution of molecular mechanisms that establish and maintain symbiotic microbial colonization. Bacteroidetes is one of the most numerically abundant Gram-negative phyla in the mammalian gastrointestinal tract[7]. Studies in the genus *Bacteroides* have revealed species that induce glycosylation of the intestinal epithelium[8], produce glycoside hydrolases that digest carbohydrates for host nutrient use[9], direct host immune maturation[10] and protect animals from inflammation in experimental models of inflammatory bowl disease and multiple sclerosis[11–13]. To explore the dynamics of microbiome assembly, we sequentially introduced *Bacteroides* species to germ-free mice and monitored colonization via colony-forming units (c.f.u.) in faeces. Animals were readily colonized with *B. fragilis* followed by *Bacteroides thetaiotaomicron* (Fig. 1a) or *Bacteroides vulgatus* (Fig. 1b), and altering the sequence of microbial exposure does not affect results (Supplementary Fig. 1a). Notably, however, animals colonized with *B. fragilis* and then exposed to the same species (marked by an antibiotic resistance gene) are resistant to super-colonization and clear the challenging strain (Fig. 1c). This novel observation of 'colonization resistance' by the same species is conserved in three other *Bacteroides* (Supplementary Fig. 1b–d), regardless of the antibiotic resistance markers used (Supplementary Fig. 1e), but not in *Escherichia coli* (Supplementary Fig. 1f). As conventional mice typically harbour 10^{11}–10^{12} c.f.u. per gram of caecal content[14] (100-fold greater than *Bacteroides* in mono-association), there does not seem to be a shortage of space or nutrients under these conditions, using a nutrient-rich standard diet. We thus proposed that individual *Bacteroides* species colonize the gut by saturating a limited and unique niche. Indeed, treatment of *B. fragilis* mono-associated mice with erythromycin to displace the existing strain permits colonization by an erythromycin-resistant challenge strain (Fig. 1d). These data suggest that *Bacteroides* colonize the gut in a species-specific and saturable manner.

We developed a functional *in vivo* screen to identify genetic factor(s) from *B. fragilis* that are sufficient to mediate species-specific colonization. Mice were mono-associated with *B. vulgatus*, then challenged with a library of *B. vulgatus* clones that each contained a fragment of *B. fragilis* genomic DNA (schematic in Supplementary Fig. 2a). We reasoned that only those clones containing genes that conferred stable gut colonization by *B. fragilis* would persist, with the remainder being cleared. We screened 2,100 clones each containing 9–10 kilobases of DNA, providing a 3.8-fold coverage of the *B. fragilis* genome and 98% probability that a given DNA sequence is present in the library (Supplementary Equation (1)). Notably, 30 days after orally gavaging the library into animals, only two clones sustained colonization. The inserts from both clones mapped to the same locus on the *B. fragilis* genome, BF3579–BF3583 (Supplementary Fig. 2b).

On the basis of predicted protein sequences, BF3583 and BF3582 constitute a sigma (σ) factor/anti-σ factor gene pair. BF3581 is a member of the SusC family of outer membrane proteins. BF3580 is a homologue of SusD, a lipoprotein often paired with SusC. These Sus-like systems have been shown to bind and import a range of oligosaccharide molecules[15–18]. BF3579 encodes a putative chitobiase, suggesting a possible polysaccharide substrate for this system[19] (Fig. 1e). Comparative genomic analysis using the Integrated Microbial Genomes database (http://img.jgi.doe.gov/cgi-bin/w/main.cgi) reveals conservation of similar clusters of genes among sequenced intestinal *Bacteroides* species (Supplementary Fig. 3). Sus-like systems are numerous in *Bacteroides* within polysaccharide utilization loci (PULs), which are gene cassettes used to harvest dietary sugars and/or forage host glycans during nutrient deprivation[15,20,21]. As PULs have not previously been implicated in saturable niche colonization, the locus we have identified encodes a unique pathway in *Bacteroides* for species-specific gut association; we named the genes *ccfA–E*, for commensal colonization factors (Fig. 1e). Furthermore, as deletion of the most closely related genes from *B. fragilis* (BFΔ0227–0229; Supplementary Fig. 3) do not affect colonization dynamics (Supplementary Fig. 4) we suggest that the CCF system represents a functionally unique subset of PULs that evolved to promote long-term symbiosis.

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA. [2]Division of Inflammation Biology, La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

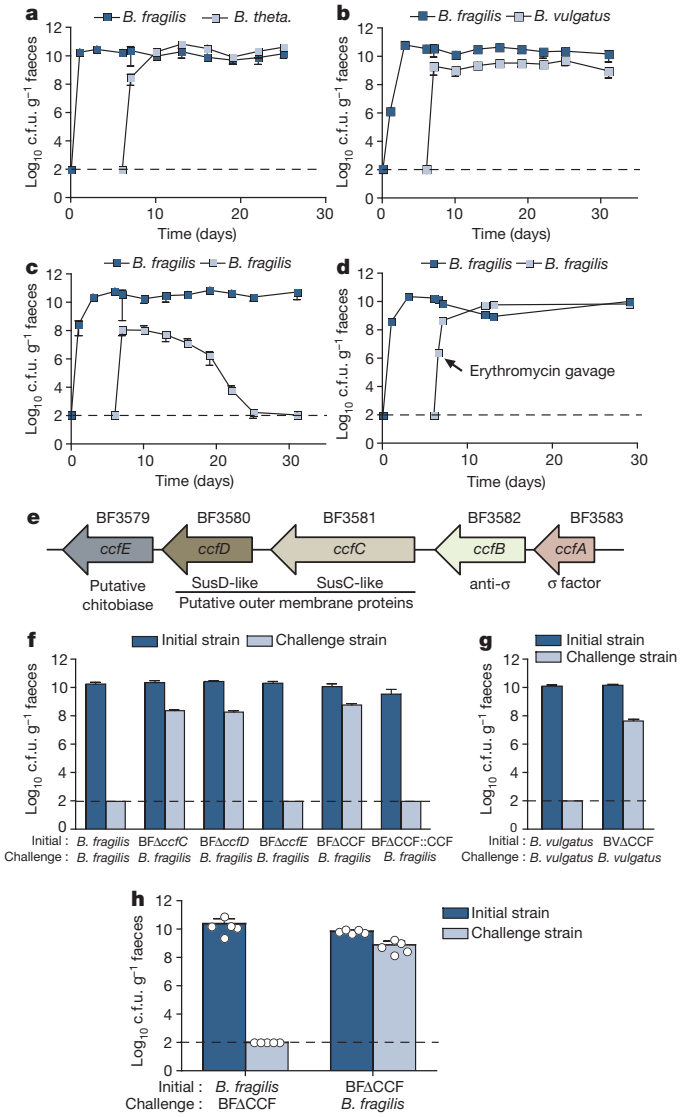


Figure 1 | *Bacteroides* species occupy species-specific niches in the gut via an evolutionarily conserved genetic locus. **a**–**c**, Germ-free mice were mono-associated with *B. fragilis* and challenged orally with *B. thetaiotaomicron* (**a**), *B. vulgatus* (**b**) or *B. fragilis* (**c**). **d**, Mice were mono-associated with erythromycin-sensitive *B. fragilis*, and subsequently challenged with erythromycin-resistant *B. fragilis*. Erythromycin was administered where indicated. **e**, Genomic organization of the *ccf* locus. **f**, Mice were mono-associated with either wild-type *B. fragilis*, mutant strains deleted in *ccfC*, *ccfD*, *ccfE*, and *ccfC–E* (BFΔCCF), or complemented strain (BFΔCCF::CCF) and challenged with wild-type *B. fragilis*. c.f.u. were determined after 30 days. **g**, Mice were mono-associated with wild-type *B. vulgatus* or a mutant strain deleted in *ccfC–E* genes (BVΔCCF), and challenged with wild-type *B. vulgatus*. c.f.u. were determined after 30 days. In all sequential colonization studies, results are representative of at least two independent trials ($n = 3$–4 animals per group). **h**, Cross-colonization between wild-type *B. fragilis* and BFΔCCF mono-associated mice at 7 days after encounter measured by c.f.u. of the initially colonizing and the horizontally transmitted (challenge) strains ($n = 2$ animals per encounter, 5 independent trials). All graphs: dashed line indicates the limit of detection at 100 c.f.u. g^{-1} faeces, and error bars indicate s.d.

To test whether the putative structural genes (*ccfC–E*) are required for gut colonization, we generated in-frame deletion mutants of *B. fragilis*: Δ*ccfC*, Δ*ccfD* and Δ*ccfE*. All strains exhibit normal morphology on solid agar medium and unimpaired growth in laboratory culture (data not shown). As shown previously, animals mono-colonized with wild-type *B. fragilis* completely clear the wild-type challenge strain after 30 days (Fig. 1f; first bars). However, animals mono-associated with Δ*ccfC* or Δ*ccfD* are permissive to colonization by wild-type bacteria (Fig. 1f; second and third bars), unlike the *ccfE* mutant (Fig. 1f; fourth bars). A deletion mutant in all three genes (BFΔCCF) also allows wild-type *B. fragilis* to colonize (Fig. 1f; fifth bars). Trans-complementation of the BFΔCCF strain with *ccfA–E* restores colonization resistance (Fig. 1f; sixth bars). Similarly, a mutant in the *B. vulgatus ccfC–E* orthologues (ΔBVU0946–BVU0948) also permits wild-type *B. vulgatus* to colonize (Fig. 1g), demonstrating conservation in *Bacteroides* species. When we tested horizontal transmission of bacteria between wild-type *B. fragilis*

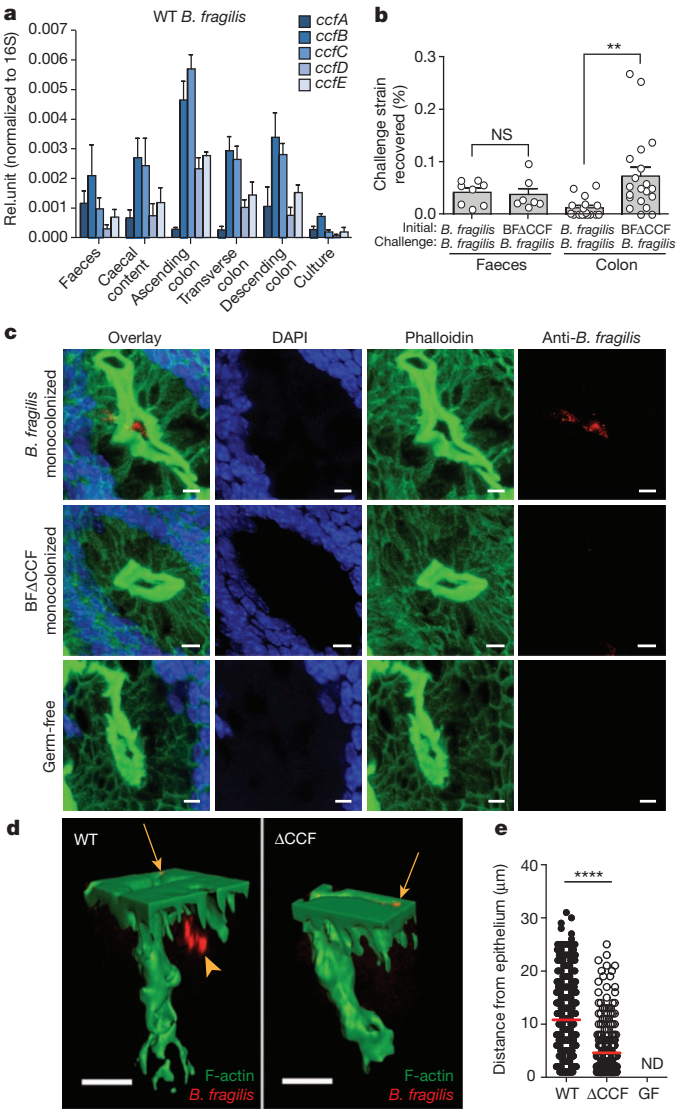


Figure 2 | *B. fragilis* colonization of the colonic crypts is mediated by the CCF system. **a**, Quantitative PCR with reverse transcription (qRT–PCR) of *ccf* gene expression levels normalized to 16S rRNA ($n = 3$ animals, 2 trials). **b**, Mice were mono-associated with either wild-type *B. fragilis* or BFΔCCF, and challenged with wild-type *B. fragilis*. The percentage of challenge strain was determined in the lumen (faeces) and colon after 1 day ($n = 8$ animals per group). **c**, Confocal micrographs of germ-free, wild-type *B. fragilis* or BFΔCCF mono-associated mice colon whole-mount. Crypts are visualized by DAPI (4′,6-diamidino-2-phenylindole; nuclei, blue) and phalloidin (F-actin, green). Bacteria (red) are stained with IgY polyclonal antibody raised against *B. fragilis*. Images are representative of seven different sites analysed from at least two different colons. Scale bars, 5 μm. **d**, Three-dimensional reconstructions of colon crypts from wild-type *B. fragilis* or BFΔCCF mono-associated mice. Bacteria are detected on the apical surface of the epithelium (arrows) and in the crypt space (arrowhead). Scale bars, 10 μm. **e**, Quantification of bacterial penetration, measured as distance from the epithelial surface per crypt. Error bars indicate s.e.m. GF, germ-free; ND, not detected; NS, not significant; rel.unit, relative unit; WT, wild type. $^{**}P < 0.01$; $^{****}P < 0.0001$.

and BFΔCCF mono-associated mice, only wild-type bacteria cross-colonized (Fig. 1h). Thus, the CCF system is involved in colonization resistance by *Bacteroides*.

Building on the previous discovery that a population of *B. fragilis* associates with mucosal tissues[22], we show that *ccfB–E* are preferentially expressed by bacteria in contact with the colon, with lower levels in caecal content and faeces (Fig. 2a). There is virtually no expression in laboratory culture. Thus, *in vivo* expression of *ccf* in gut tissue may be critical for colonization. Indeed, in contrast to laboratory-grown bacteria (see Fig. 1c), sustained colonization is conferred to bacteria recovered directly from animals (Supplementary Fig. 5). To examine regulation of gene expression, we deleted the σ factor *ccfA*, which led to highly reduced expression of all five genes during animal colonization (Supplementary Fig. 6a). Accordingly, germ-free mice mono-colonized with the *B. fragilis* Δ*ccfA* mutant are permissive of super-colonization by wild-type *B. fragilis*, demonstrating a functional defect in the saturable niche occupancy (Supplementary Fig. 6b). On the basis of this tissue-associated expression pattern, we tested whether the CCF system promotes bacterial localization to mucosal tissue. Mice were mono-associated with either the wild-type or the *ccf* deletion strain, and both groups were subsequently challenged with wild-type *B. fragilis*. Twenty-four hours after challenge, we observed the same numbers for challenge strains in faeces of both groups (Fig. 2b first and second bars, and Supplementary Fig. 7a, b). By contrast, *ccf*-mutant-associated animals show higher levels of challenge strain in colon tissue, suggesting a colonization defect by the mutant strain specifically at the mucosal surface (Fig. 2b third and fourth bars, and Supplementary Fig. 7a, b). Therefore, CCF-mediated colonization fitness seems to involve physical association with the gut.

Recent studies have revealed microbial communities that colonize intestinal crypts of conventional mice in the absence of disease[23], and we have shown that *B. fragilis* occupies the colonic crypts of mono-associated mice[22]. Discovering a role for *ccf* genes near mucosal tissue led us to explore the intriguing hypothesis that the CCF system mediates crypt occupancy. We mono-colonized mice with *B. fragilis* and visualized bacterial localization in colon tissue by whole-mount confocal microscopy. Indeed, wild-type *B. fragilis* co-localize with crypts from the ascending colon, appearing to be located in the centre of crypt opening. Notably, BFΔCCF mono-associated mice display virtually no crypt occupancy (Fig. 2c and Supplementary Fig. 8). Colon cross-section imaging also reveals that only wild-type bacteria are crypt associated (Supplementary Fig. 9). Two-photon imaging of colon explants clearly demonstrates the presence of wild-type *B. fragilis* on the surface of the epithelium and inside the crypt. Although both wild-type and mutant strains of *B. fragilis* associate with the surface of the epithelium, only wild-type bacteria are able to penetrate deep into the colonic crypts of mice (Fig. 2d and Supplementary Video 1). Measuring the distance from the surface of the epithelium to bacterial signals in a survey of crypts reveals significantly greater tissue penetration by wild-type bacteria (Fig. 2e). Collectively, these data reveal that the CCF system allows *B. fragilis* to reside in a specific niche within crypts during steady-state colonization.

We next investigated the effects of the CCF system in the context of a complex microbiota. Wild-type *Bacteroides* species do not readily colonize most strains of specific pathogen-free (SPF) mice, namely BALB/c, Swiss Webster and C57BL/6, despite oral administration of high inoculums (Supplementary Fig. 10a, c and data not shown). Furthermore, transfer of an SPF microbiota to mono-colonized mice leads to clearance of wild-type *B. fragilis* (Supplementary Fig. 10b, d). To overcome this obstacle, we tested various additional genetic backgrounds and empirically determined that C57BL/6 $Rag^{-/-}$ mice (which lack an adaptive immune system) and non-obese diabetic (NOD) mice can be stably colonized by *B. fragilis* with a single oral gavage. We introduced either wild-type or *ccf* mutant *B. fragilis* at equal inoculums into separate groups of animals, and measured colonization. Only wild-type *B. fragilis* stably colonizes SPF $Rag^{-/-}$ mice, whereas BFΔCCF establishes a significantly lower colonization in the gut (Fig. 3a). Co-inoculation of equal numbers of wild-type and *ccf* mutant bacteria into $Rag^{-/-}$ mice also results in rapid clearance of the mutant strain from the gut (Fig. 3b), demonstrating a cell-intrinsic defect that could not be complemented in *trans* by wild-type bacteria. NOD animals are also preferentially colonized by wild-type *B. fragilis* compared to *ccf* mutants in separate groups of animals (Fig. 3c) or in equal co-inoculation (Fig. 3d). These data show that deletion of the *ccf* genes compromises *B. fragilis* colonization of hosts with a complex microbiota.

During symbiosis with mammals, the microbiota may be confronted by rapid environmental changes with potentially adverse consequences to bacteria, such as enteric infections or antibiotic exposure. Gastroenteritis is commonly experienced by humans and is known to perturb the microbiota. To test whether resilience of *B. fragilis* colonization is CCF-dependent, we used *C. rodentium* infection of mice to mimic human gastrointestinal tract infection[24]. Using an antibiotic treatment protocol that does not sterilize the gut but promotes colonization of SPF mice by *Bacteroides*[25], we were able to simultaneously colonize mice with equivalent levels of wild-type and *ccf* mutant bacteria. Mice were subsequently infected orally with *C. rodentium*, and colonization of *B. fragilis* was monitored. Wild-type bacteria decline in number at first, but return to maximal levels 3–4 weeks after infection (Fig. 3e). Importantly, the BFΔCCF strain is completely cleared from the mouse

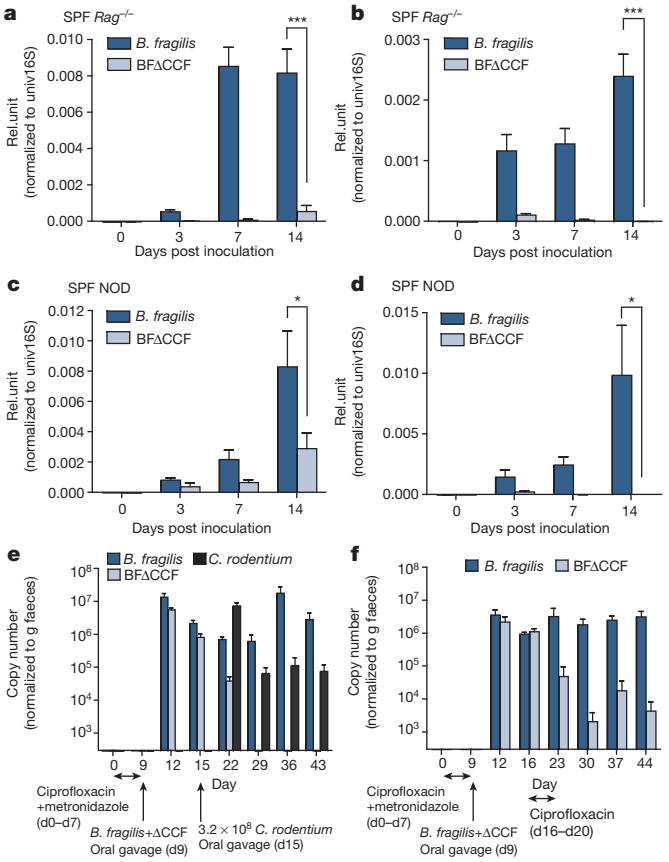


Figure 3 | *B. fragilis* requires the *ccf* genes for stable and resilient colonization of mice. a, Groups of SPF $Rag^{-/-}$ mice were gavaged with either wild-type *B. fragilis* or BFΔCCF. **b**, SPF $Rag^{-/-}$ mice were given a 1:1 co-inoculum of wild-type *B. fragilis* and BFΔCCF by single gavage. **c**, SPF NOD mice were gavaged with either wild-type *B. fragilis* or BFΔCCF. **d**, SPF NOD mice were given a 1:1 co-inoculum of wild-type *B. fragilis* and BFΔCCF by single gavage. **e**, SPF mice were co-associated with wild-type *B. fragilis* and BFΔCCF, and infected with *C. rodentium*. **f**, SPF mice were co-associated with wild-type *B. fragilis* and BFΔCCF, and given ciprofloxacin in drinking water for the time period shown. For all analyses, bacterial colonization levels were assessed by qRT–PCR from stool DNA ($n = 4$ animals per group). Results are representative of at least two independent trials per experiment. Error bars indicate s.e.m. $^*P < 0.05$; $^{***}P < 0.001$.

gut after gastroenteritis (Fig. 3e), but not when animals are left uninfected (Supplementary Fig. 11a). Next, we challenged mice that were co-colonized with wild-type and BFΔCCF with oral antibiotics and observe a colonization defect only in *ccf* mutant bacteria (Fig. 3f and Supplementary Fig. 11b). These results reveal that the CCF system establishes resilient colonization by gut *Bacteroides* following disruption of the microbiome. Finally, when SPF mice colonized with wild-type *B. fragilis* were given an antibiotic treatment that cleared faecal bacteria, crypt-associated microbial populations persisted (Supplementary Fig. 12), suggesting that symbiotic bacteria occupy a protected niche that creates a reservoir for stable gut colonization.

Co-evolution has bound microbes and man in an inextricable partnership, resulting in remarkable specificity and stability of the human microbiome[2,3]. Our findings reveal a novel pathway required for persistent colonization of the mammalian gut by the *Bacteroides*. Homology to the Sus family of proteins suggests a role for CCF in uptake and use of glycans. Although certain Sus-containing PULs in *B. thetaiotaomicron* mediate foraging of host mucus[15], their contributions to microbial colonization have been previously described only during nutrient deprivation conditions[20]. Our discovery of CCF-dependent colonization in mice fed a nutrient-rich diet suggests a new role whereby *Bacteroides* evolved specific Sus-like systems to use non-dietary glycans during homeostasis. On the basis of findings that *ccf* genes are preferentially expressed in proximity to mucosal tissues and *B. fragilis* associates with colonic crypts, we find it likely that host factors may promote expression of the CCF system. In support of this notion, *N*-acetyl-D-lactosamine (LacNAc)—a component of host mucus—induces the *ccf* genes and its homologues in *B. fragilis* and *B. thetaiotaomicron*[20] (Supplementary Fig. 13). But because a closely related PUL (BF0227–31) responds to LacNAc but does not mediate saturable niche colonization (Supplementary Fig. 4), LacNAc alone may be an inducer but is not the substrate used by CCF systems. We propose that specific glycan structures within colonic crypts serve as nutrient sources for individual *Bacteroides* species, and that CCF systems provide a molecular mechanism for a hypothesis proposed decades ago, "that populations of most indigenous intestinal bacteria are controlled by substrate competition, i.e., that each species is more efficient than the rest in utilizing one or a few particular substrates and that the population level of that species is controlled by the concentration of these few limiting substrates"[26]. Future work will aim to identify the precise glycan(s) for CCF systems from various *Bacteroides*. Finally, our data suggest that the *ccf* genes encode a specific subset of PULs that evolved the novel activity of promoting stable and resilient colonization, and crypt-associated bacterial reservoirs may represent 'founder' cells that repopulate the gut following disruption of the microbiome by enteric infections or antibiotic exposure. Discovery of a molecular mechanism for colonization fitness by gut bacteria provides a glimpse into the evolutionary forces that have shaped the assembly and dynamics of the human microbiome.

METHODS SUMMARY

All germ-free mice were bred and housed in flexible film isolators until 8 weeks of age, then transferred to microisolator cages and maintained with autoclaved food, bedding and water supplemented with gentamicin and erythromycin. Mice were mono-associated with gentamicin- and erythromycin-resistant *Bacteroides* strains by single oral gavage. Colonization level was determined over time by stool serial dilution plating on selective agar media. For qRT–PCR, total RNA was extracted from laboratory bacterial culture, faecal and caecal content (ZR Soil/Fecal RNA MicroPrep, Zymo Research) and colon tissues (Trizol, Invitrogen) from mono-associated animals, converted to first strand complementary DNA and analysed by qPCR using Power SYBR Green PCR Master Mix (Applied Biosystems). For colon whole-mount imaging, tissues were collected from germ-free or single strain mono-associated animals, fixed with 4% paraformaldehyde and stained with a *B. fragilis* specific antibody, DAPI and phalloidin. The colon crypts were visualized by confocal microscopy and two-photon microscopy. SPF mice were colonized with *B. fragilis* and/or BFΔCCF by single oral gavage and the colonization level was determined over time by stool DNA extraction (ZR Fecal DNA MiniPrep, Zymo Research) and qPCR using strain specific primers.

Full Methods and any associated references are available in the online version of the paper.

Received 6 September 2012; accepted 11 July 2013.
Published online 18 August 2013.

1. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl Acad. Sci. USA* **110,** 3229–3236 (2013).
2. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
3. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).
4. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
5. Palmer, C., Bik, E. M., Digiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5,** e177 (2007).
6. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6,** e280 (2008).
7. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308,** 1635–1638 (2005).
8. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host-microbial interactions in an open mammalian ecosystem. *Science* **273,** 1380–1383 (1996).
9. Xu, J. *et al.* A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299,** 2074–2076 (2003).
10. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–118 (2005).
11. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–625 (2008).
12. Round, J. L. & Mazmanian, S. K. Inducible Foxp3$^+$ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **107,** 12204–12209 (2010).
13. Ochoa-Repáraz, J. *et al.* Central nervous system demyelinating disease protection by the human commensal *Bacteroides fragilis* depends on polysaccharide A expression. *J. Immunol.* **185,** 4101–4108 (2010).
14. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124,** 837–848 (2006).
15. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nature Rev. Microbiol.* **10,** 323–335 (2012).
16. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284,** 24673–24677 (2009).
17. Shipman, J. A., Berleman, J. E. & Salyers, A. A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **182,** 5365–5372 (2000).
18. Schauer, K., Rodionov, D. A. & de Reuse, H. New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends. Biochem. Sci.* **33,** 330–338 (2008).
19. Kawada, M. *et al.* Chitinase 3-like-1 enhances bacterial adhesion to colonic epithelial cells through the interaction with bacterial chitin-binding protein. *Lab. Invest.* **88,** 883–895 (2008).
20. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4,** 447–457 (2008).
21. Sonnenburg, J. L. *et al.* Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–1959 (2005).
22. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332,** 974–977 (2011).
23. Pédron, T. *et al.* A crypt-specific core microbiota resides in the mouse colon. *MBio* **3,** e00116–12 (2012).
24. Mundy, R., MacDonald, T. T., Dougan, G., Frankel, G. & Wiles, S. *Citrobacter rodentium* of mice and man. *Cell. Microbiol.* **7,** 1697–1706 (2005).
25. Bloom, S. M. *et al.* Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9,** 390–403 (2011).
26. Freter, R., Brickner, H., Botney, M., Cleven, D. & Aranki, A. Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infect. Immun.* **39,** 676–685 (1983).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Thron and S. McBride for the maintenance of germ-free animals, J. Selicha for assisting with the experimental procedures and G. Chodaczek for help with confocal and two-photon microscopy. We are grateful to E. C. Martens and members of the Mazmanian laboratory for critical review of the manuscript. S.M.L. and G.P.D. were supported by a pre-doctoral training grant (GM007616). This work was supported by grants from the National Institutes of Health (GM099535 and DK078938) and the Crohn's and Colitis Foundation of America to S.K.M.

Author Contributions S.M.L. and S.K.M. conceived the project. S.M.L. performed most of the experiments; G.P.D., Z.M. and S.B. contributed data. S.M.L., G.P.D., Z.M., K.L. and S.K.M. interpreted the data. K.L. and S.K.M. secured funding. S.M.L. and S.K.M. wrote the manuscript. G.P.D., Z.M. and K.L. edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.K.M. (sarkis@caltech.edu).

METHODS

Bacterial strains, plasmids and culture conditions. Bacterial strains and plasmids are described in Supplementary Table 1. *Bacteroides* strains were grown anaerobically at 37 °C for 2 days in brain heart infusion broth supplemented with 5 μg ml^{-1} hemin and 0.5 μg ml^{-1} vitamin K (BHIS), with gentamicin (200 μg ml^{-1}), erythromycin (5 μg ml^{-1}), chloramphenicol (10 μg ml^{-1}) and tetracycline (2 μg ml^{-1}) added where appropriate. *E. coli* JM109 containing recombinant plasmids were grown in luria broth (LB) with ampicillin (100 μg ml^{-1}) or kanamycin (30 μg ml^{-1}). *C. rodentium* DBS100 strain was grown in LB at 37 °C for 24 h. For the induction of *susC/D* homologues, *B. fragilis* and *B. thetaiotaomicron* were grown in minimal medium with either glucose or N-acetyllactosamine as the sole carbon source as described previously[20].

Mice. 8–10-week old male and female germ-free Swiss Webster mice were purchased from Taconic Farms and bred in flexible film isolators. For gnotobiotic colonization experiments, germ-free mice were transferred to freshly autoclaved microisolator cages, fed *ad libitum* with a standard autoclaved chow diet and given autoclaved water supplemented with 10 μg ml^{-1} of erythromycin and 100 μg ml^{-1} of gentamicin. Male SPF C57BL/6 mice and Swiss Webster mice were purchased from Taconic Farms. Male SPF NOD/ShiLtJ mice and *Rag*$^{-/-}$ C57BL/6 mice were purchased from the Jackson Laboratory. No randomization or blinding was used to allocate experimental groups. Sample size and standard deviation were based on empirical data from pilot experiments. All procedures were performed in accordance with the approved protocols using IACUC guidelines of the California Institute of Technology.

Construction of chromosomal library and screen. Genomic DNA was isolated from overnight culture of *B. fragilis* using a commercial kit (Wizard Genomic DNA Purification Kit, Promega). 20 μg of genomic DNA was incubated with 4 U of Sau3AI for 5, 10, 15 or 20 min at 37 °C in 50 μl volume and the partially digested genomic DNA was separated by electrophoresis on 0.7% agarose gel. 9–10-kb fragment DNA was excised and recovered from the agarose gel (Zymoclean Gel DNA Recovery Kit, Zymo Research). Insert DNA was ligated to BglII site of plasmid vector (pFD340-catBII, Supplementary Table 1), transformed into *E. coli* and amplified on LB-ampicillin plate. Individual clones from the plasmid library were mobilized from *E. coli* to *B. vulgatus* by conjugal helper plasmid RK231 generating a library of *B. vulgatus* hosting *B. fragilis* chromosomal DNA fragments consisting of approximately ~2,100 clones. To screen the library *in vivo*, pools of 96 clones (10^6 c.f.u. of each clone) were gavaged into 22 germ-free Swiss Webster mice (10^8 c.f.u. per animal) pre-colonized with *B. vulgatus* pFD340 for 1–2 weeks. Two weeks after gavage, fresh faecal samples were plated on BHIS agar plate containing chloramphenicol to select for clones with colonization phenotype.

Generation of *ccfA*, *ccfC*, *ccfD*, *ccfE*, *ccfC–E* (ΔCCF) and ΔBF0227–0229 deletion mutants. ~2-kb DNA segments flanking the region to be deleted were PCR amplified using primers listed in Supplementary Table 2. Reverse primer of the left-flanking DNA and forward primer of the right-flanking DNA were designed to be partially complementary at their 5′ ends by 18–21 bp. Fusion PCR was performed using the left and right flanking DNA (~300 ng each after gel purification) as DNA template and forward primer of the left-flanking DNA and reverse primer of the right-flanking DNA[27]. The fused PCR product was cloned into BamHI or SalI site of the *Bacteroides* conjugal suicide vector pNJR6 and mobilized into *B. fragilis*. Colonies selected for erythromycin resistance (Em^r), indicating integration of the suicide vector into the host chromosome were passaged for 5 days and then plated on nonselective medium (BHIS). The resulting colonies were replica plated to BHIS containing Em, and Em^s (erythromycin sensitive) colonies were screened by PCR to distinguish wild-type revertants from strains with the desired mutation. The same strategy was used to generate Δ*ccfC–E* deletion mutants from *B. vulgatus*.

qRT–PCR. Total RNA was extracted from mid-log phase bacterial culture using ZR Fungal/Bacterial RNA MiniPrep (Zymo Research), faeces and caecal content from mice using ZR Soil/Fecal RNA MicroPrep (Zymo Research), and mouse colon tissues after removing luminal content by gently scraping the mucosal surface and PBS rinse using Trizol (Invitrogen). Complementary DNA was made using an iSCRIPT cDNA synthesis kit per manufacturer's instructions (Bio-Rad). All qRT–PCR reactions were performed in ABI PRISM 7900HT Fast Real-Time PCR System (Applied Biosystems) using Power SYBR Green PCR Master Mix (Applied Biosystems). Gene-specific primers are described in Supplementary Table 2.

Immunofluorescent staining of colon whole-mounts and frozen sections. For whole-mount staining, colons were fixed in buffered 4% paraformaldehyde, washed with PBS and subjected to indirect immunofluorescence. Tissues were made permeable by incubation with 0.5% (wt/vol) saponin, 2% (vol/vol) FBS and 0.09% (wt/vol) azide in PBS for at least 18 h. The same buffer was used for subsequent incubations with antibodies. Colon fragments were incubated with a primary polyclonal chicken IgY anti-*B. fragilis* antibodies for 12–16 h at room temperature followed by 1–2 h incubation at 37 °C. Following PBS washes, samples were reacted with goat anti-chicken IgY secondary antibodies (Alexa Fluor 488 or Alexa Fluor 633, 2 μg ml^{-1}, Molecular Probes), fluorescently labelled phalloidin (fluorescein or AF568, 2 U ml^{-1}, Molecular Probes) and DAPI (2 μg ml^{-1}, Molecular Probes) for 1 h at room temperature. Tissues were mounted in Prolong Gold (Invitrogen) and allowed to cure for at least 48 h before imaging. In some experiments, anti-*B. fragilis* antibodies were pre-absorbed on tissue fragments derived from either germ-free mice (up to 18 h at room temperature) or SPF mice (1 h at room temperature).

For frozen sections, colon tissues were embedded in OCT Compound (Sakura Finetek), frozen on dry ice and stored at −80 °C. Frozen blocks were cut with a thickness of 10 μm using a Microm HM505E cryostat, and sections were collected on positively charged slides (Fisher Scientific) for staining. Slides were fixed with 4% buffered paraformaldehyde for 10 min and washed 2 × 10 min with PBS. Tissue sections were blocked with 10% normal goat serum and 0.5% bovine serum albumin in PBS for 1 h at room temperature. Sections were incubated with anti-*B. fragilis* antibodies for at least 8 h at 4 °C, washed twice for 10 min with PBS, reacted with secondary reagents and mounted as described above. In some experiments, anti-*B. fragilis* antibodies were pre-absorbed for 1 h at room temperature on tissue sections derived from germ-free mice.

Fluorescence microscopy. An SP5 resonant laser-scanning confocal and two-photon microscope (both scanning heads mounted on the same DM 6000 upright microscope, Leica Microsystems) with a 40× oil objective (numerical aperture 1.4) or 63× oil objective (numerical aperture 1.4) were used for fluorescence microscopy. Images used for three-dimensional reconstructions were acquired using dual confocal–two-photon mode. For confocal imaging, 488-nm and 543-nm excitation wavelengths were used for Alexa Fluor 488-labelled bacteria and Alexa Fluor 568-labelled phalliodin, and signals were detected with internal photomultiplier tubes. Two-photon imaging was performed with four nondescanned detectors (Leica Microsystems) and a Chameleon Ultra Ti: Sapphire laser (Coherent) tuned at 700–800 nm for acquisition. Emitted fluorescence was split with three dichroic mirrors (496 nm, 560 nm and 593 nm) and passed through an emission filter (Semrock) at 585/40 nm. Images (512 × 512) acquired with a 0.5-μm *Z* step were smoothed by median filtering at kernel size 3 × 3 pixels. Three-dimensional reconstructions of crypts and bacteria were performed using Imaris software (version 7.5.1 × 64; Bitplane AG). Crypt structures were visualized by DAPI and phalloidin signals. Images used for quantification were acquired with FluoView FV10i confocal microscope (Olympus) using 60× (numerical aperture 1.35) oil objective.

Image analysis. For bacterial localization with respect to the epithelial layer, frames of 512 × 512 pixels were acquired with 1-μm *Z* steps in the crypt length axis. Images were processed using ImageJ software (NIH). Background was subtracted (rolling ball method), images were smoothed by median filtering (3 × 3 pixels), segmented by threshold and position of the signal in the *Z* stack was recorded. Data did not follow normal distribution and were analysed by non-parametric two-sided Mann–Whitney *U*-test.

For quantification of crypt-associated bacterial signals from antibiotic-treated animals, stacks of 512 × 512 pixels by eight frames (1 μm per frame) were flattened by maximum intensity projection and filtered by median (3 × 3 kernel size). Images were segmented by thresholding. Number of positive spots per 1,000 μm^2 and area occupied by individual spots were analysed. Data were not normally distributed and were analysed by Mann–Whitney or Kruskal–Wallis followed by Dunn's multiple comparisons test where appropriate. 11-13 stacks/group were examined. Total area that was analysed within the group of stacks was between 0.08–0.2 mm^2.

Gnotobiotic animal colonization experiments. 8–12-week-old germ-free Swiss Webster mice were gavaged once with a 100 μl of bacterial suspension for mono-association (~10^8 c.f.u. of each bacterial strain collected from a log-phase culture and re-suspended in PBS with 1.5% $NaHCO_3$). For sequential colonization, germ-free mice were mono-associated with an initial strain for 6–7 days and subsequently gavaged with a 100 μl suspension of a challenge strain. All *Bacteroides* strains used to colonize germ-free animals were resistant to gentamicin inherently, and to erythromycin by plasmid. Unless otherwise indicated, the initial strains carried pFD340-*cat* (chloramphenicol resistant; Cm^r) and the challenge strains, pFD340-*tetQ* (tetracycline resistant; Tet^r). For horizontal transfer by encounter experiment, two single-housed mice that were mono-associated with either wild-type *B. fragilis* pFD340-*tetQ* or BFΔCCF pFD340-*cat* for at least 3 weeks were co-housed in a fresh sterile cage for 4 h and then separated. At each time point, fresh faecal samples were collected, weighed, homogenized and serially diluted in PBS (or BHI broth) for plating on selective media to determine bacterial c.f.u. per g of faeces.

SPF animal colonization experiments. 7–8-week-old male SPF mice (C57BL/6, Swiss Webster, NOD and *Rag*$^{-/-}$) were given a single inoculum of 1 × 10^8 c.f.u. of either wild-type *B. fragilis*, BFΔCCF or 1:1 mixture of the two strains by oral

gavage. At each time point, bacterial genomic DNA from faecal samples were isolated using a commercial kit (ZR Fecal DNA MiniPrep, Zymo Research) following the manufacturer's instructions and the relative densities of bacteria were determined by qPCR using strain-specific primers (Supplementary Table 2).

***C. rodentium* infection.** 8-week-old female SPF Swiss Webster mice were treated with metronidazole (100 mg kg^{-1}) by oral gavage every 24 h and ciprofloxacin dissolved in drinking water (0.625 mg ml^{-1}; Hikma Pharmaceuticals) for 7 days; mice were transferred to a fresh sterile cage every 2 days. 2 days after the cessation of antibiotic treatment, mice were orally gavaged with a single inoculum of 1:1 mixture of wild-type *B. fragilis* and BFΔCCF ($\sim 5 \times 10^8$ c.f.u. total per animal). 6–7 days after *B. fragilis* gavage, mice were either infected orally with $\sim 5 \times 10^8$ c.f.u. of overnight culture *C. rodentium* or PBS-gavaged as control. The relative densities of bacteria were determined by faecal bacterial DNA extraction and qPCR.

Antibiotic treatment. 8-week-old female SPF Swiss Webster mice were treated with metronidazole (100 mg kg^{-1}) by oral gavage every 24 h and ciprofloxacin dissolved in drinking water (0.625 mg ml^{-1}) for 7 days; mice were transferred to a fresh sterile cage every 2 days. 2 days after the cessation of antibiotic treatment, mice were orally gavaged with a single inoculum of 1:1 mixture of wild-type *B. fragilis* and BFΔCCF ($\sim 5 \times 10^8$ c.f.u. total per animal). 6–7 days after *B. fragilis* gavage, one group of mice were treated with ciprofloxacin for 4 days dissolved in drinking water (1 mg ml^{-1}) and another group were left untreated. The relative densities of bacteria were determined by faecal bacterial DNA extraction and qPCR.

Antibiotic treatment for colon whole-mount imaging. 8-week-old female SPF Swiss Webster mice were treated with metronidazole (100 mg kg^{-1}) by oral gavage every 24 h and ciprofloxacin dissolved in drinking water (0.625 mg ml^{-1}) for 7 days; mice were transferred to a fresh sterile cage every 2 days. 2 days after the cessation of antibiotic treatment, mice were orally gavaged with a 100 µl inoculum of *B. fragilis* ($\sim 10^8$ c.f.u.) or PBS. 7 days after bacterial gavage (day 16), PBS or *B. fragilis* inoculated mice were treated with ciprofloxacin in drinking water (1 mg ml^{-1}) for 7 days and one group of *B. fragilis* inoculated mice were left untreated. At the end of the ciprofloxacin treatment (day 23), faeces were collected for stool DNA extraction and colon tissues were collected and fixed with 4% paraformaldehyde for whole-mount imaging.

27. Wurch, T., Lestienne, F. & Pauwels, P. J. A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. *Biotechnol. Tech.* **12,** 653–657 (1998).

Cell Host & Microbe
Resource

Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat

Andrew L. Goodman,[1] Nathan P. McNulty,[1] Yue Zhao,[1] Douglas Leip,[1] Robi D. Mitra,[1] Catherine A. Lozupone,[1,2] Rob Knight,[2] and Jeffrey I. Gordon[1,*]
[1]Center for Genome Sciences, Washington University School of Medicine, St. Louis, MO 63108, USA
[2]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
*Correspondence: jgordon@wustl.edu
DOI 10.1016/j.chom.2009.08.003

SUMMARY

The human gut microbiota is a metabolic organ whose cellular composition is determined by a dynamic process of selection and competition. To identify microbial genes required for establishment of human symbionts in the gut, we developed an approach (insertion sequencing, or INSeq) based on a mutagenic transposon that allows capture of adjacent chromosomal DNA to define its genomic location. We used massively parallel sequencing to monitor the relative abundance of tens of thousands of transposon mutants of a saccharolytic human gut bacterium, *Bacteroides thetaiotaomicron*, as they established themselves in wild-type and immunodeficient gnotobiotic mice, in the presence or absence of other human gut commensals. In vivo selection transforms this population, revealing functions necessary for survival in the gut: we show how this selection is influenced by community composition and competition for nutrients (vitamin B_{12}). INSeq provides a broadly applicable platform to explore microbial adaptation to the gut and other ecosystems.

INTRODUCTION

Our indigenous microbial communities play critical roles in shaping myriad features of our biology. The distal gut hosts the majority of our microbes; these include representatives of all three domains of life, plus their viruses. The density of organisms occupying this habitat is astonishing, exceeding 10^{12} cells/mL. Most phylogenetic types (phylotypes) observed in the guts of humans and other mammals belong to just two bacterial divisions (phyla)—the Firmicutes and the Bacteroidetes (Ley et al., 2008). Microbial community (microbiota) exchange experiments indicate that gut community members are dynamically selected: for example, transplantation of a Proteobacteria-dominated zebrafish gut microbiota into germ-free mice transforms this community so that it comes to resemble a mouse gut microbiota, while transplantation of a mouse microbiota into germ-free zebrafish has the opposite effect, yielding a community that has the phylum-level characteristics of the native zebrafish microbiota (Rawls et al., 2006).

Within the Firmicutes and Bacteroidetes, hundreds to thousands of phylotypes partition available niches (professions) to create a community able to maintain itself in this continuously perfused ecosystem despite shifts in host diet, regular ingestion of foreign bacteria, intense resource competition, high bacteriophage levels, and immune surveillance. Comparisons of the sequenced genomes of cultured representatives of major gut phylogenetic lineages provide a means for identifying genomic features potentially important for colonization and competition in the gut. However, the recent surge in microbial genome sequencing projects has far outpaced development of broadly applicable tools for directly testing the role of genes in determining fitness in this habitat. The paucity of tools is unfortunate, as fundamental questions connecting genome content to function remain unexplored. For example, how are the determinants of fitness related to nutrient availability, and how closely do the genes required for fitness in vivo mirror those required for maximizing growth rate in vitro? Does community structure influence this map of genetic requirements, or is competition largely "within species"? Do the major recognized components of the host immune system play a dominant role in determining which genes are critical for symbiont fitness in vivo? To address these questions, we integrated a simple and broadly applicable genetic tool with second-generation DNA sequencers and gnotobiotic mouse models to identify fitness determinants in the genome of a human gut mutualist.

Mariner transposon mutagenesis is an attractive forward genetic strategy for connecting phenotype to gene because stable random insertions can be generated in a recipient genome without specific host factors: the ability of these transposons to serve as mutagenic agents is well established in members of all three domains of life (Lampe et al., 1996; Mazurkiewicz et al., 2006). After alignment of the inverted repeat (IR) sequences that delimit *mariner* family transposons, we noted that a single G-T transversion at a nonconserved position would create a recognition sequence for the type IIs restriction enzyme MmeI. When directed to this location, MmeI would cleave 16 bp outside of the transposon, capturing a genomic fragment that identifies the insertion site. Moreover, if the genome sequence of the recipient organism were known, the short genomic DNA sequences captured by this MmeI digestion would be sufficient to uniquely map transposon location. We reasoned that in a mixed population of transposon mutants produced in a given recipient bacterial species, the relative abundance of each MmeI-liberated genomic fragment, identified after limited PCR amplification and massively parallel sequencing, would in

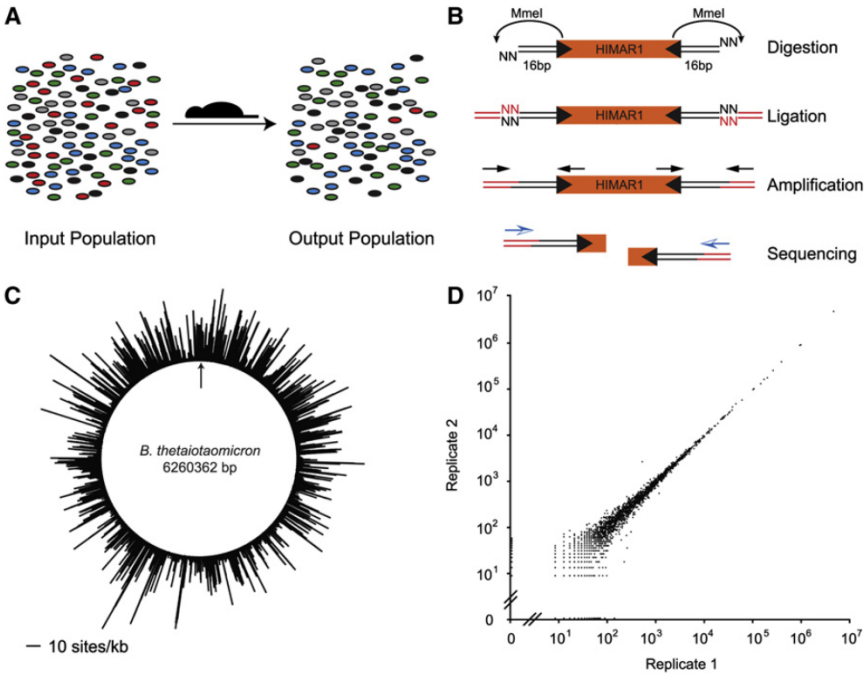


Figure 1. Mapping and Quantifying Tens of Thousands of Transposon Insertion Strains by High-Throughput INSeq

(A) A negative selection scheme for identification of genes required for colonization in vivo. Mutants in genes important for competitive growth (red) are expected to decrease in relative abundance in the output population.

(B) Preparation of an INSeq library. Genomic DNA is extracted from the mutagenized bacterial population, digested with MmeI, and separated by polyacrylamide gel electrophoresis (PAGE). Transposon-sized fragments are appended with double-stranded oligonucleotide adapters by ligation. Limited cycles of PCR create the final library molecules for sequencing.

(C) Map of insertion sites in the *B. thetaiotaomicron* genome. An arrow marks the origin of replication.

(D) Reproducibility of library preparation and sequencing protocols. Technical replicates were prepared and sequenced from a single transposon mutant population. Each point represents the abundance of insertions in a single gene; the coefficient of determination, R^2, on log-transformed abundance values is 0.92.

principle mirror the abundance of each mutant in the population. Putting this population under potentially selective conditions (e.g., colonization of the intestines of gnotobiotic mice) could then be used to highlight mutants that change in relative abundance, and thereby identify genes and pathways critical for fitness under these conditions (Figure 1A).

Using this insertion-sequencing (INSeq) approach, we subjected a mutagenized population of the prominent human gut symbiont, *Bacteroides thetaiotaomicron*, to varying selective pressures: in vitro growth in continuous flow chemostats, monoassociation of wild-type and knockout germ-free mice lacking major branches of their innate or acquired immune systems, and as one component of several defined in vivo communities of human gut-derived microbes. The results provide evidence that human gut symbionts, like their pathogenic counterparts, possess dedicated mechanisms critical for interaction with their host and each other. The relative importance of these mechanisms is not static but instead is shaped by other members of the microbiota. As a strategy for functional characterization of newly sequenced genomes in general, and of the human gut microbiome in particular, INSeq extends existing techniques in several important aspects.

RESULTS

We constructed pSAM, a sequencing-adapted *mariner* transposon delivery vector with three major features: an antibiotic resistance cassette flanked by MmeI-modified *mariner* IRs, a multiple cloning site immediately upstream of the *himar1C9* mariner transposase (Lampe et al., 1999), and machinery for replication in the donor strain and transfer by conjugation (see Figures S1A–S1C available online). *B. thetaiotaomicron* was chosen as the recipient species to test this approach for several reasons (Figure S2A). First, it is highly adapted to life in the distal human gut (Zocco et al., 2007). A prominent member of the gut microbiota, this mutualist is richly endowed with a broad arsenal of genes encoding glycoside hydrolases and polysaccharide lyases not represented in the human genome (Xu et al., 2003). These genes are incorporated together with genes encoding nutrient sensors and carbohydrate transporters into 88 polysaccharide utilization loci (PULs) representing 18% of the organism's genome (Martens et al., 2008). Thus equipped, *B. thetaiotaomicron* functions as a flexible forager of otherwise indigestible dietary glycans, as well as host glycans when dietary polysaccharides are not available (Martens et al., 2008). Second, >200 GeneChip data sets of *B. thetaiotaomicron*'s transcriptome have been collected during growth in vitro under a variety of conditions, after monocolonization of germ-free mice fed different diets, as well as after cocolonization with another human gut bacterial or archaeal species (NCBI GEO archive). Third, functional genomic studies conducted in gnotobiotic mice have shown that monoassociation with *B. thetaiotaomicron* can recapitulate a number of host responses evoked by a complete mouse gut microbiota. Fourth, a limited number of in vivo competition experiments conducted in gnotobiotic mice colonized with isogenic wild-type and mutant *B. thetaiotaomicron* have identified a few fitness determinants that could serve as reference controls for the present study (Peterson et al., 2007).

Construction and Characterization of a Transposon Mutant Population by INSeq

We found that transfer of the MmeI-modified *mariner* transposon into the genome of *B. thetaiotaomicron* occurs with high efficiency (Figure S2). To identify the site of transposon insertion and the relative abundance of each mutant in an otherwise isogenic population, we developed a straightforward procedure to extract the two 16 bp genomic sequences adjacent to each

transposon, append sequencing adapters to these fragments, and separate the desired molecules from genomic background (Figure 1B and Supplemental Experimental Procedures). Sequencing these tags using an Illumina Genome Analyzer II produced ~8 million raw reads from a single flow cell lane: ~90% of these reads contained the transposon (Figure S2D). We designed a software package, MapSAM, to filter out low-quality sequences, quantify and pair reads generated from either side of an insertion, and assign these paired reads to a specific location in the target genome (Figure 1C). Examination of technical replicates indicated that the library preparation, sequencing, and mapping strategies were highly reproducible (Figure 1D and Figure S3). The proportion of reads that were unambiguously mapped (98%) matched predictions from an in silico model of random transposon insertion (Figure S4A); no insertion sequence bias beyond the known "TA" dinucleotide requirement (Bryan et al., 1990) was apparent (Figures S4B and S4C).

We first characterized a mutant population containing ~35,000 *B. thetaiotaomicron* transposon insertion strains. Insertions were well distributed across the genome at an average density of 5.5 insertions/kb (Figure 1C). After filtering out insertions in the distal (3′) 10% of any coding region (because such insertions could possibly still permit gene function), we found that 3435 of the 4779 predicted open reading frames in the genome (72%) had been directly disrupted in the mutant population. Inclusion of genes disrupted by upstream (polar) mutations in a predicted operon increased this number to 78%; rarefaction analysis suggested that this population is approaching saturation (Figure S4D).

To identify *B. thetaiotaomicron* genes unable to tolerate transposon insertion, we generated and mapped a second, independent mutant population. We combined both data sets and applied a Bayesian model to account for the number of informative insertion sites in each gene (Lamichhane et al., 2003). The results yielded a conservative list of 325 candidate essential genes for growth under anaerobic conditions when plated onto rich (tryptone-yeast extract-glucose; TYG) medium (Table S1). These genes were significantly enriched (Benjamini-Hochberg corrected $p < 0.05$) for Clusters of Orthologous Groups (COG) categories representing cell division (category D), lipid transport/metabolism (I), translation/ribosomal structure/biogenesis (J), and cell-wall/membrane biogenesis (M). This is consistent with genome-wide mutagenesis studies of *Escherichia coli* and *Pseudomonas aeruginosa* (Baba et al., 2006; Jacobs et al., 2003). For nonessential genes, insertion frequency showed some correlation (R^2 of log-transformed values = 0.33) with GeneChip-defined expression levels during mid-log phase growth of the parental wild-type strain in batch fermentors containing TYG medium (Figure S4E). The reason for this relationship is not known, although studies of *mariner* transposition in vitro suggest that the enzyme has a preference for bent or bendable DNA (Lampe et al., 1998).

Combinatorial Mapping of Individual Insertion Strains from an Archived Mutant Collection

A mutant population of this complexity contains insertions in most of the coding potential of the genome and can facilitate forward genetic approaches for identifying genotypes connected with a phenotype of interest. These mixed populations, however, are less amenable to reverse genetics: specific genotypes are not individually retrievable. Arrayed transposon mutant collections, in which strains of known genotype are archived individually, provide an important avenue for retrieval and further study of specific strains of interest. To date, such collections have been created by using a strain-by-strain procedure that typically consists of cell lysis, removal of cellular debris, multiple rounds of semirandom or single-primer PCR, DNA cleanup, and individual Sanger sequencing of each amplicon. As an alternative, we developed a combinatorial technique for simultaneously mapping thousands of individually archived transposon mutant strains in parallel (Figure 2A, Figure S5, and Supplemental Experimental Procedures).

This approach consists of three basic steps: (1) culturing and archived storage of randomly picked mutant colonies in individual wells of 96-well (or higher density) plates, (2) placement of each of these strains into pools in unique patterns, and (3) sequencing of these pools by INSeq in order to associate each transposon insertion location with a strain in the original arrayed multiwell plates. Because n pools can contain 2^n unique presence/absence patterns, a small number of pools can uniquely identify a large number of strains. To this end, a benchtop liquid-handling robot was used to distribute archived transposon mutant strains across a subset of pools in a pattern selected to minimize the likelihood of mistaking one strain for another or incorrectly mapping clonal strains (Figure 2B and Supplemental Experimental Procedures). Libraries were then prepared from each pool using the same method described in Figure 1B, except that a pool-specific, barcoded adaptor (Table S2) was used in the ligation step. These libraries were combined into a single sample that can be sequenced with an Illumina Genome Analyzer II using just one lane of the instrument's eight-lane flow cell. Reads were first mapped to the reference *B. thetaiotaomicron* genome to determine insertion sites; to assign an insertion site to a specific archived strain, the pool-specific barcodes associated with a given insertion location on the chromosome were then matched with the patterns assigned to the strains in the original archived set of plates.

Using this strategy, we were able to identify the insertion coordinates for over 7000 individually archived *B. thetaiotaomicron* transposon mutant strains in parallel (Table S3). To verify the accuracy of these assignments, we first used ELISA to test strains predicted to have lost reactivity to two monoclonal antibodies of known specificity (Peterson et al., 2007). We also amplified transposon-genome junctions of test strains by semirandom PCR and sequenced the amplicons. In total, 179 of 183 strains tested (98%) produced the anticipated results (Figure 2C), confirming that combinatorial barcoding and INSeq can be used to efficiently and economically generate archived, sequence-defined, mutant collections.

Identification of Genes Required for Fitness In Vitro

To identify genes that contribute to exponential growth in nutrient-rich conditions in vitro, we maintained a 35,000-strain mutant population in this growth phase (OD_{600} 0.1–0.4), under anaerobic conditions, in chemostats that were continuously supplied with fresh TYG medium. Output populations were

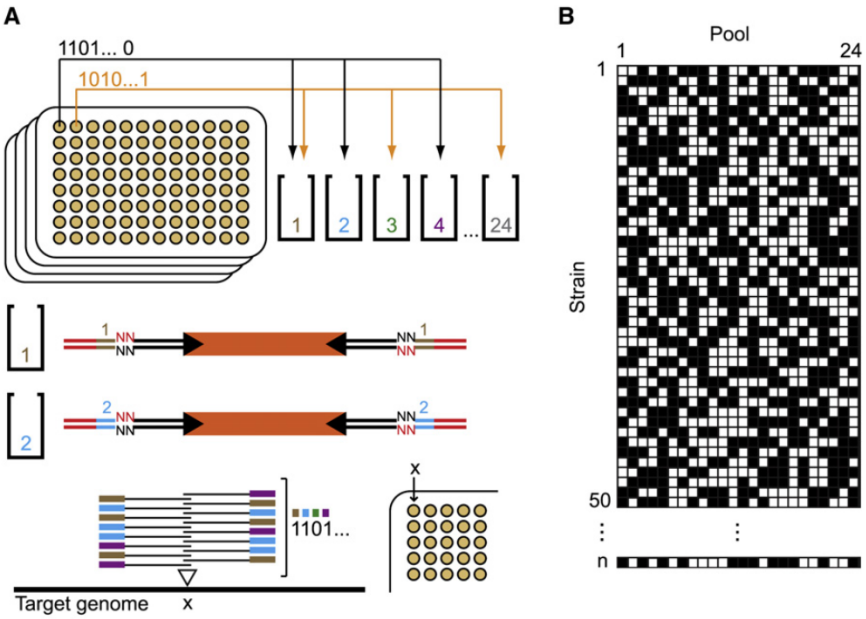


Figure 2. Mapping an Archived Strain Collection by Combinatorial Pooling and High-Throughput Sequencing

(A) Individual strains are archived in a 96-well format and then placed into a subset of 24 pools according to a unique 24-bit binary string assigned to each strain. Libraries are prepared from each of these pools using the workflow shown in Figure 1B, except that one of 24 pool-specific barcoded dsDNA adaptors is employed in the ligation step. Libraries are then combined and sequenced in a single run. Reads mapping to a specific insertion location are compiled and the associated pool-specific barcodes are identified to recreate the 24-bit string and, with it, the physical location of the corresponding strain in the archived collection. For details, see the Supplemental Experimental Procedures.

(B) Sample pool distribution patterns for archived strains. A given strain (row) is placed in pools (columns) designated with white boxes and omitted from pools marked with black boxes. The patterns generated are each distinct in at least six positions and do not overlap to produce another pattern in the set. For details, see the Supplemental Experimental Procedures.

(C) Confirmation of strain-insertion assignments.

C

Method	Tested	Confirmed
ELISA	144	141
Direct Sequencing	35	34
PCR	4	4
Total	183	179 (98%)

sampled from four independent chemostats; two after ~15 hr of continuous exponential growth and two after ~45 hr of growth. DNA was prepared from each of these populations, and transposon-adjacent genomic fragments were identified by INSeq. After removing insertions in the 3′ 10% of each gene, a z test was applied to identify genes that show a significantly altered representation from the overall distribution of output:input abundance ratios after q value correction for multiple hypothesis testing ($q < 0.05$; see the Supplemental Experimental Procedures). To test this approach for identifying fitness determinants, we also examined insertions in intergenic "neutral loci" (Figure S6A). None of these 80 control regions passed the statistical cutoff for underrepresentation (three increased in abundance). In contrast, 477 genes (~14% of the genes represented in the input population) showed a statistically significant change in abundance after in vitro selection (265 underrepresented/212 overrepresented; Figure S6B and Table S4). Consistent with selection for maximal growth rate in rich medium, the list of factors important for fitness under these conditions was significantly enriched in genes annotated as being in COG category C (energy production and conversion).

As a proof of principle, we asked whether an observed enrichment in a broad functional group (COG category), represented among genes required for fitness, could be altered by manipulating environmental conditions. To do so, we harvested exponentially growing cells from mutant populations grown in minimal defined medium in the presence or absence of exogenous amino acids and quantified the abundance of transposon mutants by INSeq (n = 4 replicate populations, each assayed independently). Gratifyingly, the set of genes required for fitness specifically in the amino acid-depleted condition was most highly enriched ($p < 0.0005$) in COG category E (amino acid transport/metabolism) (Figure S7 and Table S5).

Genes Required for Establishment of *B. thetaiotaomicron* within the Distal Gut of Monoassociated Gnotobiotic Mice

To survey the *B. thetaiotaomicron* genome for genes critical for fitness in a mammalian gut ecosystem, we colonized germ-free mice with a single gavage of approximately 10^8 colony-forming units (CFUs) of the 35,000-strain mutant population (n = 15 animals representing three independent experiments, each involving a cohort of five 8- to 12-week old C57BL/6J males; experiments were performed ~3 months apart). The five animals in each cohort were caged individually in a shared gnotobiotic isolator and fed a standard, autoclaved, polysaccharide-rich, low-fat chow diet ad libitum. The relative abundance of each mutant strain in the cecal bacterial population was defined at the time of sacrifice 14 days after gavage; this interval between gavage and sacrifice encompassed several cycles of turnover of the mucus layer and the underlying gut epithelium, and is sufficient to allow mobilization of innate and adaptive immune responses (Peterson et al., 2007).

All recipients of the gavage harbored equivalent levels of *B. thetaiotaomicron* at the time of sacrifice (~10^{11}–10^{12} CFU/mL cecal contents as quantified by plating and by qPCR). Moreover, the relative representation of mutants was consistent

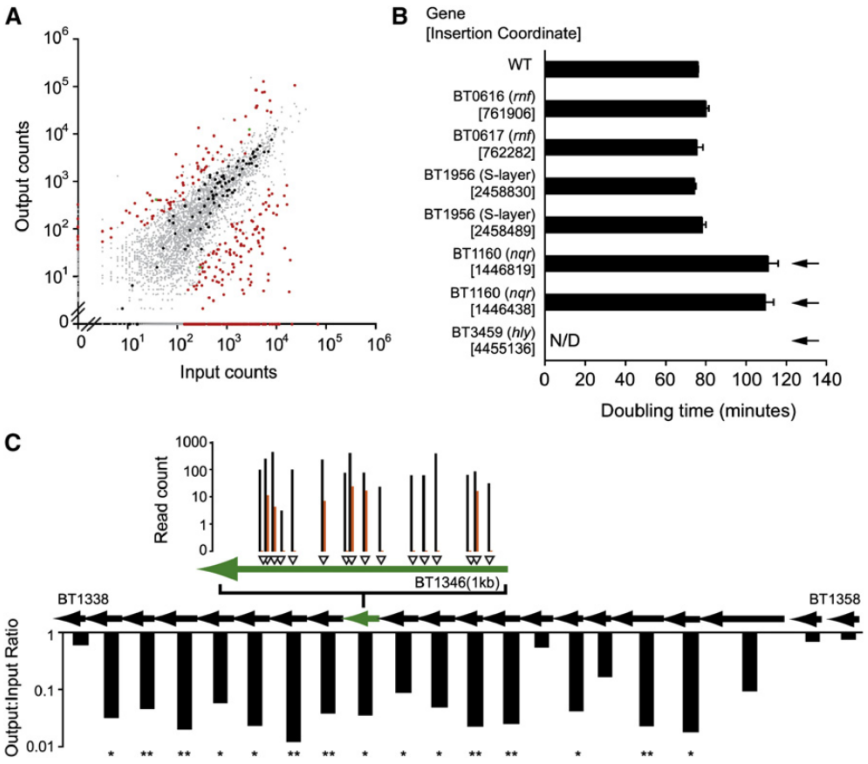


Figure 3. Identification of Genetic Determinants of Fitness In Vivo

(A) The transposon mutant population is largely stable in vivo. The relative abundance of mutations in each gene (points) was compared between input and output (median from wild-type monoassociated mice, n = 15) populations. Genes that show a statistically significant change ($q < 0.05$) in representation in all three cohorts of mice are shown in red, others in gray. The relative abundances of 80 gene-sized neutral loci are shown in black (no significant change) and green (three neutral loci that pass the significance criteria).

(B) Individual strains retrieved from the archived collection demonstrate that genes critical for fitness in vivo are dispensable in vitro. Each strain was cultured individually in TYG medium, and doubling time was calculated from OD_{600} measurements; error bars represent one standard deviation based on triplicate experiments. Mutants predicted by INSeq to have in vitro growth defects are marked with arrows. N/D, no growth detected. *rnf*, Na^+-transporting NADH:ubiquinone oxidoreductase (BT0616-22); S layer, putative S layer locus (BT1953-7); *nqr*, Na^+-translocating NADH-quinone reductase (BT1155-60); *hly*, hemolysin A (BT3459).

(C) Insertions in the *CPS4* locus (BT1338-58) show a consistent in vivo fitness defect. Individual insertion locations (open arrowheads) in a representative gene (BT1346; green arrow) are shown at top. Read counts for input (black) and output (orange) samples at each insertion location are indicated (output counts represent the median from the ceca of wild-type monoassociated mice, n = 15). Median output:input ratios for each gene (black/green arrows) across the *CPS4* locus are shown below. Asterisks indicate the average FDR-corrected p value (q) across three experimental cohorts (n = 5 mice/cohort): *$q < 0.05$; **$q < 0.01$.

between the ceca of individual mice (Tables S9–S11) and for most genes reflected their abundance in the input population (Figure 3A). However, compared to the input population, mutants in 370 genes showed significantly ($q < 0.05$) altered representation (90 overrepresented/280 underrepresented) in all three cohorts of mice (Table S6; note that the largest category of genes identified in this screen encode hypothetical or conserved hypothetical proteins). Only one of the 80 "neutral intergenic" controls described above was significantly underrepresented in these populations (two were overrepresented).

While the smaller group of 90 genes that produce a competitive advantage in the cecum when mutated (highlighted in Table S6 in blue and green) are not significantly enriched in any broad predicted functional (COG) categories, only half can be explained by corresponding behavior in chemostats containing rich medium. The underrepresented genes show a similar trend: half (146) of the 280 genes critical for in vivo fitness can be predicted from growth defects in rich medium; the remainder (134/280), which are highlighted in Table S6 in yellow, do not show a defect after prolonged exponential growth in vitro. These include loci with diverse predicted functions, including assembly of polysaccharide- and protein-based surface structures (BT1339-55, BT1953-7), synthesis and utilization of vitamin B_{12}-dependent cofactors (BT2090-1, BT2760), and an *rnf*-like oxidoreductase complex (BT0616-22) (e.g., Figure S8).

To confirm that the requirement for these genes in vivo could not be explained by general growth defects, we analyzed individual mutant strains retrieved from the archived strain collection. This collection contained sequence-defined, single-insertion transposon mutants in ~70% of the genes designated as critical for survival in the distal gut in vivo (~80% if predicted polar effects are included) (Table S3). After validating the site of selected transposon insertions by semirandom PCR and Sanger sequencing, we determined the exponential doubling time of representative strains individually (Figure 3B). Strains carrying transposon insertions in genes uniquely required in vivo had an in vitro growth rate similar to wild-type *B. thetaiotaomicron*, further suggesting that the critical function of these genes in vivo cannot be simply explained by a necessity for sustaining exponential growth in rich medium. In contrast, mutants that exhibited a competitive defect in the 35,000-strain population both in vitro and in vivo had a slower doubling time when cultured individually, suggesting that these genes play a basic role in bacterial cell physiology.

An earlier report from our group used a targeted mutagenesis strategy to disable expression of genes encoding capsular polysaccharide (CPS) 4 in this organism; this strain was rapidly displaced by wild-type *B. thetaiotaomicron* after initial inoculation as a 1:1 mixture into germ-free mice (Peterson et al., 2007). The 35,000-strain transposon mutant population recapitulated this observation (Figure 3C and Figure S9). The transposon mutant population included over 1100 independent insertions across 143 genes that span all eight CPS loci encoded in the *B. thetaiotaomicron* genome. None of the other CPS loci

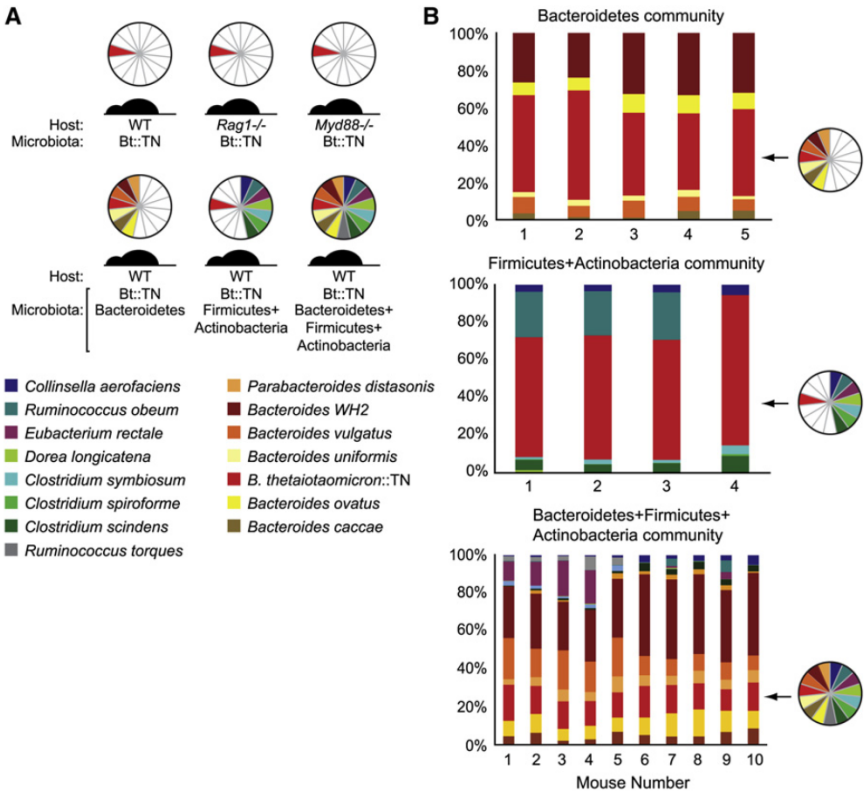


Figure 4. Identification of Environmental and Microbial Factors that Determine the Fitness Landscape for *B. thetaiotaomicron*

(A) The relative abundance of each mutant in the *B. thetaiotaomicron* population was evaluated in multiple host genotypes (wild-type C57Bl/6; *Rag1*$^{-/-}$; *Myd88*$^{-/-}$) and microbial contexts.

(B) qPCR assays of cecal microbial community composition in gnotobiotic mice at the time of sacrifice. The ~35,000-strain *B. thetaiotamicron* population is indicated with arrows.

were required for fitness in vivo, indicating that *CPS4* plays a unique role for *B. thetaiotaomicron* in the gut environment of monoassociated gnotobiotic mice fed a standard plant polysaccharide-rich chow diet. Moreover, our observation that transposon inactivation of genes in any single PUL did not confer a competitive disadvantage in vivo is consistent with *B. thetaiotaomicron*'s capacity for adaptive foraging of a broad range of glycans present in this diet (a total of 5137 distinct transposon mutants, involving 810 genes in all 88 PULs).

The Impact of Host Genotype and Community Structure on Selection In Vivo

We next asked whether a broad view of the mutant population as a whole could help address some basic questions in mammalian gut microbial ecology described in the Introduction: are the genetic determinants of fitness influenced by the bacterial community (microbiota) context; do inter-specific competition and intra-specific competition play distinct roles in shaping the selective pressures on a genome; is this selection primarily maintained by elements of the host immune system?

To explore these questions, we manipulated two features of the host habitat: (1) the immune system, by introducing the *B. thetaiotaomicron* mutant population into germ-free mice with genetically engineered defects in innate or adaptive immunity (*Myd88*$^{-/-}$ and *Rag1*$^{-/-}$, respectively); or (2) microbial composition, by including this mutant population as one component of three different types of defined communities, one consisting of six other sequenced human gut-associated Bacteroidetes, another composed of eight sequenced human gut-associated Firmicutes and Actinobacteria, and a third consisting of all 14 of these representatives of the human distal gut microbiota (n = 4–5 animals per treatment group per experiment; Figure 4A and Table S7). Animals were sacrificed 14 days after gavage and their cecal contents harvested. qPCR assays of cecal DNA, using species-specific primers, revealed that gavage with a given multispecies input community yielded consistent cecal "output" community compositional profiles (Figure 4B and Table S8).

Each of these treatments shifted the *B. thetaiotaomicron* population from its input distribution (Tables S9–S12). To search for functional trends in these shifts, we assigned the genes underrepresented in output populations to functional (COG) categories. The in vivo fitness determinants were significantly enriched in different predicted functions compared to essential genes, or to the genes required for maximal exponential growth in rich medium in vitro (Figure 5). For example, the predicted essential genes are most prominently enriched in COG categories J (translation, ribosome structure/biogenesis) and D (cell-cycle control and cell division), neither of which is enriched among the in vivo fitness determinants. Instead, the genes required in vivo are biased toward energy production/conversion (category C) and amino acid and nucleotide transport/metabolism (COG categories E and F, respectively). This represents an expansion beyond the single category (C) enriched after selection for maximal growth rate in vitro and is consistent across all in vivo treatment groups.

Closer examination of enriched COGs and carbohydrate-active enzyme (CAZy) families highlights the role of polysaccharide synthesis for competitive fitness in this environment; overrepresented functions include UDP-glucose-4-epimerases and glycosyltransferases (specifically, GT2 and GT4 families [Cantarel et al., 2009]) (Table S13). Together, these observations suggest that at the level of statistical enrichment of broad functional groups, the in vivo fitness requirements were distinct from those derived in vitro but that these enrichments were consistent across in vivo treatment conditions. Many of the annotated fitness determinants (e.g., *CPS4* and the *rnf*-like oxidoreductase) followed this pattern: dispensable in vitro but critical across all in vivo conditions tested.

This functional category enrichment analysis depends critically on genome annotation (~50% of genes were not assignable to COG categories), while not accounting for

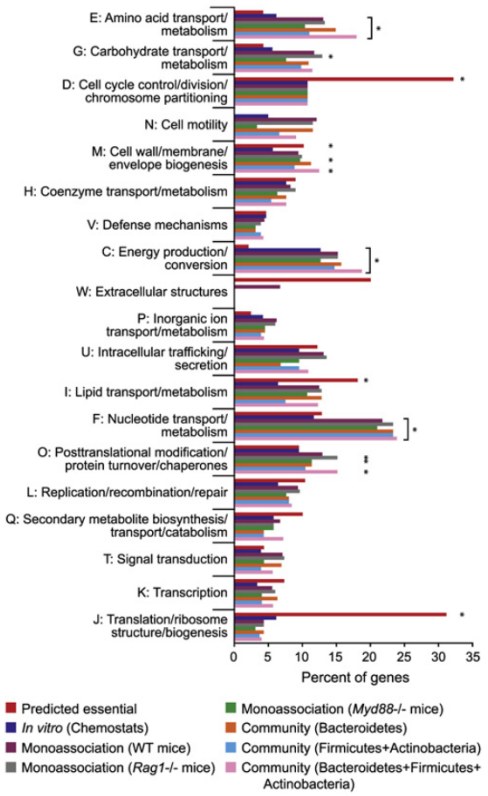


Figure 5. COG Category-Based Classification of Genes Critical for Fitness In Vitro and In Vivo

Percent representation was calculated as (number of genes in a COG category/number of genes underrepresented in output population). Significant enrichments in specific COG categories, assessed by comparing these percentages to a null expectation based on the size of the list of selected genes and the representation of a given category in the genome, are marked with asterisks (Benjamini-Hochberg corrected, $p < 0.05$).

differences in selection magnitude between treatments. To circumvent these limitations, we used unsupervised hierarchical clustering to evaluate the relationships between all in vitro and in vivo output populations (48 samples in total). In this analysis, samples are clustered based on the relative abundance of mutants in each gene independent of annotation. A dendrogram of the clustering results separates the in vitro populations from those shaped by in vivo selection (Figure 6A). The populations recovered from monoassociated wild-type, *Rag1*$^{-/-}$, and *Myd88*$^{-/-}$ mice formed shared branches on this tree of mutant population structures, suggesting that *B. thetaiotaomicron* may have multiple, redundant pathways for countering immune pressures, at least under the conditions of these experiments. In contrast, the transposon mutant populations recovered from mice that also harbored other Bacteroidetes species were fully separable from the population structures selected in *B. thetaiotaomicron* monoassociations or those resulting from the presence of a Firmicutes+Actinobacteria consortium. These patterns were recovered in >99% of bootstrapped dendrograms and also by unsupervised hierarchical clustering of samples based only on the genes that showed a significantly altered output:input ratio by z test (data not shown).

To further evaluate the robustness of the clustering algorithm, we conducted a principal coordinates analysis on these 48 transposon mutant populations (Figure 6B and Figure S10). Similar to the hierarchical clustering dendrograms, the first principal coordinate separates in vitro from in vivo mutant populations. The second coordinate separates these populations by microbial context: Bacteroidetes-containing communities shape a *B. thetaiotaomicron* mutant population that is distinct from that produced in monoassociations, or in the Firmicutes+Actinobacteria consortia, further indicating that although the same broad functional categories are enriched among *B. thetaiotaomicron* fitness determinants under a range of in vivo conditions, these mutant populations are additionally shaped by changes in microbial community composition.

We applied a random forest classifier (Breiman, 2001) to identify genes that were responsible for the observed separation of mutant populations in monoassociated mice from those mutant populations present in Bacteroidetes cocolonized mice. This machine-learning algorithm serves to estimate the importance of predictor variables (i.e., genes) for differentiating between classes (i.e., the monoassociation versus Bacteroidetes cocolonized groups, which were distinguished by unsupervised clustering and principal coordinates analysis). This approach identified a total of 220 genes as important for differentiating these groups (Table S14). Mutants in 144 of these predictor genes had lower output:input ratios in the monoassociations: in other words, these genes were more important when other Bacteroidetes were not present. Mutants in 76 genes had lower output:input ratios in mice cocolonized with other Bacteroidetes. These 76 genes, which provide a signature of functions under increased selection for *B. thetaiotaomicron* in the presence of other Bacteroidetes, are enriched for components of amino acid biosynthetic pathways, suggesting that although these functions are required in all in vivo conditions, other Bacteroidetes may outcompete *B. thetaiotaomicron* for exogenous amino acids.

The Functional Requirement for a Vitamin B_{12}-Regulated Locus Is Modulated by Community Composition

We identified 165 independent transposon insertions, mapping to five adjacent genes (BT1957-53), that conferred a drastic fitness disadvantage during monoassociation of germ-free mice yet had no impact on exponential growth in vitro. Moreover, their effect on fitness was influenced by community context: the Bacteroidetes-only community exacerbated the competitive defect, while the Firmicutes+Actinobacteria consortium fully nullified the requirement for these genes. Introducing all 14 of these species resulted in an intermediate phenotype (Figures 7A and 7B).

Examination of *B. thetaiotaomicron* transcriptional profiles (NCBI GEO archive) disclosed that expression of genes in this locus (spanning BT1957–BT1949) is strongly upregulated in vivo compared to growth in vitro under a variety of conditions. Moreover, in two closely matched experiments conducted in defined minimal medium that differed in five components, expression was modulated >10-fold (Table S15). This observation was validated by qRT-PCR assays of BT1954 and BT1956 (data not shown). Systematic addition of each variable

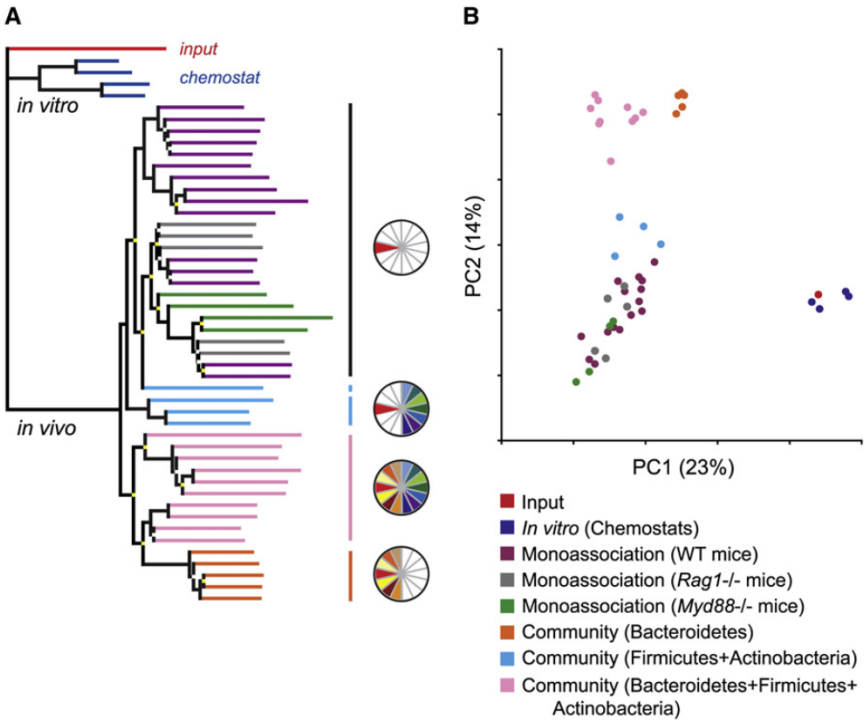


Figure 6. Clustering of *B. thetaiotaomicron* Mutant Population Structures after Manipulation of Host Environment or Microbial Context

(A) Unsupervised hierarchical clustering of *B. thetaiotaomicron* mutant population structures in vitro and in vivo. Branches are colored by treatment. Pie charts indicate the microbial context of the mutant population in samples from different branches of the tree. Bootstrap support is indicated by a square at each node: >50% (white), >90% (yellow), >99% (black/unmarked).

(B) Principal coordinates analysis based on the representation of transposon-disrupted genes.

component revealed that locus transcription is induced in response to reduced levels of vitamin B_{12} (Figure 7C). Moreover, the ABC transporter encoded by BT1952-50 shares homology with the BtuFCD B_{12} acquisition system of *S. typhimurium* LT2 (23%, 35%, and 25% identity, respectively).

Vitamin B_{12} is critically involved in normal mammalian physiology yet is synthesized exclusively by microbes (Krautler, 2005). Because the complete genome sequence of each member of the defined microbial communities used in our experiments was known, we were able to use BLAST to identify homologs to known B_{12} synthesis, transport, or utilization genes in the synthetic human gut microbiomes (Tables S16 and S17). As described for the related species *Porphorymonas gingivalis* (Roper et al., 2000), members of the Bacteroidetes community (including *B. thetaiotaomicron*) were missing some or all of the genes necessary for synthesis of B_{12} or its direct precursors but encoded predicted transporters and likely have an obligate B_{12} requirement for growth. In contrast, the Firmicute/Actinobacteria group contained several members that harbored complete B_{12} biosynthetic pathways. To test these predictions, we attempted to culture each species on defined medium in the presence and absence of vitamin B_{12}. *R. obeum*, a Firmicute that encodes a complete B_{12} biosynthetic pathway and is predicted to require the vitamin for methionine synthesis, grew robustly without B_{12} in the medium, while the Bacteroidetes were auxotrophic (Table S16). These observations suggest that the capacity for vitamin B_{12} biosynthesis is determined by community structure (Figure 7D) and that *B. thetaiotaomicron* responds to changes in community membership via the gene products encoded by BT1957-49. Consistent with this observation, the fitness defect of BT1957-3 mutants correlated well with variation in *R. obeum* levels between mice in vivo (R^2 of log-transformed values = 0.77; Figure 7E).

DISCUSSION

By simultaneously profiling the relative abundance of tens of thousands of *B. thetaiotaomicron* mutants across multiple conditions, we identified hundreds of genes that are critical for the fitness of this prominent human gut symbiont in vivo. A large number of these genes are not determinants of exponential growth in nutrient-replete medium in vitro.

This approach (INSeq) for functional genome-wide analysis of organisms for which a genome sequence (and possibly little else) is known is generally applicable and extends existing techniques in several important ways. First, a single transposon replaces the sets of individually barcoded variants needed for signature-tagged mutagenesis (Hensel et al., 1995). Second, high-throughput sequencing provides a general alternative to the species-specific DNA microarrays required for hybridization-based mutant profiling (Mazurkiewicz et al., 2006). Third, this sequencing-based strategy identifies the precise genomic location and provides a "digital" count-based abundance readout of individual insertions in both coding and noncoding regions. In this way, independent insertions with shared behavior serve to validate gene-level fitness effects. Finally, because *mariner* family transposon activity has been demonstrated in Bacteria, Archaea, and Eukarya, this method is generalizable. Further, the barcoded pooling strategy used to create a sequence-defined archived strain collection allows for retrieval of individual strains of interest for follow-up studies of the impact of individual gene disruptions on various microbial functions and adaptations. In this way, a forward genetic tool (a mutagenized cell population that can be screened for phenotypes en masse) can also serve as a platform for reverse genetics (a collection of isogenic, sequence-defined mutations in most of the coding potential of the target genome).

Surprisingly, mice lacking major branches of the immune system did not exhibit noticeably restructured *B. thetaiotaomicron* mutant populations compared to wild-type animals. It is possible that examination of microbial populations in closer contact with the host mucosa, or populations from host animals exposed to intentional immune stimulation, would aid the identification of genes differentially required for fitness in response to

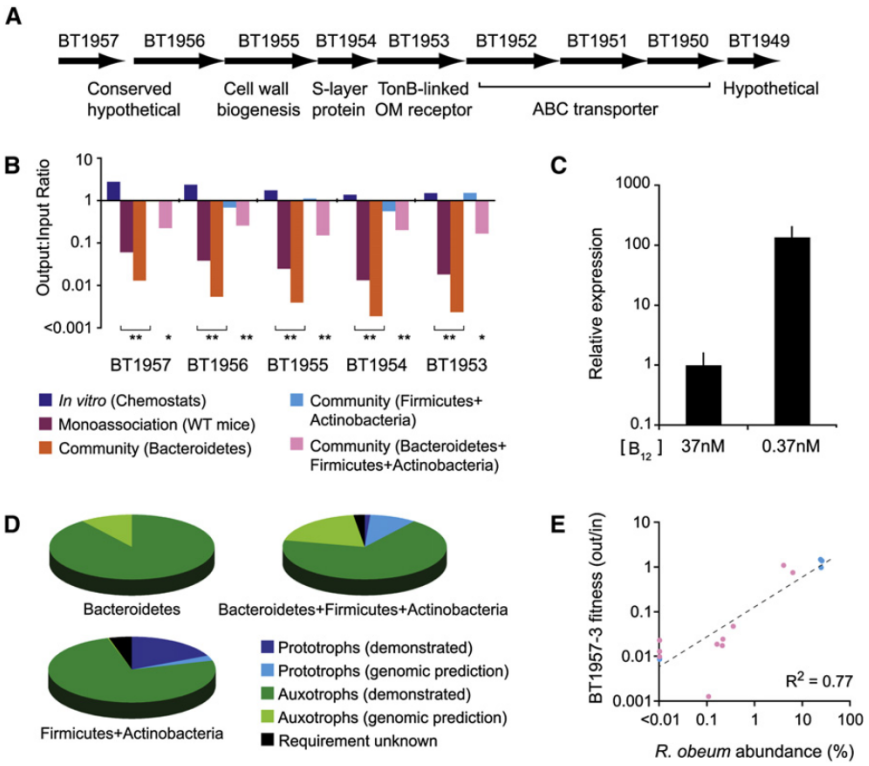


Figure 7. Community Context Modulates the Fitness Requirement for a Vitamin B_{12}-Regulated Locus in *B. thetaiotaomicron*

(A) Genetic organization and predicted annotation of the BT1957-49 locus.

(B) The relative abundance of transposon insertion mutants in input and output mutant populations is dependent on community context. Asterisks indicate FDR-corrected p value (q) for each cohort: *q < 0.05; **q < 0.01. Note that few insertions/reads were identified in the input community for BT1952-49 (Table S9); therefore, these genes are not included in the analysis shown in this panel.

(C) BT1956 gene expression is regulated by vitamin B_{12}. Error bars represent one standard deviation based on triplicate qRT-PCR experiments. Similar results were observed for BT1954 (data not shown). Interestingly, BT1956 insertion mutants do not exhibit growth defects in medium containing either 37 or 0.37 nM B_{12}, suggesting that multiple loci are involved in acquisition of this vitamin in vitro (data not shown). Consistent with this observation, other loci annotated as being potentially involved in B_{12} uptake were coordinately upregulated with BT1957-49 in genome-wide transcriptional profiling experiments.

(D) Pie charts showing that the defined microbial communities characterized in this study vary in their capacity for vitamin B_{12} biosynthesis. Color codes are as follows: dark blue, species with a predicted complete biosynthetic pathway (see Tables S16 and S17 for annotations) that are able to grow in defined medium lacking B_{12} ("demonstrated prototrophs"); light blue, organisms with a predicted complete biosynthetic pathway able to grow on rich medium but not on defined medium with or without B_{12} ("predicted prototrophs"); dark green, species without a complete pathway whose growth in defined medium requires B_{12} ("demonstrated auxotrophs"); light green, species without a complete biosynthetic pathway unable to grow on defined medium with or without B_{12} ("predicted auxotrophs"); black, species that do not grow on defined medium with or without B_{12} but that possess a B_{12}-independent methionine synthase (the presence of such an enzyme implies the absence of a B_{12} requirement). The relative proportions of prototrophs and auxotrophs shown represent the average in each community in vivo as determined by species-specific qPCR of cecal contents.

(E) The *B. thetaiotaomicron* fitness requirement for BT1957-3 correlates with levels of the B_{12}-prototrophic species *Ruminococcus obeum* in the microbial community. Each point represents an individual mouse containing a defined multispecies microbiota in addition to the *B. thetaiotaomicron* transposon mutant population. Color code is as follows: blue, Firmicutes+Actinobacteria community; pink, Bacteroidetes+Firmicutes+Actinobacteria. The relative abundance of *R. obeum* (determined by qPCR analysis of cecal contents at the time of sacrifice) is plotted against the average output:input ratio of *B. thetaiotaomicron* transposon mutants in genes BT1957-3 in each individual.

host immune surveillance. INSeq already can be applied to relatively small amounts of starting material: further decreases should help address the largely unexplored question of the relationship between the activities of the innate and/or adaptive immune systems and the biogeography of the microbiota.

Our study underscores how selection is shaped by microbial context. Because the INSeq strategy specifically targets transposon-adjacent chromosomal fragments, it is possible to monitor changes in the structure of a mutagenized population (in either wild-type or genetically manipulated gnotobiotic mice) even if this population constitutes a small fraction of a larger microbial community. For example, using unsupervised hierarchical clustering and principal coordinates analysis to evaluate the relationships between *B. thetaiotaomicron* mutant population structures in defined microbial communities in vivo, we observed that the presence of other Bacteroides was an important determinant of the selective pressures acting on the *B. thetaiotaomicron* genome. Additionally, we identified a mechanism by which *B. thetaiotaomicron* senses and responds to changes in gut microbial community structure: this species employs the products of BT1957-49 in response to variations in vitamin B_{12} levels that result from changes in community composition. These genes are dispensable in rich medium in vitro and become increasingly critical for fitness in vivo as the levels of B_{12} prototrophic species (such as *R. obeum)* are reduced. The mechanism by which *R. obeum* relieves this selective pressure, the role of the surface (S layer) proteins encoded by this locus, and the consequences of microbial B_{12} competition on host physiology await further study.

In principle, INSeq can be extended to an analysis of communities of defined species composition introduced into gnotobiotic mice together with, before, or after introduction of one or more transposon-mutagenized species. Subsequent INSeq-based time series studies of these deliberately constructed microbial populations offer the opportunity to address a wide range of unanswered questions about properties of our gut microbiota, ranging from the characteristics, determinants, and ecologic principles underlying its initial assembly, to the genomic correlates of niche partitioning among its members, to the genetic and metabolic factors that determine the persistence

and impact of various probiotic and enteropathogen species (and degree to which the functions required for persistence of gut mutualists overlap with those of pathogens [Hendrixson and DiRita, 2004; Lalioui et al., 2005; Liu et al., 2008; Shea et al., 2000]).

In summary, INSeq can be applied to a variety of phylotypes to identify factors that shape their adaptations to myriad environments, and further, to readily retrieve mutants in the genes encoding such factors. As such, INSeq, and the rapidly evolving capacity for massively parallel DNA sequencing that supports its application, should be a useful platform for microbial genetics, genomics, and ecology.

EXPERIMENTAL PROCEDURES

Bacterial Culture Conditions

Escherichia coli S-17 λ pir strains (Cowles et al., 2000) were grown at 37°C in LB medium supplemented with carbenicillin 50 μg mL^{-1} where indicated in the Supplemental Experimental Procedures. *B. thetaiotaomicron* VPI-5482 (ATCC 29148) was grown anaerobically at 37°C in liquid TYG medium (Holdeman et al., 1977) or on brain-heart-infusion (BHI; Becton Dickinson) agar supplemented with 10% horse blood (Colorado Serum Co.). Antibiotics (gentamicin 200 μg mL^{-1} and/or erythromycin 25 μg mL^{-1}) were added as indicated in the Supplemental Experimental Procedures. Other human gut-derived species were cultured in supplemented TYG (TYG_S: see the Supplemental Experimental Procedures).

Genetic Techniques

DNA purification, PCR, and restriction cloning were performed by using standard methods. Primer sequences are provided in Table S19. pSAM construction and mutagenesis protocols are described in the Supplemental Experimental Procedures.

Preparation and Sequencing of Transposon Population Libraries

A detailed protocol is provided in the Supplemental Experimental Procedures. Genomic DNA was purified, digested with MmeI, and separated by PAGE. Transposon-sized fragments were extracted from the gel and ligated to a double-stranded DNA adaptor bearing a 3′-NN overhang. PAGE-purified adaptor-ligated library molecules were PCR amplified for 18 cycles using a transposon-specific and an adaptor-specific primer. The 125 bp product was purified by PAGE and sequenced using an Illumina Genome Analyzer as described in the user's manual. Sequence images were converted into raw reads using Illumina software with default settings. Filtered, normalized, and mapped sequencing results from all samples are provided in Tables S9–S12 and S18.

Gnotobiotic Husbandry

All experiments using mice were performed using protocols approved by the animal studies committee of Washington University. Germ-free mice were maintained in gnotobiotic isolators and fed a standard autoclaved chow diet (B&K Universal, East Yorkshire, UK) ad libitum. Animals were sacrificed 14 days after gavage and cecal contents frozen immediately at −80°C.

ACCESSION NUMBERS

All of the sequencing results from this study are available from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) at accession number GSE17712.

SUPPLEMENTAL DATA

Supplemental Data include 10 figures, Supplemental Experimental Procedures, Supplemental Protocol, Supplemental MapSAM Software, Supplemental References, and 19 tables and spreadsheets and can be found with this article online at http://www.cell.com/cell-host-microbe/supplemental/S1931-3128(09)00281-9.

ACKNOWLEDGMENTS

We thank David O'Donnell, Maria Karlsson, Sabrina Wagoner, Nicole Koropatkin, Daniel Peterson, Pankaj Pal, Laura Langton, Jessica Hoisington-López, Xuhua Chen, Laura Kyro, and James Dover for assistance, plus Gary Stormo, Jay Shendure, Ryan Kennedy, and the Gordon laboratory for helpful suggestions. This work was supported by National Institutes of Health grants DK30292 and 1F32AI078628-01 (to A.G.).

Received: March 30, 2009
Revised: June 8, 2009
Accepted: August 13, 2009
Published: September 16, 2009

REFERENCES

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006 0008.

Breiman, L. (2001). Random Forests. Mach. Learn. *45*, 5–32.

Bryan, G., Garza, D., and Hartl, D. (1990). Insertion and excision of the transposable element mariner in Drosophila. Genetics *125*, 103–114.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. *37*, D233–D238.

Cowles, C.E., Nichols, N.N., and Harwood, C.S. (2000). BenR, a XylS homologue, regulates three different pathways of aromatic acid degradation in Pseudomonas putida. J. Bacteriol. *182*, 6339–6346.

Hendrixson, D.R., and DiRita, V.J. (2004). Identification of Campylobacter jejuni genes involved in commensal colonization of the chick gastrointestinal tract. Mol. Microbiol. *52*, 471–484.

Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E., and Holden, D.W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. Science *269*, 400–403.

Holdeman, L.V., Cato, E.D., and Moore, W.E.C. (1977). Anaerobe Laboratory Manual (Blacksburg, VA: Virginia Polytechnic Institute and State University Anaerobe Laboratory).

Jacobs, M.A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., et al. (2003). Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. USA *100*, 14339–14344.

Krautler, B. (2005). Vitamin B12: chemistry and biochemistry. Biochem. Soc. Trans. *33*, 806–810.

Lalioui, L., Pellegrini, E., Dramsi, S., Baptista, M., Bourgeois, N., Doucet-Populaire, F., Rusniok, C., Zouine, M., Glaser, P., Kunst, F., et al. (2005). The SrtA Sortase of Streptococcus agalactiae is required for cell wall anchoring of proteins containing the LPXTG motif, for adhesion to epithelial cells, and for colonization of the mouse intestine. Infect. Immun. *73*, 3342–3350.

Lamichhane, G., Zignol, M., Blades, N.J., Geiman, D.E., Dougherty, A., Grosset, J., Broman, K.W., and Bishai, W.R. (2003). A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. USA *100*, 7213–7218.

Lampe, D.J., Churchill, M.E., and Robertson, H.M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro. EMBO J. *15*, 5470–5479.

Lampe, D.J., Grant, T.E., and Robertson, H.M. (1998). Factors affecting transposition of the Himar1 mariner transposon in vitro. Genetics *149*, 179–187.

Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J., and Robertson, H.M. (1999). Hyperactive transposase mutants of the Himar1 mariner transposon. Proc. Natl. Acad. Sci. USA *96*, 11428–11433.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., et al. (2008). Evolution of mammals and their gut microbes. Science *320*, 1647–1651.

Liu, C.H., Lee, S.M., Vanlare, J.M., Kasper, D.L., and Mazmanian, S.K. (2008). Regulation of surface architecture by symbiotic bacteria mediates host colonization. Proc. Natl. Acad. Sci. USA *105*, 3951–3956.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe *4*, 447–457.

Mazurkiewicz, P., Tang, C.M., Boone, C., and Holden, D.W. (2006). Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. Nat. Rev. Genet. *7*, 929–939.

Peterson, D.A., McNulty, N.P., Guruge, J.L., and Gordon, J.I. (2007). IgA response to symbiotic bacteria as a mediator of gut homeostasis. Cell Host Microbe *2*, 328–339.

Rawls, J.F., Mahowald, M.A., Ley, R.E., and Gordon, J.I. (2006). Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. Cell *127*, 423–433.

Roper, J.M., Raux, E., Brindley, A.A., Schubert, H.L., Gharbia, S.E., Shah, H.N., and Warren, M.J. (2000). The enigma of cobalamin (Vitamin B12) biosynthesis in Porphyromonas gingivalis. Identification and characterization of a functional corrin pathway. J. Biol. Chem. *275*, 40316–40323.

Shea, J.E., Santangelo, J.D., and Feldman, R.G. (2000). Signature-tagged mutagenesis in the identification of virulence genes in pathogens. Curr. Opin. Microbiol. *3*, 451–458.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V., and Gordon, J.I. (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science *299*, 2074–2076.

Zocco, M.A., Ainora, M.E., Gasbarrini, G., and Gasbarrini, A. (2007). Bacteroides thetaiotaomicron in the gut: molecular aspects of their interaction. Dig. Liver Dis. *39*, 707–712.

Critical symbiont signals drive both local and systemic changes in diel and developmental host gene expression

Silvia Moriano-Gutierrez[a,b,c], Eric J. Koch[a,c], Hailey Bussan[c], Kymberleigh Romano[d,1], Mahdi Belcaid[a], Federico E. Rey[d], Edward G. Ruby[a,c], and Margaret J. McFall-Ngai[a,c,2]

[a]Pacific Biosciences Research Center, University of Hawai'i at Mānoa, Honolulu, HI 96822; [b]Department of Molecular Biosciences and Bioengineering, University of Hawai'i at Mānoa, Honolulu, HI 96822; [c]Department of Medical Microbiology and Immunology, University of Wisconsin–Madison, Madison, WI 53706; and [d]Department of Bacteriology, University of Wisconsin–Madison, Madison, WI 53706

Contributed by Margaret McFall-Ngai, January 25, 2019 (sent for review November 27, 2018; reviewed by Jeffrey I. Gordon and Liping Zhao)

The colonization of an animal's tissues by its microbial partners creates networks of communication across the host's body. We used the natural binary light-organ symbiosis between the squid *Euprymna scolopes* and its luminous bacterial partner, *Vibrio fischeri*, to define the impact of colonization on transcriptomic networks in the host. A night-active predator, *E. scolopes* coordinates the bioluminescence of its symbiont with visual cues from the environment to camouflage against moon and starlight. Like mammals, this symbiosis has a complex developmental program and a strong day/night rhythm. We determined how symbiont colonization impacted gene expression in the light organ itself, as well as in two anatomically remote organs: the eye and gill. While the overall transcriptional signature of light organ and gill were more alike, the impact of symbiosis was most pronounced and similar in light organ and eye, both in juvenile and adult animals. Furthermore, the presence of a symbiosis drove daily rhythms of transcription within all three organs. Finally, a single mutation in *V. fischeri*—specifically, deletion of the *lux* operon, which abrogates symbiont luminescence—reduced the symbiosis-dependent transcriptome of the light organ by two-thirds. In addition, while the gills responded similarly to light-organ colonization by either the wild-type or mutant, luminescence was required for all of the colonization-associated transcriptional responses in the juvenile eye. This study defines not only the impact of symbiont colonization on the coordination of animal transcriptomes, but also provides insight into how such changes might impact the behavior and ecology of the host.

squid–vibrio | symbiosis | development | daily rhythm | bioluminescence

Recent studies of animal and plant microbiomes have demonstrated that they can have far reaching effects, influencing both the internal and external environments of the host (1, 2). For example, the human microbiota impacts both tissues with which it directly interacts and more remote tissues of the body (3), as well as the built and natural environment in which the human host resides (4, 5). These complex microbial networks profoundly influence host development, from embryogenesis through senescence, while maintaining physiological homeostasis along this trajectory (2).

The best-studied nexus of these complex interactions is the mammalian gut microbiota (6), which affects not only the gut tissues themselves, but also the immune system (7), brain (8), liver (9), heart (10, 11), kidney (12), lung (13, 14), and eye (15–17). The microbiota also helps coordinate the activities of these tissues and organs: for example, the strong association of the gut microbiota with the control of host circadian rhythms (18, 19). Furthermore, axes of influence between the gut and other organs have revealed that dysbiosis of the microbiota is a critical driver of seemingly unrelated diseases (3).

Thus far, the mechanisms underlying these wide-ranging effects remain poorly studied. The integration of the gut microbiota into host biology is reflected in the transcriptomic regulation of genes in tissues both in direct contact with (20–22) and distant from (23, 24) the microbial assemblage. Available data suggest that the metabolomes of the blood, sweat, and urine carry products of the gut microbiota, such as short-chain fatty acids and microbe-associated molecular patterns (25, 26), to which these remote tissues respond. The complexity of the mammalian gut microbiota, however, renders it difficult to investigate the impact of a particular microbe on host biology under natural conditions, because the responses of adjacent and remote host tissues are the result of the cumulative effects of microbe–microbe and host–microbe interactions with hundreds to thousands of microbial phylotypes. In contrast, here we use the binary light-organ symbiosis between the Hawaiian bobtail squid, *Euprymna scolopes*, and its luminous bacterial partner, *Vibrio fischeri* (27, 28), to define the impact of a single symbiotic partner on the transcriptomic responses of host tissues, both those housing the symbiont population and those remote from the symbionts.

Significance

Biologists now recognize that animal microbiomes have strong impacts not only on the organs with which they associate, but also anatomically remote tissues; however, the precise triggers underlying these impacts remain unknown. Here, using the squid–vibrio light-organ association, which affords unparalleled resolution of a natural binary partnership, we report both near-field (light organ) and far-field (eye and gill) symbiont-driven effects on host gene expression. Colonization by the symbiont results in unique transcriptional signatures for each organ. Furthermore, distinct organ-specific patterns arise over the day/night cycle, and across the host's developmental trajectory. Most strikingly, the loss of a single genetic locus in the symbiont, that encoding bioluminescence, triggers a dominant and biologically relevant change of gene expression across the host's body.

Author contributions: S.M.-G., K.R., M.B., F.E.R., E.G.R., and M.J.M.-N. designed research; S.M.-G., E.J.K., H.B., K.R., and F.E.R. performed research; S.M.-G. and F.E.R. contributed new reagents/analytic tools; S.M.-G., E.J.K., K.R., M.B., F.E.R., E.G.R., and M.J.M.-N. analyzed data; and S.M.-G., K.R., M.B., F.E.R., E.G.R., and M.J.M.-N. wrote the paper.

Reviewers: J.I.G., Washington University School of Medicine in St. Louis; and L.Z., Rutgers University.

Conflict of interest statement: L.Z. and M.J.M.-N. are coauthors on a 2015 Comment article.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database, the National Center for Biotechnology Information Sequence Read Archive (accession nos. PRJNA473394, PRJNA498343, and PRJNA498345).

See Commentary on page 7617.

[1]Present address: Department of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44106.

[2]To whom correspondence should be addressed. Email: mcfallng@hawaii.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1819897116/-/DCSupplemental.

Published online March 4, 2019.

The squid host acquires its symbiont each generation from the surrounding environment and, similar to the mammalian gut microbiota, the bacteria reside extracellularly along the apical surfaces of epithelium-lined crypts (29). Also analogous to the mammalian gut microbiota, the squid–vibrio symbiosis undergoes significant development and maturation. Within hours following initial colonization of the juvenile animal, the symbionts trigger the regression of superficial ciliated fields of cells that promote light-organ inoculation (28). The symbionts also induce development of the crypt cells with which they directly associate throughout the animal's life, notably an increase in microvillar density and a swelling of the cells lining the crypts (Fig. 1*A*). A dark mutant derivative of *V. fischeri* (Δ*lux*), defective in light production, the principal "currency" of the symbiosis, is also defective in the induction of this latter hallmark event of early light-organ development (30–32).

Development of the light organ also involves the onset of diel cycles. Beginning during the first day of colonization and thereafter, ∼90% of the symbiont population is vented each day at dawn into the surrounding environment (33). Furthermore, in response to luminous (but not Δ*lux*) symbionts, the organ's cryptochrome-encoding clock gene, e*scry1*, begins a day/night cycling, and the host concomitantly imposes a cycling of the symbiont's luminescence levels, which peak in the hours of the early evening (1900–2000 hours), when the nocturnal squid host begins to forage. Then, after 3–4 wk of colonization, the symbiosis undergoes a final maturation, with the onset of a strong daily rhythm of metabolic processes (34, 35), not unlike the circadian rhythms of metabolism described in the mammalian gut symbioses (18, 36). Specifically, the animal becomes fully nocturnal, and the symbiont metabolism begins a day/night fluctuation between respiration and fermentation in response to a change in nutrients provided by the host.

Here we compared the transcriptomes of three highly vascularized organs of the squid host: the light organ itself and the eye and gill, manipulating both their symbiotic state and the genetics of the bacterial partner, in both juvenile and adult animals, and over the day/night cycle (Fig. 1*A*). The eye was chosen because, like the symbiotic organ, it is a light-sensing organ, and shows convergence in morphology, biochemistry (37), molecular biology (38, 39), and developmental pathways (40). As an immune organ, the gill, like the light organ, responds to bacterial colonization (41). Here, we present evidence that both light-organ colonization and luminescence influence gene regulation of not only symbiotic tissue, but also host organs remote from the symbionts.

Results

De Novo Transcriptome Assembly, Annotation, and Validation Provide the Resources for Analyses of Symbiosis Effects on Host Gene Expression. To determine the extent to which symbiotic colonization impacts host gene expression, we sequenced transcripts isolated from the squid light organ, eye, and gill (Fig. 1 *A* and *B*). Samples were collected for RNA-sequencing (RNA-seq) analysis from both juvenile (24-h

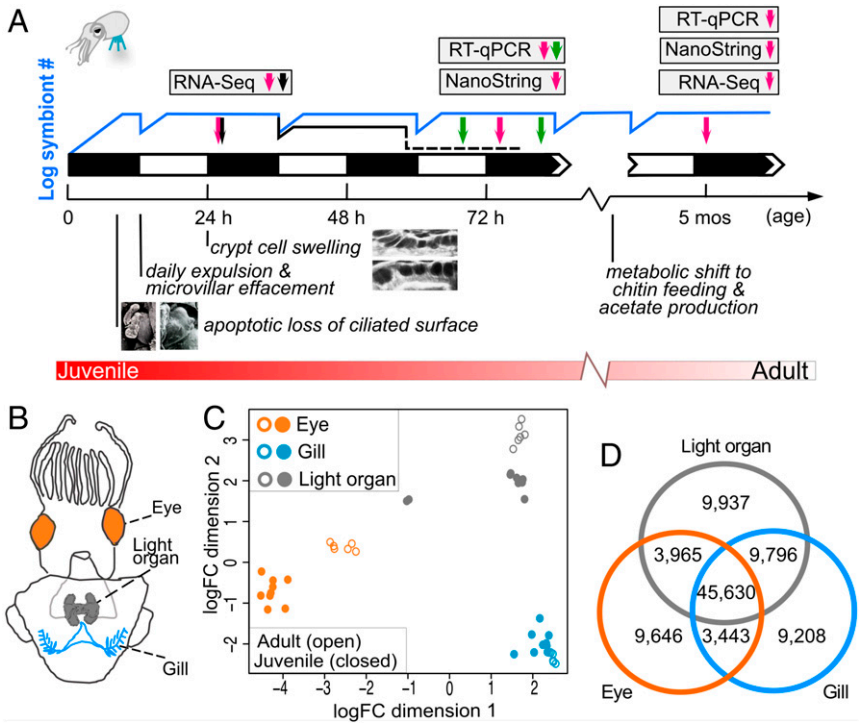


Fig. 1. Transcriptome profiling of *E. scolopes* organs. (*A*) Transcriptome sampling scheme during symbiotic development. On the night (black squares) that they hatch, juvenile squid become inoculated by *V. fischeri* cells, which proliferate, (blue line), filling the light-organ crypts and producing bioluminescence. Colonization triggers developmental events in the light organ's tissues, including apoptosis of the surface epithelium (*Left Inset*, APO; *Right Inset*, SYM), and edematous swelling of the crypt epithelial cells (*Upper Inset*, APO; *Lower Inset*, SYM). Each dawn, the nocturnally active host effaces the crypt-cell microvilli and expels most of its symbiont population, which grows back up by noon. A dark mutant (Δ*lux*) colonizes normally, but is unable to persist in the organ (dotted black line), and doesn't induce normal crypt cell swelling. After 1 mo, the host begins providing chitin to the symbionts, which ferment it to acetate. Transcriptomes of light organ, eye, and gill were constructed by RNA-seq or NanoString from organs sampled at 2000 hours (magenta arrows) in APO and SYM hosts of juvenile and adult animals. SYM-dark hosts, colonized by a dark-mutant strain, were also sampled (black arrow). For day/night comparisons, APO and SYM organs were sampled at 1600 and 0400 hours (green arrows) as well. Levels of transcripts of interest were confirmed by NanoString and qRT-PCR. (*B*) Schematic drawing of *E. scolopes* indicating tissue types collected from both juvenile and adult squid. (*C*) Multidimensional scaling plot of gene expression for the *E. scolopes* reference transcriptome. (*D*) Venn diagram of the number of shared and specific genes, expressed by tissue type. A gene is considered expressed when FPKM (fragments per kilobase million) > 0.5 in at least two samples per tissue.

posthatch) and mature adult (5-mo-old) animals, under two colonization states: symbiont-free (i.e., aposymbiotic, or APO) or colonized by the wild-type *V. fischeri* light-organ isolate ES114 (i.e., symbiotic, or SYM). An additional condition that was analyzed in juvenile animals was colonization by an isogenic, nonluminous (Δ*lux*) mutant of ES114 (i.e., SYM-dark) (42) (see *SI Appendix*, *SI Results* for details). The 2.2 billion paired-end reads obtained by Illumina sequencing, were de novo assembled to create a reference transcriptome (*SI Appendix*, Fig. S1 and Dataset S1). The large number of assembled transcripts is a common trait found in other de novo assembled transcriptomes of *E. scolopes* (43, 44). Cephalopods are known to expand certain gene families (45, 46). This feature, together with the high levels of heterozygosity and transcript editing (47) that are known to challenge assembly software (48), contributed to the high number of observed expressed transcripts. Transcriptomic profiles clustered by tissue type and, within each tissue type, by developmental stage, with a higher degree of variation among juvenile replicates (Fig. 1*C* and *SI Appendix*, Fig. S2). Irrespective of developmental stage, eye-derived samples showed the most divergent transcriptional profile. The light organ and gill displayed a more highly correlated expression pattern (Fig. 1*C* and *SI Appendix*, Fig. S2*A*), and shared more total expressed genes than either did with the eye (Fig. 1*D*), perhaps because they are both predominantly composed of epithelial tissue.

When considering the total number of genes that are expressed in each organ, depending on the host developmental stage and symbiotic state (*SI Appendix*, Fig. S3*A*), on average juvenile samples expressed 20% more detectable genes than their adult counterparts. Only 7,464 genes were expressed in all three organs, from both juveniles and adults, and in both the SYM and APO state (*SI Appendix*, Fig. S3*B*), suggesting that these genes encode core or "housekeeping" functions (for more details of transcriptomic patterns, see Dataset S2). In contrast, when we determined tissue-specific genes (i.e., those that were expressed at least eightfold higher in one organ relative to the other two), a total of 5,587 genes were identified (*SI Appendix*, Fig. S2*E* and Dataset S2). Not surprisingly, gene ontology (GO) terms enriched for the eye were related to visual perception or synaptic signaling, while for the gill were linked to gas exchange or pH regulation; similarly, the light organ was enriched in GO terms related to the expected activities of oxidative stress (49, 50) and chitin-associated processes (51) (*SI Appendix*, Fig. S2 *B–D* and Dataset S3). Furthermore, we validated the RNA-seq dataset by two methods, qRT-PCR, and the NanoString nCounter XT platform (*SI Appendix*, Fig. S4 and Dataset S4), which all had a high degree of congruity. These results provide strong evidence that the transcriptional patterns in response to colonization are robust, and clearly differentiate the three tissue types and their developmental states.

Adult Gene Expression in the Light Organ and Remote Tissues Responded Uniquely to Symbiont Colonization. To identify whether and how the light organ, eye, and gill responded to colonization of the light organ, we compared the gene-expression patterns of these three organs when sampled from adult APO and SYM squid (Fig. 1*A*). According to the values for the differentially expressed transcripts, the samples clustered by condition (APO or SYM) within each organ (Fig. 2*A*). Transcripts having expression levels that differed significantly between APO and SYM were identified as up- or down-regulated by symbiosis (Fig. 2*B*). Unsurprisingly, the light organ, which harbors the symbionts, had the strongest transcriptional response to its colonization, with a total of 206 genes significantly differentially regulated, which clustered into five distinct expression profiles (Fig. 2 and *SI Appendix*, Fig. S5). Although they are in anatomically remote organs (Fig. 1*B*), the transcriptomes of both eye and gill also responded to colonization of the light organ. Because of the greater similarity between the number of total transcripts in the light organ and gill (Fig. 1 *C* and *D* and *SI Appendix*, Fig. S2*A*), and because both of these organs respond to bacteria as part of their normal function, we anticipated that, compared with the eye, more symbiotically responsive genes would be detected in the gill, and they would overlap more significantly with the light organ. However, the eye had twice as many symbiotically regulated genes as the gill (84 vs. 42) (Fig. 2*B*, *SI Appendix*, Fig. S5, and Dataset S5). Furthermore, each organ had a distinctive transcriptional response to light-organ colonization: only one gene (annotated as angiotensin-converting enzyme or ACE) was up-regulated in two of the organs (eye and gill).

To further analyze not only the possible functions of these symbiosis-responsive genes, but also whether there were shared functions (if not genes) among the three organs, we conducted a GO-enrichment analysis for all of the differentially expressed genes. This analysis identified overrepresented terms in each organ, using the entire transcriptome as the background for the enrichment analysis. In APO animals, we found 40, 32, and 29 overrepresented functions in eye, gill, and light organ, respectively. In contrast, overrepresented functions in response to symbiosis were highest in the light organ, followed by gill and eye (Dataset S6). In addition, each of the three organs expressed genes within a unique set of top 10 enriched biological processes, in a symbiosis-dependent manner (Fig. 2*C*), a further indication of the distinct ways in which they react to the presence of bacteria in the light organ. For example, in the light organ itself, the three major responses to symbiosis, encompassing 9 of the 20 enriched biological functions (Fig. 2*C*), could be associated with: (*i*) vascularization and an increased oxygen demand driven by the symbiont's bioluminescence; (*ii*) tissue stress from the presence of the symbionts; and (*iii*) an easing of innate immune responses once the organ is colonized. All of these functions are consistent with previous studies (40, 52, 53). In contrast, in the eye, light-organ colonization resulted in an up-regulation of genes encoding structural proteins, and down-regulation of genes encoding elements of sensory perception and oxidative stress, while the gill exhibited an increased expression of genes encoding stress responses and transcriptional regulation.

Because a robust systemic response to colonization was observed that included functions associated with light-perception in the squid eye (Fig. 2*C*), an organ convergent in form and function with the mammalian eye (39), we asked whether and how eyes of another well-studied symbiosis model, the mouse, respond to host colonization; to our knowledge, the impact of the gut microbiota on the transcriptomic profile of the mouse eye has not been reported. We compared, by RNA sequencing, the expression profile of the eye of conventionalized mice (i.e., mice in which the gut microbiota was present) to that of germ-free mice. Adult stage mice and squid were compared to minimize any effects due to differences in their developmental rates. Applying the same level of stringency as used for the squid eye (i.e., an adjusted $P < 0.05$) only five genes were detected as differentially expressed in the mouse eye in response to conventionalization (Dataset S7). One predicted gene was down-regulated, and four genes were up-regulated, including lactotransferrin, which was previously reported as present in the transcriptome of mouse eye (54), IFN-activated gene 205, a mitochondrial tRNA, and a noncoding RNA of the RIKEN family. Although the evolutionarily convergent eyes of cephalopods and vertebrates share a large number of conserved genes with similar expression levels (55, 56), we detected no shared symbiosis-regulated genes in the eyes of these two organisms.

In summary, in the mature squid symbiosis: (*i*) functionally distinct and anatomically distant tissues are influenced by the presence of symbiotic bacteria; (*ii*) unlike the total expression profile for each of the three organs, the transcriptional responses to symbiosis, and their functional annotations, were specific and

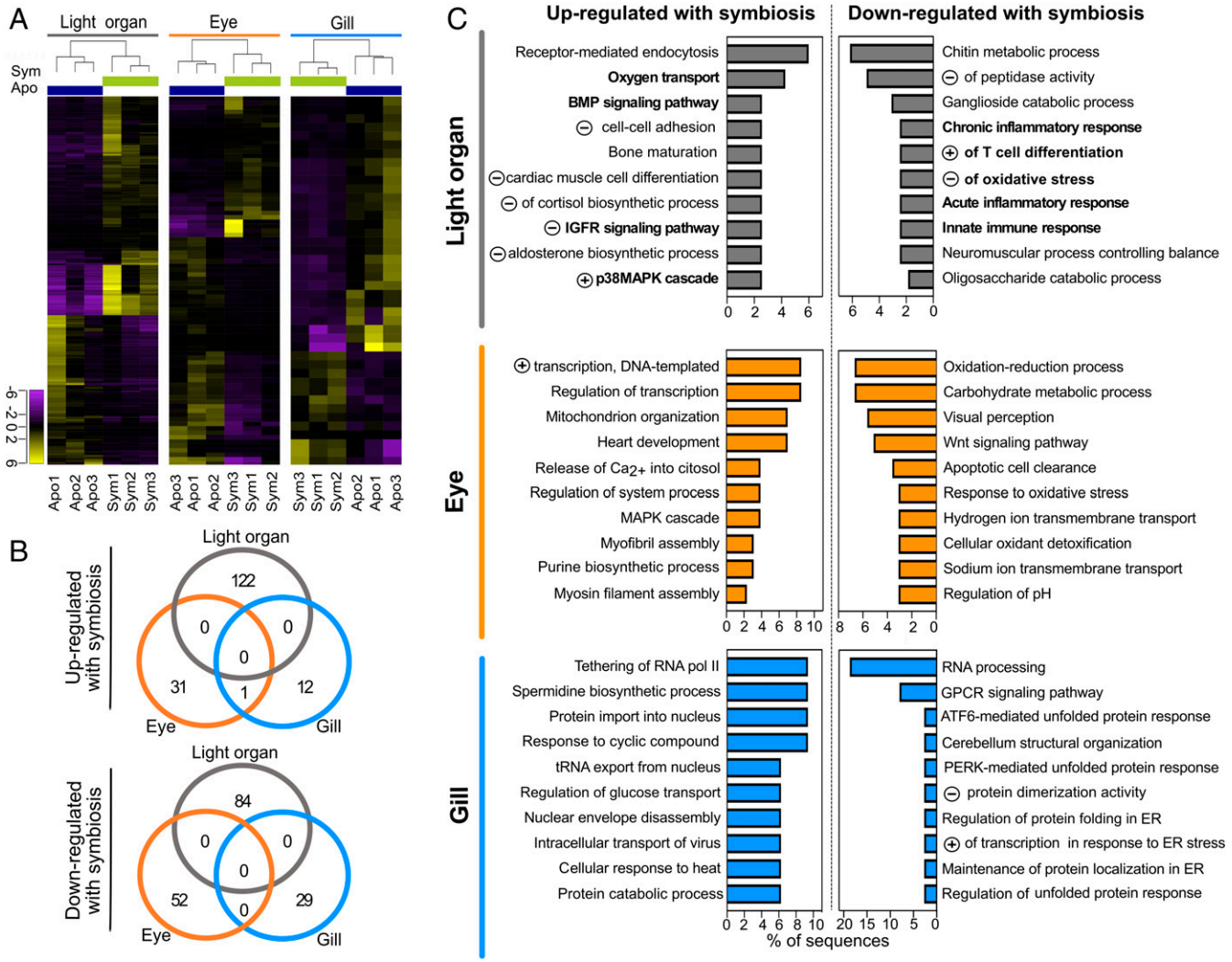


Fig. 2. Impact of light-organ symbiosis on gene expression in different adult organs. (*A*) Heat map of expression values, $\log_2$-transformed and median centered, for genes significantly differentially expressed (>twofold, $P_{adj} < 0.05$) in adult light organ, eye, and gill. Apo: aposymbiotic, in dark blue; Sym: symbiotic (colonized by wild-type *V. fischeri*) in green. (*B*) Venn diagrams indicating the numbers of significantly differentially expressed genes (>twofold, $P_{adj} < 0.05$) in response to symbiosis. (*C*) Functional annotation of symbiosis-responsive genes in remote tissues. The differentially expressed genes were enriched in functional categories based on GO annotation. The top 10 enriched biological processes are shown ordered by percentages of sequences with that function and by its significance level (Fisher's exact test, FDR < 0.05). "Negative (or positive) regulation of…" is abbreviated by a circled minus (or plus) symbol. Complete GO-term names and codes are in Dataset S6. Bold lettering indicates GO terms described in *Results*.

nonoverlapping (Figs. 1*D* and 2*B*); and (*iii*) although the evolutionarily convergent eyes of cephalopods and vertebrates share a large number of conserved genes with similar expression levels (55, 56), no shared gene regulation was detected within eyes of squid and mouse in response to microbial colonization of distant tissues.

Colonization of Juvenile Hosts Had a Rapid Impact on Gene Expression, Even in Remote Tissues. In the adult host, transcriptional responses to symbiosis are evident both locally and systemically (Figs. 2 and 3), but how quickly during symbiotic development does this outcome appear? The transcriptional response of the light organ has been reported to occur as early as 3 h following exposure to environmental *V. fischeri* (52). To determine the manner and timing of symbiosis-specific responses in other organs, more distant from the light-organ symbionts, we compared the RNA-seq gene-expression data of light organ, eye, and gill 24 h after the initiation of symbiosis, when the bacteria have fully colonized and are brightly luminous (28). At this point, we found that the light organ already exhibits a distinct transcriptional response; specifically, when comparing APO and SYM conditions, a total of 1,919 differentially regulated genes were detected, including 17% of the 206 genes characteristic of the adult SYM light-organ response (Fig. 2*B*). Analysis of this overlapping set of 36 genes revealed that ∼40% are associated with osmoregulatory and immune functions (Datasets S8 and S9). Subclusters 3 and 4 of light-organ differential gene expression comprise highly up-regulated genes in only two of the three analyzed SYM light organs (*SI Appendix*, Fig. S6), indicating a response whose onset is either variable or transitory. Interestingly, these two subclusters contained genes related to light perception, with significantly enriched functions, such as "structural components of the lens," "visual perception," or "phototransduction" (Dataset S9), perhaps reflecting the development of the light organ's capacity to perceive light (39).

As expected, a smaller number of symbiosis-responsive genes (44 in the eye and 184 in the gill) were detected (*SI Appendix*, Figs. S6 and S7), and there was essentially no overlap with the adult response in either of these organs. Nevertheless, a trend in

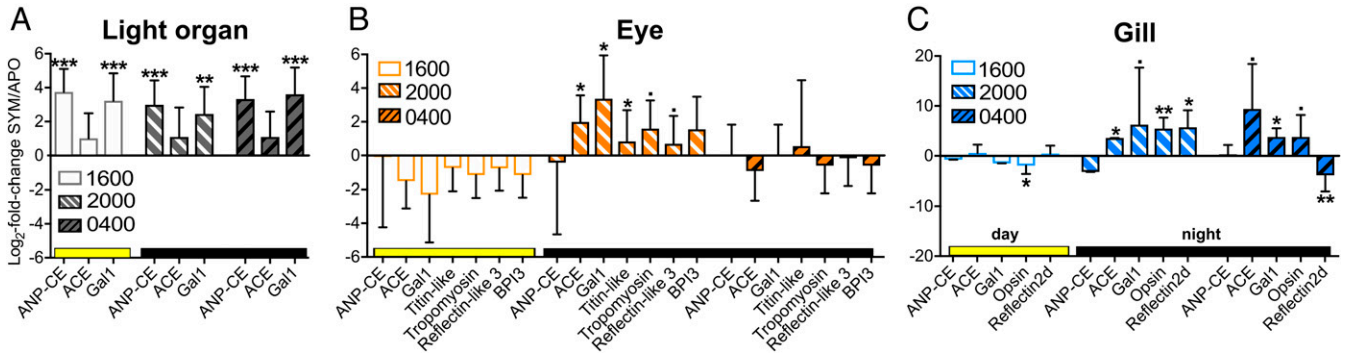


Fig. 3. Variation of symbiosis-responsive gene expression over the day/night cycle. Gene expression changes in SYM (compared with APO) light-organ (*A*), eye (*B*), and gill (*C*) tissue at different times of day: 0400 hours = 2 h before dawn; 1600 hours = 2 h before dusk; 2000 hours = 2 h after dusk. Juvenile squid were maintained for 3 d under a 12–12 light-dark schedule. Candidate genes where chosen for qRT-PCR based on expression changes observed at 72 h (*SI Appendix*, Fig. S9 and Table S1 and Dataset S4). BPI3, bactericidal/permeability-increasing protein 3. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; ·$P < 0.1$ (SYM vs. APO).

which eye (but not gill) genes clustered with colonization state was detected (*SI Appendix*, Fig. S7). Unlike the light organ, the responses of eye and gill at 24 h were highly variable between samples; thus, we hypothesized that many genes that were differentially regulated in these organs in adults may have not yet become apparent in 24-h juveniles (*SI Appendix*, Fig. S8*A*). Thus, we used the NanoString platform to determine whether the patterns of a selected set of 23 genes that were not significantly regulated at 24 h (Dataset S4), but were either trending toward induction at 24 h or would become induced in adult eye and gill had, by 72 h, became significantly differentially regulated by symbiosis. While such a temporal comparison can be made with juvenile eye or gill tissues, changes in the light organ transcriptome are confounded by this organ's substantial morphogenic transformation between 24 and 72 h (28).

Of the 13 selected adult eye-specific, symbiosis-responsive genes, 4 became clearly differentially regulated in juvenile eye tissue between 24 and 72 h postcolonization (e.g., *SI Appendix*, Figs. S8*A*′ and S9), indicating that during this period much of the transcriptional signature of the adult eye was still developing. For the gill, we chose two groups of genes: six that were significantly up-regulated in adults and four that were not, but were trending toward significance in 24-h juveniles. Of the first group, only one (ACE) had become differentially regulated by 72 h, while all four of the 24-h trending genes had. Thus, the data suggest that the gill has a more juvenile-specific response that is not retained in adults. In summary, the analysis of juvenile organs indicates that: (*i*) a robust transcriptional response to symbiosis appeared in the light organ within 24 h postcolonization, (*ii*) a smaller systemic response by eye and gill also became apparent, and (*iii*) by 24 h, the juvenile eye began to show an adult-like response, which became more significant at 72 h.

Expression of Some Symbiosis-Responsive Genes Was Regulated over the Day/Night Cycle. Because the light organ has a well-described daily rhythm of transcriptional regulation (35) that is reflected in crypt-cell ultrastructural remodeling, symbiont luminescence, and metabolic activity in both partners (34, 35), we asked whether the symbiosis-regulated gene expression detected in remote organs also changed over the day (Fig. 1*A*). Based on the NanoString data for 72-h juveniles at 2000 hours (*SI Appendix*, Fig. S9), we characterized symbiosis-responsive gene expression from juvenile organs by qRT-PCR at three times: 2 h before dusk (1600 hours, at ∼70-h postcolonization) and 2 h before dawn (0400 hours), both times when the host is quiescent and symbiont luminescence is reduced, compared with 2 h after dusk (2000 hours, at ∼74-h postcolonization) (Fig. 1*A*), when the host is active and the symbionts are brightly luminous (34).

Expression levels of three genes [atrial natriuretic peptide-converting enzyme (ANP-CE), ACE, and galaxin 1 (Gal1)] were determined across all of the organs. In the light organ, while ACE was not significantly regulated by symbiosis, ANP-CE and Gal1 remained up-regulated in SYM relative to APO at all times tested (Fig. 3*A*). In contrast, Gal1 and ACE were up-regulated by symbiosis in the eye only at 2000 hours although, in gill, ACE was up at both 2000 and 0400 hours (Fig. 3 *B* and *C*). Thus, ANP-CE is specifically regulated in the light organ, as is ACE regulated only in the eye and gill.

Expression of an additional four eye-specific and two gill-specific genes that were symbiosis-regulated at 2000 hours were likewise dependent on time of day. While there are trends of down-regulation of these genes in the eye at 1600 hours, no significant differences appeared beyond 2000 hours (Fig. 3*B*). Similarly, in gill, opsin is up-regulated at 2000 relative to 1600 hours, while reflectin 2 d becomes down-regulated (relative to APO) at 0400 hours (Fig. 3*C*). Cephalopods are noted for extraocular photoreceptors (57), but these structures are associated with the surface of the animal, and not with internal organs, such as the gills (58).

In summary, symbiosis-responsive genes that were regulated in one organ at one time of day can be differentially expressed in other organs at a different time, emphasizing the time- and context-dependency of the response. In addition, among the genes examined here, the symbiosis-dependent up-regulation of expression in gill, and especially in eye, was generally most prominent early in the evening (2000 hours), when the host is ecologically active. In contrast, in the light organ, no pattern was observed for these genes (Fig. 3*A*), although other genes show strong patterns of temporal regulation (35).

Symbiont Luminescence Was the Principal Driver of Transcriptomic Patterns in both the Light Organ and the Eye. Because symbiosis-induced up-regulation of gene expression occurred at night, coincident with high levels of symbiont luminescence, we asked whether light-emission itself is a factor driving gene expression. To this end, the gene-expression profiles of the juvenile light organ, eye, and gill were compared when the light organ was colonized by either a wild-type, light-producing strain (SYM) or a nonluminous Δ*lux* mutant derivative (SYM-dark). Because such dark mutants can only maintain normal levels of colonization for the first day postinoculation (32), we focused our analyses on 24 h after symbiosis had initiated.

At this time, under normal conditions of SYM colonization, 1,919 genes are regulated in the light organ compared with APO (Fig. 4*A*). Comparison of the SYM expression profile with that of the SYM-dark animals revealed that at 24 h the light organ has a strong transcriptional response, independent of light production.

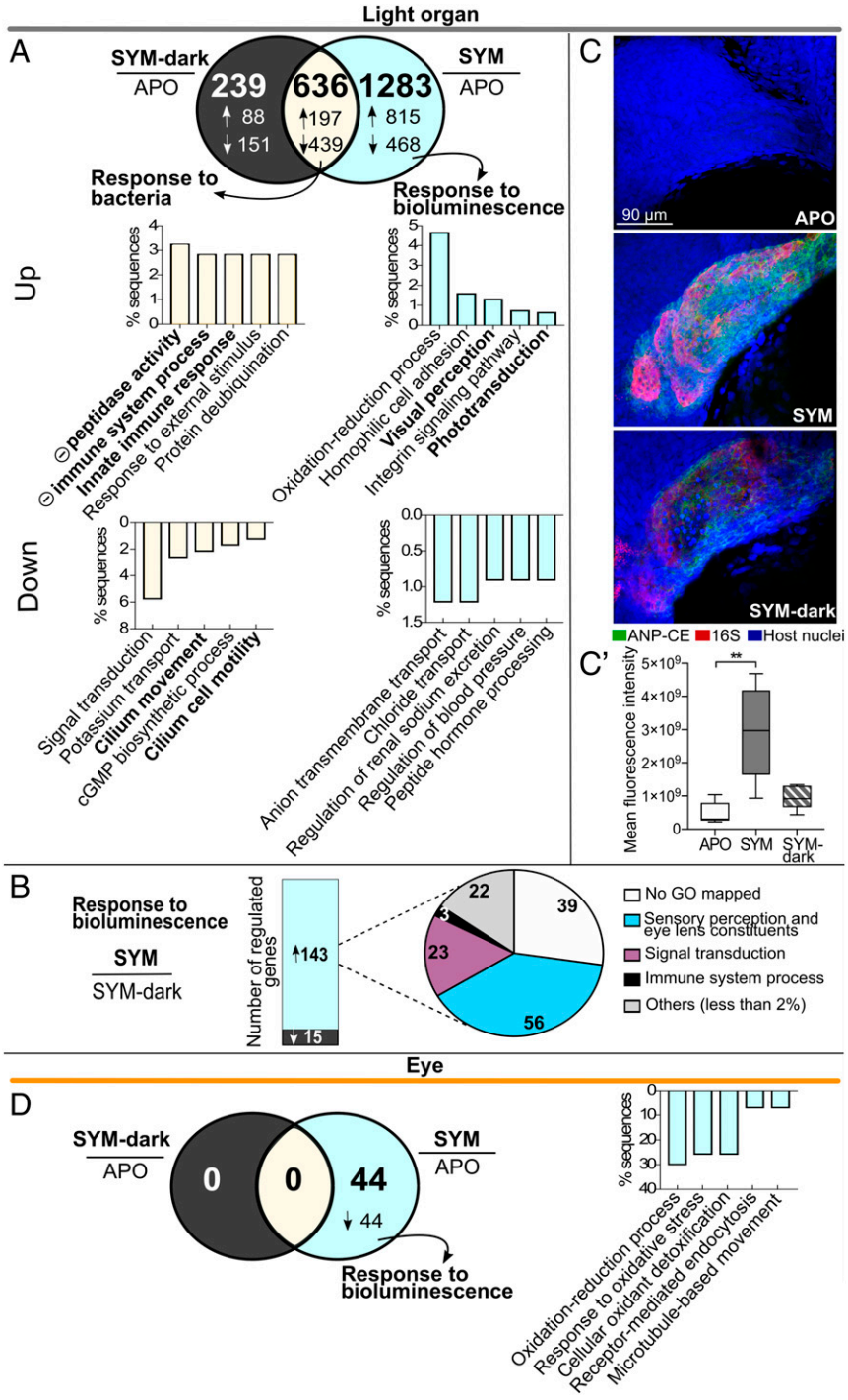


Fig. 4. Impact of symbiont bioluminescence on juvenile gene expression. (*A*) Venn diagram of numbers of differentially expressed genes in the light organ, 24 h after colonization by either SYM (wild-type) or SYM-dark (Δ*lux*) strains, compared with APO (>twofold, $P_{adj} < 0.05$). Arrows indicate either up (↑) or down (↓) regulation. Bar graphs: functional enrichment of genes significantly up-regulated and down-regulated with symbiosis. For each set, the top five biological process terms are represented (Fisher's exact test, FDR < 0.01). Notations as in Fig. 2. (*B*, *Left*) number of genes significantly up- or down-regulated in SYM compared with SYM-dark colonized light organs; (*Right*) proportion of annotated biological processes accounting for >2% of up-regulated genes. (*C*) Visualization of ANP-CE transcript in whole-mount light organs 24 h after colonization. Representative confocal images showing ANP-CE expression in crypt epithelium of APO, SYM, or SYM-dark colonized juvenile squid; merged mid-section of *z*-stack of crypt #1; ANP-CE (green), 16S RNA (symbionts, red), and host nuclei (TOPRO, blue) (*SI Appendix*, Fig. S10 and Movies S1–S3). (*C'*) Quantification of ANP-CE signal by fluorescence intensity from *z*-stacks of crypt #1 in five light organs. *P* values were calculated using Kruskal–Wallis test and Dunn's multiple comparison test. Error bar: SD (**$P < 0.01$). (*D*) Venn diagram of differentially expressed genes in the eye, 24 h after colonization by either SYM (wild-type) or SYM-dark (Δ*lux*) strains, compared with APO (>twofold, $P_{adj} < 0.05$). Arrow indicates down (↓) regulation. Bar graph: notations as in *A* (Datasets S8 and S9).

Specifically, a total of 636 genes were regulated by both strains, over two-thirds of which were down-regulated. The functional annotation of >25% of these down-regulated genes was dominated by GO categories associated with maintenance of ciliary structure and function (Fig. 4*A*), which is not surprising because both SYM and SYM-dark bacteria induce the cell death and loss of the ciliated surface that mediates initial colonization (28). Also enriched in this shared set were genes up-regulated in immune response and stress (Fig. 4*A*), which is also not unexpected, as the morphogenesis of the ciliated surface is driven largely by symbiont microbe-associated molecular patterns.

However, it is most striking that over two-thirds of the 1,919 genes regulated by colonization required that the symbionts be luminescent, underscoring how critical a role *V. fischeri* bioluminescence plays in shaping the symbiosis. In direct contrast to the luminescence-independent response, nearly 70% of the genes of the luminescence-specific response were up-regulated. Notably, the 39 GO categories of up-regulated genes included "visual perception," "phototransduction," "photoreceptor activity," and "structural constituent of eye lens," all involved with light perception or modulation, as well as homophilic cell adhesion and oxidative-reduction processes (Fig. 4*A* and Dataset S9). The light organs that were colonized by the dark mutant not only failed to regulate these luminescence-specific genes, but also had an expression signature of their own. The dark mutant

regulated 875 genes, only 46% the number regulated by the wild-type strain (1,919), with just over one-quarter of their regulated genes specific to the SYM-dark colonization, compared with the two-thirds of genes specific to the luminous SYM colonization. Furthermore, unlike the SYM animals, the SYM-dark condition resulted principally in down-regulated genes, with no significant functional enrichment in any GO category (Fig. 4*A*). Finally, when gene regulation was compared directly between SYM and SYM-dark animals, 143 annotated genes were up-regulated in SYM, and only 15 down-regulated (Fig. 4*B*). Thirty-nine percent of these up-regulated genes were associated with sensory perception of light stimulus or modulation of light (e.g., lens proteins) (Dataset S9).

A particularly interesting difference between colonization conditions was the ∼sevenfold up-regulation in SYM compared with SYM-dark of ANP-CE, which regulates cell volume and inflammation in a variety of systems (59). One of the key developmental features of the light organ is the SYM-induced swelling of the crypt cells with which the bacteria directly associate (Fig. 1*A*); however, the dark mutant is defective in inducing this phenotype (32). If ANP-CE were involved in such a cell-swelling and inflammation phenotype, we would predict that the transcript for this protein would specifically localize to the crypt epithelium in the SYM host, and be at a higher abundance than in SYM-dark–colonized animals. Using hybridization chain reaction-FISH, we compared the localization of the ANP-CE transcript in light organs at 24-h postinoculation. Abundant transcripts localized specifically to the cytoplasm of the crypt cells in SYM-colonized animals (*SI Appendix* and Movies S1–S3). In contrast, only low levels were detected in either APO or SYM-dark animals, with no significant difference between these conditions (Fig. 4 *C* and *C′*).

Unlike the light organ, no difference in colonization-dependent gene expression was detected between gill tissue of SYM and SYM-dark juveniles at 24 h, perhaps due to this organ's high variability in development at this time point. However, the eye showed a uniform, down-regulation in the expression of all 44 of the genes that responded in any way to colonization by the luminous symbiont (Fig. 4*D*); in contrast, no significant change in expression of any of those genes was detected when the symbiont was the dark mutant. These data suggest that, like the light organ, the eye's principal reaction to symbiotic colonization is mediated by the presence of light production. Interestingly, eye genes down-regulated by symbiont colonization were enriched in biological processes related to oxidation state or tissue reorganization (Fig. 4*D* and Dataset S9). Only 4 of the 44 genes down-regulated in the eye were shared with the light-organ's response: specifically, ANP-CE, MAM/LDL-receptor class A domain-containing 2-like, dynein heavy-chain axonemal, as well as the hypothetical protein KGM_03810, which is also regulated in the adult light organ. However, unlike with the juvenile light organ, the first three of these eye genes are up- (not down-) regulated in response to symbiosis. In summary, symbiosis-dependent gene expression in both the light organ and eye was more dependent on the existence of bacterium's luminescence than on the presence of the bacteria themselves.

Discussion

The data presented here demonstrate that, beginning within hours of the onset of the *E. scolopes* light-organ symbiosis, localized colonization of tissues by the specific symbiont *V. fischeri* creates a network of communication across the host's body that reprograms transcription system-wide. The coordination of this network persists throughout the developmental trajectory of the association, beginning with early symbiosis-induced changes in light-organ form and function, and continuing well into the maturation of the partnership. Furthermore, the network reprograms remote organs to respond to the daily rhythms set by the *V. fischeri* population within the light organ. Finally, genetic manipulation of the symbiosis revealed that bioluminescence, the principal currency of the symbiont, while having no effect on the gill response, is not only the major symbiosis-dependent driver of transcriptional regulation in the light organ, but also the only driver in the eye.

Transcriptomes of the Host's Organs Reflect Their Biological Functions Through Development. The transcriptomes examined in this study, which segregated both by organ and by life stage, reflect known functions of the light organ, eye, and gill: that is, control of symbiont luminescence, vision, and respiration, respectively (*SI Appendix*, Fig. S3 and Dataset S3). The light organ, eye, and gill transcriptomes clustered separately between juveniles and adults (Fig. 1*C* and *SI Appendix*, Fig. S2*A*), a finding that may reflect the dramatic change in the animal's ecology upon its maturation. Briefly, from hatching to ∼4-wk postcolonization, the juvenile's behavior is not controlled by a daily rhythm, being either active or quiescent at all times of day and night; however, by about 1 mo, the animal has assumed a profound diel rhythm of burying in the substrate during the day and coming out to forage at night, a behavior that will persist throughout its ∼1-y life (60). This change in lifestyle coincides with a dramatic shift in the daily cycling of host and symbiont metabolism (34), a shift that promotes a brighter luminescence of the bacteria in the evening, when the squid is active and using the light emission of the symbionts to camouflage itself by "counterilluminating" (28, 61). We hypothesize that the eye may be responding transcriptionally not only to its commitment to a diel rhythm of environmental light, but also to the requirement that the eye coordinate its function with the light organ's emission. Specifically, during counterillumination, the light organ modulates its luminescence in response to the intensity of down-welling moonlight and starlight, which is monitored by the eye, through as yet undefined mechanisms. With such developmental changes in day/night behavior, it is not surprising that the transcriptomes of both the light organ and eye adopt new patterns as these organs mature. Similarly, the respiratory and immune functions of gill tissue, like its transcriptome (Fig. 1*C*), change between juvenile and adults as the animal begins to bury in the substrate each day by 4 wk of age.

Symbiosis-Induced Changes in Squid–Host Gene Expression Show Similarities with Those Reported for the Mammalian Microbiota. As with the global patterns of squid gene expression (Fig. 1*C*), both the light organ and remote tissues reacted robustly to colonization by *V. fischeri*, and several features of this life-stage and organ-specific transcriptomic response appear to be evolutionarily conserved between the squid light-organ and mammalian-gut symbioses. For example, a comparison of intestinal epithelial cell transcriptomes from germ-free and conventionalized mice found that the response of intestinal epithelial cells to the presence of the microbiota was only a fraction of the genes expressed in these cells and, as in squid, little overlap in this response occurs between juvenile and adult mice (62). In the squid, such largely stage-specific patterns of symbiosis reprogramming apply not only to the colonized tissue (i.e., the light organ), but also to the anatomically remote eye and gill. As yet, we know little about how remote tissues receive information about the colonization state or activity of symbiotic organs, but two modes are possible. The first is a chemical signal, such as a bacterial metabolite, delivered through the circulation (63, 64); one such metabolite, acetate, is generated in the light organ as byproduct of symbiont metabolism (34). For example, the presence of *V. fischeri* in the light organ has a systemic effect on hemocyte signaling (65). The second mode is neural: cephalopods, in particular, produce both targeted and systemic responses via their nervous system, similar to mammals, where the vagus nerve conveys information about the gut microbiota to the brain (66). However they are delivered, the transcriptional

changes appearing during postnatal development in the mouse and squid organs reveal only part of the impact of symbiosis. In fact, many of these effects are deployed during embryogenesis, when symbionts are not yet present: that is, both in the squid–vibrio system and in mammals, the developing host creates specific target-tissue conditions that are poised to respond to the eventual arrival of their symbiotic partners (28). Conversely, in a synoptic study comparing digestive system transcriptomes of four regions of the mouse intestine and liver (67), as with the squid, very little overlap in the transcriptional response to symbiosis occurred across the body.

To our knowledge, an integrated comparison both across tissues and through development has not yet been addressed in the mouse; thus, whether the trajectory of the robust mammalian transcriptional response to symbiont colonization over early development varies among organs, as it did in the squid (i.e., first the light organ, then eye, then gill) (Datasets S4 and S8) remains to be determined. However, taken together, these similarities in life-stage and organ/tissue-specific responses to symbiosis in the two distantly related animal taxa suggest that, rather than having a generalized response to bacteria or their products, as they develop each tissue or organ near to or distant from sites of colonization integrates the partnership into its specific function. For example, in the squid association, we observed an up-regulation of genes involved in vascular development of tissues (Fig. 2*C* and Dataset S6), which has also been reported for the nutrition-based gut symbioses of mammals (68). The finding that the squid system, wherein bioluminescence and not nutrition is the principal benefit to the host, suggests, not surprisingly, that increased vascularization is important for other aspects of this symbiosis, such as facilitating the support of the bacterial population and their dialogue with host tissues (34, 53).

Symbiotic Regulation of Genes Encoding Specific Functions Occurs on a Diel Cycle. In addition to sharing life-stage and tissue-specific responses to symbiosis, in both mammalian (69) and squid hosts, there exist diel transcriptomic rhythms in colonized and remote tissues that are influenced by their bacterial partners. For example, although the overall transcriptomes of the three squid organs examined here differed greatly (Fig. 1*C*), the juvenile light organ shared with eye and gill a symbiosis-dependent diel regulation of two genes: ACE and Gal1. ACE occurs widely among animals, controlling blood pressure and electrolyte balance, although the conserved function of this protein family is immune modulation in both vertebrates (70, 71) and invertebrates (59, 72). Immunity is also likely to be its ancestral role because ACE occurs even in taxa without a closed circulatory system; however, in the squid, whose circulation is closed, this protein may serve both an immune and a vascular function. In contrast, galaxins are invertebrate-specific proteins, first identified in corals; in *E. scolopes*, Gal1 is an antimicrobial peptide present in the light organ (73). Although the genes encoding ACE and Gal1 were both regulated (either up or down) by symbiosis, each had a different daily rhythm of expression depending on the host organ (Fig. 3).

What biological purposes might underlie the diel regulation of these genes? At all times analyzed—that is, 0400, 1600, and 2000 hours—the light organs were fully colonized (74); however, at 1600 and 0400 hours, the per cell luminescence of the symbionts is relatively low, compared with its maximum at 2000 hours (75). The light organ expressed ACE and Gal1 constitutively throughout the day, whereas ACE and Gal1 expression was up-regulated in the eye only when luminescence was highest (2000 hours). In addition, the light organ had symbiosis-induced expression of ANP-CE, which transforms a propeptide to the edema-related peptide ANP (76). Because ANP-CE and ACE often offset one another's functions in immunity and vascular homeostasis (77), these findings suggests that an ACE-driven modulation of ANP-CE activity in the light organ releases ANP into the host blood stream, modulating the biochemistry of the eye and gill during the day/night cycle. Taken together, the transcriptional data are consistent with an organ-specific expression of an increased immune potential at particular times of day.

Light Organ and Eye Are More Similar in Their Responses to Symbiosis. The repertoire of gene expression characteristic of light organ and gill tissues overlapped more in gene identity and number (Fig. 1*D*), consistent with both their similar relationship to the environment (i.e., both are bathed with bacteria-rich seawater during ventilation) and their shared immune function. In contrast, the interior portions of the eye examined here are protected from any direct exposure to environmental microbes. However, compared with the gill, the differential gene expression in symbiosis responses of the squid's eye and light organ were more similar in magnitude (Fig. 2*B*), reinforcing the hypothesis proposed above, that a coordination between these organs facilitates the host's counterillumination behavior within the ambient light field (61).

The squid eye also had a stronger relative response to light-organ colonization (84 genes) (Fig. 2*B*) than the mouse eye did to colonization of the gut by the microbiota (five genes) (Dataset S7). This difference in the scale of transcriptional response was unexpected because, while the mice were an inbred strain, the squid were genetically diverse, reared from wild-caught parents, and such genetic variation should lead to an underestimation of significant transcriptomic differences. However, the relatively weak reprogramming of the mouse by the presence of its gut microbiota may reflect the vertebrate eye's status as a site of immune privilege (17). Nevertheless, colonization has been reported to influence the mouse eye's lipid content (16), and metabolomics studies have revealed a gut–retina axis that correlates with a proclivity toward development of macular degeneration (15, 78). In short, the difference between squid and mouse responses may be either due to differences in how the immune system interacts with the eye, or because of a greater need for light-organ/eye coordination in bioluminescent symbiotic associations. The latter hypothesis could be tested by a study of fishes with light-organ symbioses, which have both the vertebrate immune privilege of the eye, and the need to modulate luminescence by the symbiotic organ (79, 80).

Symbiont Luminescence Has a Disproportionately Large Transcriptomic Effect. A striking feature of symbiont-induced gene expression was discovered during a comparison of the organ transcriptomes of squid colonized with either the wild-type or the Δ*lux* mutant symbiont. The results highlight the remarkable dominance of luminescence in reprogramming gene regulation in not only the symbiont-containing light organ (Fig. 4 *A* and *B*), but also the anatomically remote eye (Fig. 4*D*). In many other associations, studies of such systemic consequences of losing the symbiont's principal activity have been clouded by the resultant physiological effects on the host. For example, in the rhizobium-legume symbiosis, bacterial mutants defective in nitrogen fixation similarly have a differential affect on transcription in both nodule tissue (81) and other plant organs (82); however, it was difficult to separate functions within the nodule from their general nutritional role. Similarly, animal–microbe interactions within gut tracts and bacteriomes also revolve around the symbiont's provision of an important nutrition function (83). That is, unlike bioluminescence, these functions directly impact the host's general physiology and health. Because the light-organ symbiosis plays an ecological role for the squid (i.e., antipredation), under laboratory conditions, the physiology of the host is not negatively affected by carriage of a dark mutant (32). As such, this system serves as a paradigm for studies of the reprogramming of remote-tissue by other microbiota assemblages with nonnutritional functions [e.g., an antibiotic

or defensive toxin (84, 85)], like those of the skin and urogenital tract of humans (86).

ANP-CE is one of the most abundant and highly regulated genes in the squid transcriptomes; specifically, it is overexpressed in SYM relative to APO tissues (Fig. 4*C*), it's level fluctuates on a day/night cycle in eye and gill (Fig. 3 *B* and *C*), and in light organ and eye its expression is induced by the symbiont's luminescence (Dataset S8). Because ANP-CE plays a role in inducing cell edema (76), and the epithelia lining the symbiont-containing crypts swell when colonized by wild-type, but not dark-mutant, *V. fischeri* (32, 87), we predicted that ANP-CE expression would be reduced in the epithelium when the symbionts are dark mutants (Fig. 4*C*). This reduction suggested a direct link between symbiont function and host-tissue response, and may indicate a mechanism by which dark mutants are sanctioned (32, 88), possibly by withholding the nutrients typically delivered by the crypt's edematous epithelium.

The data presented here demonstrate that different tissues can have different transcriptional drivers in response to symbiosis, and future work will address the array of symbiont features that trigger the systemic response to light-organ colonization. Here we demonstrate that, among several effects on remote tissues, the most remarkable was the role of the symbiont's product (luminescence), rather than the presence of the symbiont itself, as the sole driver reprogramming the eye's transcriptome (Fig. 4*D*). As yet, it is unclear how the eye recognizes the presence of a luminous symbiont: perhaps a signal is delivered indirectly through a humoral or neural signal from the light organ (39). In any case, determining the mechanisms underlying this response presents a rich horizon for discovering shared principles of microbe–organ communication. The transcriptomic responses described here not only document symbiosis-induced molecular networks across the host, but also reveal how these networks may influence the behavior and ecology of the host: for example, mediating the counterillumination antipredatory strategy.

In conclusion, we determined three types of organ-specific transcriptional responses to symbiont colonization over the trajectory of development and over the day/night cycle: (*i*) a strong reaction to both the microbe and its primary product, luminescence (light organ); (*ii*) a response to the presence of the symbionts, independent of their luminescence (gill); and (*iii*) a response triggered solely by the luminescence product (eye). Determining the presence and mechanistic basis of the interorgan network that connects symbiotic and remote tissues, enabling a coordinated response, is a critical area of exploration, and will eventually reveal the degree to which symbioses can influence host health and homeostasis throughout life.

Materials and Methods

Sample Collection. Animals were collected 24 and 72 h after hatching (juveniles), or ~5 mo (adults) (Fig. 1*A*), and anesthetized in seawater containing 2% ethanol; juveniles were stored whole in RNAlater (Sigma-Aldrich) as previously described (52), while adult tissues were dissected before storage. With the exception of day/night cycle studies, all samples were collected at 2000 hours, 2 h after dusk.

RNA-Seq Assembly and Analysis. A total of 2.2 billion paired-end reads were de novo assembled using the Trinity-v2.4.0 RNA-Seq assembler (89) (Dataset S1), and annotated by BLASTx against the National Center for Biotechnology Information nonredundant protein database. For functional annotation of the reference transcriptome, GO mapping of the transcripts was performed with Blast2GO software (90). To estimate the relative expression value for transcripts, RSEM software (91) was used in combination with the R package edgeR (92) to identify the significantly differentially expressed transcripts. Statistical enrichment of GO terms for differentially expressed genes was performed in Blast2Go using the Fisher's exact test with a false-discovery rate (FDR) < 0.01. In addition, gene set-weighted enrichment analysis with 500 permutations and FDR < 0.1 was performed on the differentially expressed transcripts (*SI Appendix* and Dataset S10).

Transcript Quantification by qRT-PCR or NanoString nCounter Analysis. Changes in host gene expression were measured by qRT-PCR using LightCycler 480 SYBR Green I Master Mix (Roche). Ribosomal protein 19L, serine hydroxymethyl transferase, and heat-shock protein 90 were used to normalize the transcript level (*SI Appendix*, Table S1). The nCounter Custom CodeSet Kit (NanoString Technologies) was used to detect changes in gene expression (Dataset S4). Assay and spike-in controls were used for normalization based on identical amounts of input RNA. Analysis was performed with nSolverAnalysis software v3.0.

See *SI Appendix, SI Materials and Methods* for additional experimental details. All experiments involving mice were performed using protocols approved by the University of Wisconsin–Madison Animal Care and Use Committee.

ACKNOWLEDGMENTS. We thank members of the M.J.M.-N. and E.G.R. laboratories for helpful discussions, and Dr. Tara Essock-Burns for her assistance with confocal imaging. This work was funded by NIH Grants R37 AI50661 (to M.M.-N) and R01 OD11024 (to E.G.R.). F.E.R. was supported by the Office of the Vice-Chancellor for Research and Graduate Education at the University of Wisconsin–Madison, with funding from the Wisconsin Alumni Research Foundation.

1. Blaser MJ, et al. (2016) Toward a predictive understanding of Earth's microbiomes to address 21st century challenges. *MBio* 7:e00714-16.
2. Bosch TC, McFall-Ngai MJ (2011) Metaorganisms as the new frontier. *Zoology (Jena)* 114:185–190.
3. Schroeder BO, Bäckhed F (2016) Signals from the gut microbiota to distant organs in physiology and disease. *Nat Med* 22:1079–1089.
4. Hsu T, et al. (2016) Urban transit system microbial communities differ by surface type and interaction with humans and the environment. *mSystems* 1:e00018-16.
5. Kembel SW, et al. (2014) Architectural design drives the biogeography of indoor bacterial communities. *PLoS One* 9:e87093.
6. Goyal MS, Venkatesh S, Milbrandt J, Gordon JI, Raichle ME (2015) Feeding the brain and nurturing the mind: Linking nutrition and the gut microbiota to brain development. *Proc Natl Acad Sci USA* 112:14105–14112.
7. Chu H, Mazmanian SK (2013) Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat Immunol* 14:668–675.
8. Sharon G, Sampson TR, Geschwind DH, Mazmanian SK (2016) The central nervous system and the gut microbiome. *Cell* 167:915–932.
9. Schneider KM, et al. (2018) Successful fecal microbiota transplantation in a patient with severe complicated *Clostridium difficile* infection after liver transplantation. *Case Rep Gastroenterol* 12:76–84.
10. Kamo T, Akazawa H, Suzuki JI, Komuro I (2017) Novel concept of a heart-gut axis in the pathophysiology of heart failure. *Korean Circ J* 47:663–669.
11. Serino M, Blasco-Baque V, Nicolas S, Burcelin R (2014) Far from the eyes, close to the heart: Dysbiosis of gut microbiota and cardiovascular consequences. *Curr Cardiol Rep* 16:540.
12. Coppo R (2018) The gut-kidney axis in IgA nephropathy: Role of microbiota and diet on genetic predisposition. *Pediatr Nephrol* 33:53–61.
13. Budden KF, et al. (2017) Emerging pathogenic links between microbiota and the gut-lung axis. *Nat Rev Microbiol* 15:55–63.
14. Samuelson DR, Welsh DA, Shellito JE (2015) Regulation of lung immunity and host defense by the intestinal microbiota. *Front Microbiol* 6:1085.
15. Lin P (2018) The role of the intestinal microbiome in ocular inflammatory disease. *Curr Opin Ophthalmol* 29:261–266.
16. Oresic M, Seppänen-Laakso T, Yetukuri L, Bäckhed F, Hänninen V (2009) Gut microbiota affects lens and retinal lipid composition. *Exp Eye Res* 89:604–607.
17. Zárate-Bladés CR, et al. (2017) Gut microbiota as a source of a surrogate antigen that triggers autoimmunity in an immune privileged site. *Gut Microbes* 8:59–66.
18. Liang X, FitzGerald GA (2017) Timing the microbes: The circadian rhythm of the gut microbiome. *J Biol Rhythms* 32:505–515.
19. Zhao L, Zhang C (2017) Microbiome: Keeping rhythm with your gut. *Nat Microbiol* 2: 16273.
20. Barr T, et al. (2018) Concurrent gut transcriptome and microbiota profiling following chronic ethanol consumption in nonhuman primates. *Gut Microbes* 9:338–356.
21. Chowdhury SR, et al. (2007) Transcriptome profiling of the small intestinal epithelium in germfree versus conventional piglets. *BMC Genomics* 8:215.
22. Sommer F, Nookaew I, Sommer N, Fogelstrand P, Bäckhed F (2015) Site-specific programming of the host epithelial transcriptome by the gut microbiota. *Genome Biol* 16:62.
23. Thaiss CA, et al. (2016) Microbiota diurnal rhythmicity programs host transcriptome oscillations. *Cell* 167:1495–1510.e12.
24. Thion MS, et al. (2018) Microbiome influences prenatal and adult microglia in a sex-specific manner. *Cell* 172:500–516.e16.
25. Holmes E, Li JV, Athanasiou T, Ashrafian H, Nicholson JK (2011) Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol* 19:349–359.
26. Postler TS, Ghosh S (2017) Understanding the holobiont: How microbial metabolites affect human health and shape the immune system. *Cell Metab* 26:110–130.

27. McFall-Ngai M (2014) Divining the essence of symbiosis: Insights from the squid-vibrio model. *PLoS Biol* 12:e1001783.
28. McFall-Ngai MJ (2014) The importance of microbes in animal development: Lessons from the squid-vibrio symbiosis. *Annu Rev Microbiol* 68:177–194.
29. Dethlefsen L, McFall-Ngai M, Relman DA (2007) An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449:811–818.
30. Heath-Heckman EA, Foster J, Apicella MA, Goldman WE, McFall-Ngai M (2016) Environmental cues and symbiont microbe-associated molecular patterns function in concert to drive the daily remodelling of the crypt-cell brush border of the *Euprymna scolopes* light organ. *Cell Microbiol* 18:1642–1652.
31. Lamarcq LH, McFall-Ngai MJ (1998) Induction of a gradual, reversible morphogenesis of its host's epithelial brush border by *Vibrio fischeri*. *Infect Immun* 66:777–785.
32. Visick KL, Foster J, Doino J, McFall-Ngai M, Ruby EG (2000) *Vibrio fischeri lux* genes play an important role in colonization and development of the host light organ. *J Bacteriol* 182:4578–4586.
33. Graf J, Ruby EG (1998) Host-derived amino acids support the proliferation of symbiotic bacteria. *Proc Natl Acad Sci USA* 95:1818–1822.
34. Schwartzman JA, et al. (2015) The chemistry of negotiation: Rhythmic, glycan-driven acidification in a symbiotic conversation. *Proc Natl Acad Sci USA* 112:566–571.
35. Wier AM, et al. (2010) Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. *Proc Natl Acad Sci USA* 107:2259–2264.
36. Carey HV, Duddleston KN (2014) Animal-microbial symbioses in changing environments. *J Therm Biol* 44:78–84.
37. Montgomery MK, McFall-Ngai MJ (1992) The muscle-derived lens of a squid bioluminescent organ is biochemically convergent with the ocular lens. Evidence for recruitment of aldehyde dehydrogenase as a predominant structural protein. *J Biol Chem* 267:20999–21003.
38. Peyer SM, Heath-Heckman EAC, McFall-Ngai MJ (2017) Characterization of the cell polarity gene crumbs during the early development and maintenance of the squid-vibrio light organ symbiosis. *Dev Genes Evol* 227:375–387.
39. Tong D, et al. (2009) Evidence for light perception in a bioluminescent organ. *Proc Natl Acad Sci USA* 106:9836–9841.
40. McFall-Ngai M, Heath-Heckman EA, Gillette AA, Peyer SM, Harvie EA (2012) The secret languages of coevolved symbioses: Insights from the *Euprymna scolopes-Vibrio fischeri* symbiosis. *Semin Immunol* 24:3–8.
41. Gestal C, Castellanos-Martínez S (2015) Understanding the cephalopod immune system based on functional and molecular evidence. *Fish Shellfish Immunol* 46:120–130.
42. Bose JL, Rosenberg CS, Stabb EV (2008) Effects of *luxCDABEG* induction in *Vibrio fischeri:* Enhancement of symbiotic colonization and conditional attenuation of growth in culture. *Arch Microbiol* 190:169–183.
43. Casaburi G, Goncharenko-Foster I, Duscher AA, Foster JS (2017) Transcriptomic changes in an animal-bacterial symbiosis under modeled microgravity conditions. *Sci Rep* 7:46318.
44. Kremer N, et al. (2018) Persistent interactions with bacterial symbionts direct mature-host cell morphology and gene expression in the squid-vibrio symbiosis. *mSystems* 3: e00165-18.
45. Albertin CB, et al. (2015) The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524:220–224.
46. Belcaid M, et al. (January 11, 2019) Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proc Natl Acad Sci USA* 116:3030–3035.
47. Alon S, et al. (2015) The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. *eLife* 4:e05198.
48. da Fonseca RR, et al. (2016) Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Mar Genomics* 30:3–13.
49. Bose JL, et al. (2007) Bioluminescence in Vibrio fischeri is controlled by the redox-responsive regulator ArcA. *Mol Microbiol* 65:538–553.
50. Ruby EG, McFall-Ngai MJ (1999) Oxygen-utilizing reactions and symbiotic colonization of the squid light organ by *Vibrio fischeri*. *Trends Microbiol* 7:414–420.
51. Schwartzman JA, Ruby EG (2016) A conserved chemical dialog of mutualism: Lessons from squid and vibrio. *Microbes Infect* 18:1–10.
52. Kremer N, et al. (2013) Initial symbiont contact orchestrates host-organ-wide transcriptional changes that prime tissue colonization. *Cell Host Microbe* 14:183–194.
53. Nyholm SV, Stewart JJ, Ruby EG, McFall-Ngai MJ (2009) Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environ Microbiol* 11: 483–493.
54. Rageh AA, et al. (2016) Lactoferrin expression in human and murine ocular tissue. *Curr Eye Res* 41:883–889.
55. Ogura A, Ikeo K, Gojobori T (2004) Comparative analysis of gene expression for convergent evolution of camera eye between octopus and human. *Genome Res* 14: 1555–1561.
56. Yoshida MA, et al. (2015) Molecular evidence for convergence and parallelism in evolution of complex brains of cephalopod molluscs: Insights from visual systems. *Integr Comp Biol* 55:1070–1083.
57. Kingston AC, Kuzirian AM, Hanlon RT, Cronin TW (2015) Visual phototransduction components in cephalopod chromatophores suggest dermal photoreception. *J Exp Biol* 218:1596–1602.
58. Mäthger LM, Roberts SB, Hanlon RT (2010) Evidence for distributed light sensing in the skin of cuttlefish, *Sepia officinalis*. *Biol Lett* 6:600–603.
59. Takei Y (2001) Does the natriuretic peptide system exist throughout the animal and plant kingdom? *Comp Biochem Physiol B Biochem Mol Biol* 129:559–573.
60. McFall-Ngai MJ, Ruby EG (1998) Sepiolids and vibrios: When first they meet. *Bioscience* 48:257–265.
61. Jones BW, Nishiguchi MK (2004) Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca: Cephalopoda). *Mar Biol* 144:1151–1155.
62. Pan WH, et al. (2018) Exposure to the gut microbiota drives distinct methylome and transcriptome changes in intestinal epithelial cells during postnatal development. *Genome Med* 10:27.
63. Pluznick JL, et al. (2013) Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. *Proc Natl Acad Sci USA* 110:4410–4415.
64. Fujisaka S, et al. (2018) Diet, genetics, and the gut microbiome drive dynamic changes in plasma metabolites. *Cell Rep* 22:3072–3086.
65. McAnulty SJ, Nyholm SV (2017) The role of hemocytes in the Hawaiian bobtail squid, *Euprymna scolopes:* A model organism for studying beneficial host-microbe interactions. *Front Microbiol* 7:2013.
66. Bonaz B, Bazin T, Pellissier S (2018) The vagus nerve at the interface of the microbiota-gut-brain axis. *Front Neurosci* 12:49.
67. Mardinoglu A, et al. (2015) The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Mol Syst Biol* 11:834.
68. Stappenbeck TS, Hooper LV, Gordon JI (2002) Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci USA* 99: 15451–15455.
69. Tahara Y, et al. (2018) Gut microbiota-derived short chain fatty acids induce circadian clock entrainment in mouse peripheral tissue. *Sci Rep* 8:1395.
70. De Vito P (2014) Atrial natriuretic peptide: An old hormone or a new cytokine? *Peptides* 58:108–116.
71. Moskowitz DW, Johnson FE (2004) The central role of angiotensin I-converting enzyme in vertebrate pathophysiology. *Curr Top Med Chem* 4:1433–1454.
72. Salzet M, Verger-Bocquet M (2001) Elements of angiotensin system are involved in leeches and mollusks immune response modulation. *Brain Res Mol Brain Res* 94: 137–147.
73. Heath-Heckman EA, et al. (2014) Shaping the microenvironment: Evidence for the influence of a host galaxin on symbiont acquisition and maintenance in the squid-*Vibrio* symbiosis. *Environ Microbiol* 16:3669–3682.
74. Nyholm SV, McFall-Ngai MJ (1998) Sampling the light-organ microenvironment of *Euprymna scolopes:* Description of a population of host cells in association with the bacterial symbiont *Vibrio fischeri*. *Biol Bull* 195:89–97.
75. Boettcher KJ, Ruby EG, McFall-Ngai MJ (1996) Bioluminescence in the symbiotic squid *Euprymna scolopes* is controlled by a daily biological rhythm. *J Comp Physiol A* 179: 65–73.
76. Theilig F, Wu Q (2015) ANP-induced signaling cascade and its implications in renal pathophysiology. *Am J Physiol Renal Physiol* 308:F1047–F1055.
77. Boudoulas KD, Triposkiadis F, Parissis J, Butler J, Boudoulas H (2017) The cardio-renal interrelationship. *Prog Cardiovasc Dis* 59:636–648.
78. Rinninella E, et al. (2018) The role of diet, micronutrients and the gut microbiota in age-related macular degeneration: New perspectives from the gut-retina axis. *Nutrients* 10:E1677.
79. Gould AL, Dougan KE, Koenigbauer ST, Dunlap PV (2016) Life history of the symbiotically luminous cardinalfish *Siphamia tubifer* (Perciformes: Apogonidae). *J Fish Biol* 89:1359–1377.
80. Haygood MG (1993) Light organ symbioses in fishes. *Crit Rev Microbiol* 19:191–216.
81. Barnett MJ, Toman CJ, Fisher RF, Long SR (2004) A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction. *Proc Natl Acad Sci USA* 101:16636–16641.
82. O'Rourke JA, et al. (2014) An RNA-seq based gene expression atlas of the common bean. *BMC Genomics* 15:866.
83. Douglas AE (2010) *The Symbiotic Habit* (Princeton Univ Press, Princeton, NJ).
84. McCutcheon JP (2013) Genome evolution: A bacterium with a Napoleon complex. *Curr Biol* 23:R657–R659.
85. Sharp KH, Davidson SK, Haygood MG (2007) Localization of *'Candidatus Endobugula sertula'* and the bryostatins throughout the life cycle of the bryozoan *Bugula neritina*. *ISME J* 1:693–702.
86. Meisel JS, et al. (2018) Commensal microbiota modulate gene expression in the skin. *Microbiome* 6:20.
87. Sycuro LK, Ruby EG, McFall-Ngai M (2006) Confocal microscopy of the light organ crypts in juvenile *Euprymna scolopes* reveals their morphological complexity and dynamic function in symbiosis. *J Morphol* 267:555–568.
88. Koch EJ, Miyashiro T, McFall-Ngai MJ, Ruby EG (2014) Features governing symbiont persistence in the squid-vibrio association. *Mol Ecol* 23:1624–1634.
89. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652.
90. Götz S, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435.
91. Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
92. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.

LETTERS

Self versus non-self discrimination during CRISPR RNA-directed immunity

Luciano A. Marraffini[1] & Erik J. Sontheimer[1]

All immune systems must distinguish self from non-self to repel invaders without inducing autoimmunity. Clustered, regularly interspaced, short palindromic repeat (CRISPR) loci protect bacteria and archaea from invasion by phage and plasmid DNA through a genetic interference pathway[1–9]. CRISPR loci are present in ~40% and ~90% of sequenced bacterial and archaeal genomes, respectively[10], and evolve rapidly, acquiring new spacer sequences to adapt to highly dynamic viral populations[1,11–13]. Immunity requires a sequence match between the invasive DNA and the spacers that lie between CRISPR repeats[1–9]. Each cluster is genetically linked to a subset of the *cas* (CRISPR-associated) genes[14–16] that collectively encode >40 families of proteins involved in adaptation and interference. CRISPR loci encode small CRISPR RNAs (crRNAs) that contain a full spacer flanked by partial repeat sequences[2,17–19]. CrRNA spacers are thought to identify targets by direct Watson–Crick pairing with invasive 'protospacer' DNA[2,3], but how they avoid targeting the spacer DNA within the encoding CRISPR locus itself is unknown. Here we have defined the mechanism of CRISPR self/non-self discrimination. In *Staphylococcus epidermidis*, target/crRNA mismatches at specific positions outside of the spacer sequence license foreign DNA for interference, whereas extended pairing between crRNA and CRISPR DNA repeats prevents autoimmunity. Hence, this CRISPR system uses the base-pairing potential of crRNAs not only to specify a target, but also to spare the bacterial chromosome from interference. Differential complementarity outside of the spacer sequence is a built-in feature of all CRISPR systems, indicating that this mechanism is a broadly applicable solution to the self/non-self dilemma that confronts all immune pathways.

S. epidermidis strain RP62a (ref. 20) contains a CRISPR locus that includes a spacer (*spc1*) that is identical to a region of the *nickase* (*nes*) gene found in nearly all sequenced staphylococcal conjugative plasmids (Fig. 1a and Supplementary Fig. 1a), including those that confer antibiotic resistance in methicillin- and vancomycin-resistant *Staphylococcus aureus* strains[21–23]. The *S. epidermidis* CRISPR system limits conjugation between staphylococci by an *spc1*-directed interference pathway[3]. As in other species[2,7,17–19,24,25] the *S. epidermidis* CRISPR locus is transcribed and processed into crRNAs[3], and northern analysis indicates that most *spc1*-containing crRNAs are ~49 nucleotides long (Supplementary Fig. 1b). During CRISPR interference in *S. epidermidis*, target specificity seems to be achieved by direct pairing of the *spc1* crRNA with the plasmid DNA[3]. DNA targeting most likely underlies CRISPR interference in many other bacterial and archaeal species as well[2,26]. Because a sequence match also exists between the crRNA and the CRISPR locus DNA that encodes it, a central issue in CRISPR interference is therefore how the crRNA identifies bona fide targets without attacking the CRISPR locus in the host chromosome.

We demonstrated previously that CRISPR interference can prevent transformation of pC194-based plasmids that contain a target sequence[3]. To test whether CRISPR spacers have an intrinsic ability to evade interference, we cloned the repeat/spacer sequences of the *S. epidermidis* CRISPR locus, along with ~200 base pairs (bp) from either side of the repeats and spacers, into pC194 (Fig. 1a). The resulting plasmid, pCRISPR(wt), which contains three possible interference targets (*spc1*, *spc2* and *spc3*), was transformed into wild-type and Δ*crispr* cells. Unlike the *nes* protospacer-containing plasmid, pCRISPR(wt) transformation efficiency was similar in both strains (Fig. 1a; Supplementary Table 1 shows the transformation efficiency values for all plasmids described here). These results indicate that the potential targets present in the CRISPR locus are specifically exempted from CRISPR interference.

We proposed that differences between flanking regions of spacers and targets — that is, the presence or absence of repeats — could provide the basis for self/non-self discrimination. To test this, we replaced 15 bp from either side of the *nes* target with the corresponding *spc1*-flanking repeat sequences. The resulting plasmids (pNes(5′DR,15) and pNes(3′DR,15)) were tested for CRISPR interference by transformation into wild-type and Δ*crispr* cells (Fig. 1b and Supplementary Fig. 2a, b). Only pNes(5′DR,15) escaped interference, indicating that repeat sequences upstream of a target (that is, adjacent to the 5′ end of the crRNA spacer sequence) can protect that target. Similar experiments narrowed the protective region further to the eight base pairs closest to the target (Fig. 1b and Supplementary Fig. 2b). This region contains five mutations in the *nes* upstream sequence, because the 5′-AGA-3′ sequence from position −5 (that is, 5 bp upstream of the start of the *nes* protospacer) to −3 are shared with *spc1* 5′ flank (Supplementary Fig. 2a). Each of these five mutations was individually tested for its effect on CRISPR interference (Fig. 1c). Only the guanosine-to-adenosine change at position −2 (G-2A) conferred protection. These results demonstrate that repeat sequences upstream of spacers prevent CRISPR interference, and point to position −2 as an important determinant of this effect.

Short (2–4 bp), conserved sequences called 'CRISPR motifs'[12,24,25] or 'protospacer adjacent motifs' (PAMs)[27] have been found near protospacers in other CRISPR systems, and mutations in these motifs can compromise interference[12,25]. A CRISPR motif has not been defined in *S. epidermidis* protospacers, of which only two have been specified[3]. To test whether G-2 is part of a conserved motif important for interference, we tested plasmids carrying the mutations G-2C and G-2T. Surprisingly, unlike the G-2A mutation, C and T transversions had no effect on transformation efficiency (Fig. 1c). This result was corroborated in a conjugation assay[3] using pG0400 G-2A or G-2C mutants (Supplementary Fig. 3) and excludes the possibility that a G at position −2 is simply a crucial CRISPR motif residue. Instead, this observation indicates that only an A at position −2, that is, the nucleotide present in the repeats, allows protection, and that any deviation from this nucleotide enables interference.

In light of this observation, we considered the complementarity of crRNAs with target protospacers and with CRISPR sequences:

[1]Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2205 Tech Drive, Evanston, Illinois 60208, USA.

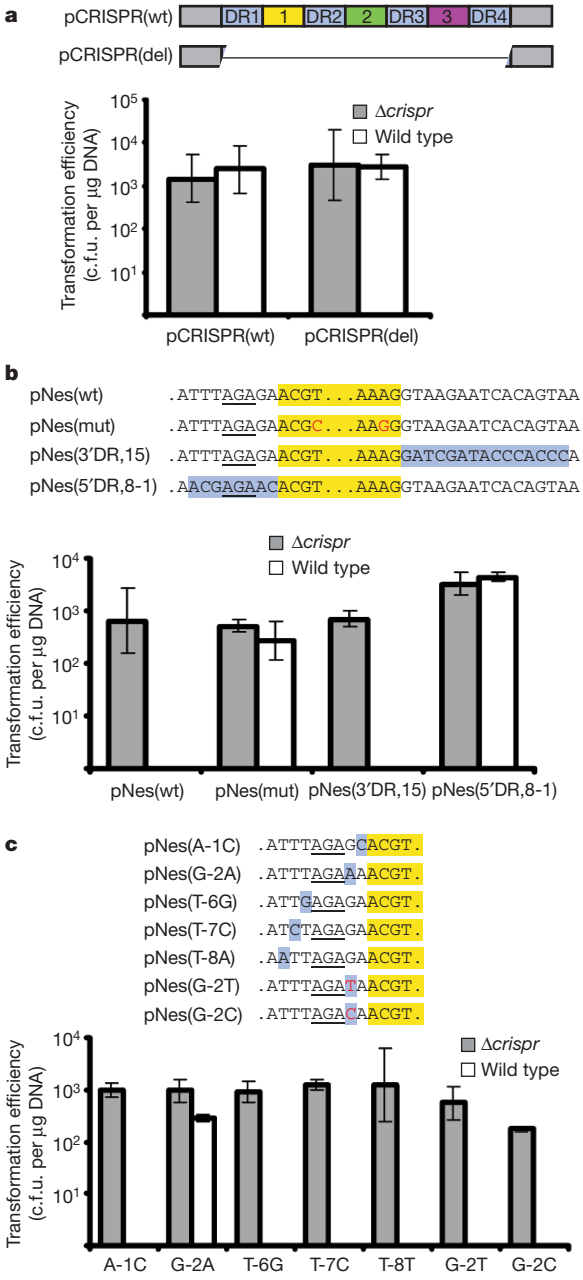


Figure 1 | Protection of *nes* target by spacer flanking sequences. **a**, Direct repeats (DR1–4, purple boxes) and spacers (1–3, coloured boxes) of the *S. epidermidis* RP62a CRISPR locus were cloned into pC194 generating pCRISPR(wt) and its deletion variant, pCRISPR(del). **b**, pNes(wt) and pNes(mut) contain wild-type and mutated *nes* target sequence of pG0400 (highlighted in yellow, with mutations in red). In pNes(5′DR,8-1) and pNes(3′DR,15) *nes* target flanks were replaced by repeat sequences present upstream and downstream of *spc1* (highlighted in purple), respectively. **c**, Individual nucleotides upstream of *nes* target were replaced by those present upstream of *spc1* (highlighted in purple). The G at position −2 was also changed to C and T (in red). The AGA sequence (underlined) is shared by both the *nes* target and spacer 5′ flanking sequences. All plasmids were transformed into *S. epidermidis* RP62a and its isogenic Δ*crispr* mutant. The average of at least three independent measures of the transformation efficiency (determined as colony forming units (c.f.u.) per µg DNA) is reported and error bars indicate 1 s.d.

although the spacer region of a crRNA can pair with target DNA (Fig. 2a) and CRISPR DNA alike, only the CRISPR DNA will be fully complementary with the CRISPR repeat sequences at the crRNA termini. We therefore proposed that specific base pairs in the crRNA/DNA heteroduplex outside of the spacer enable protection, thereby providing a mechanism to avoid autoimmunity. To explore

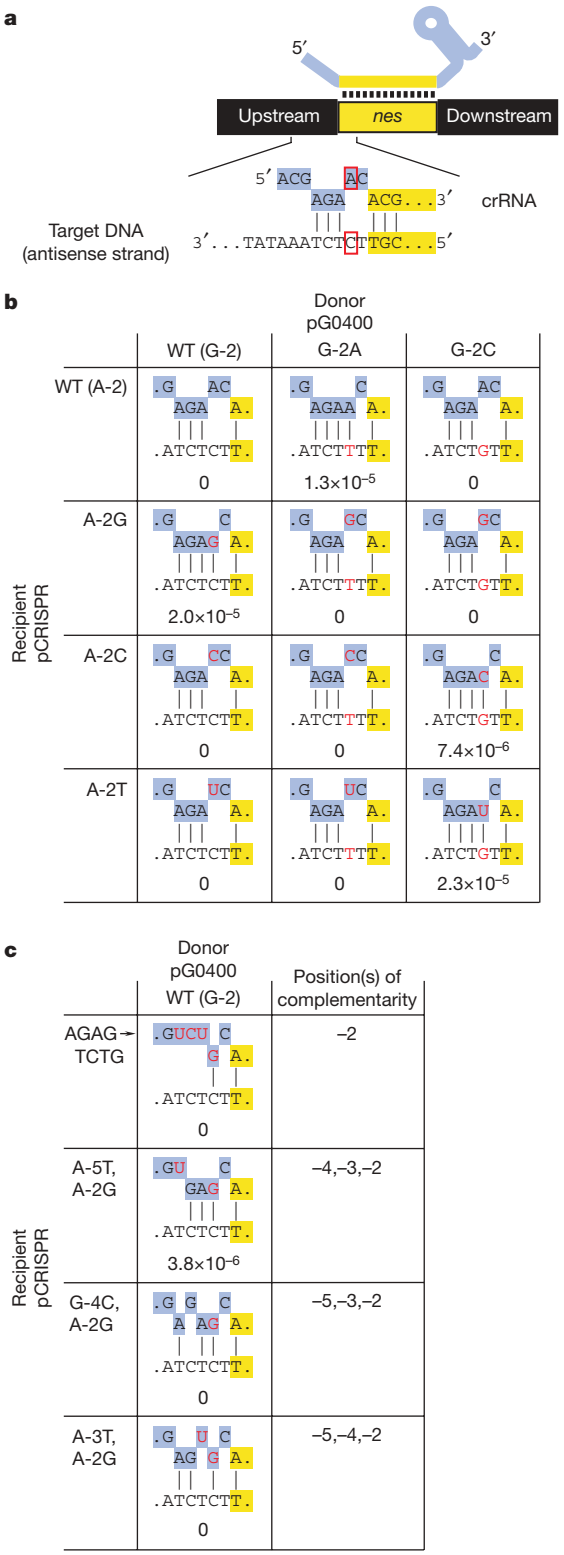


Figure 2 | Complementarity between crRNA and target DNA flanking sequences is required for protection. **a**, Schematic of the complementarity between the flanking sequences of crRNA (top, highlighted in purple) and target DNA (bottom). The red boxes indicate the nucleotides mutated in the experiments shown in **b**. **b**, **c**, Conjugation assays of pG0400 and its mutant variants, using as a recipient the Δ*crispr* strain harbouring different pCRISPR plasmids. Mutations are shown in red. Conjugation efficiency was determined as transconjugant c.f.u. per recipient c.f.u.; the average of at least three independent experiments is reported. WT, Wild type.

this possibility we introduced compensatory mutations in the crRNA (Fig. 2a, b) by changing sequences upstream of *spc1* in pCRISPR(wt), which can complement the interference deficiency of the Δ*crispr* strain. The wild-type *spc1* crRNA contains an adenosine at position −2, so we generated the A-2G, A-2C and A-2T mutants. We then tested each for interference with conjugation of wild-type pG0400 as well as its G-2A and G-2C mutant derivatives. In agreement with our proposal, the *nes* target was protected from interference only when pairing was possible at position −2: rA-dT, rG-dC and rC-dG Watson–Crick appositions each resulted in evasion of the CRISPR system by the conjugative plasmid (Fig. 2b). All crRNA mutants were functional and therefore correctly processed[2], because all pCRISPR plasmids were able to restore interference in Δ*crispr* cells with pG0400 derivatives that were mismatched at position −2 (Fig. 2b). Interestingly, a rU-dG wobble apposition also protected the conjugative plasmid from interference whereas a rG-dT wobble apposition did not, despite the greater stability of rG-dT pairs in an otherwise Watson–Crick-paired heteroduplex[28]. These results provide strong evidence that crRNA/target non-complementarity at position −2 is important for interference. Further mutagenesis of the −5 to −2 sequence of *spc1* crRNA indicated that the minimal complementarity required for protection involves positions −4, −3 and −2 (Fig. 2c and Supplementary Fig. 4). Altogether these results indicate that protection of the *nes* target during conjugation in *S. epidermidis* requires complementarity between crRNA and target upstream flanking sequences at these positions, and strongly indicate that the mechanism of protection requires base-pair formation in this region (Fig. 4a). Furthermore, the base pairing implied by our compensatory analyses reinforces our earlier conclusion[3] that the target of the *S. epidermidis spc1* crRNA is the antisense strand of the *nes* DNA locus, not the sense-oriented mRNA.

If base pairs at positions −4, −3 and −2 confer protection on an otherwise susceptible target, then abolition of base pairing in the same region should confer susceptibility on an otherwise protected CRISPR locus. Deletion analyses in pCRISPR (Fig. 3a and Supplementary Fig. 5a) demonstrated that sequences immediately upstream of spacers prevent autoimmunity. We then tested the effect of substitutions on the direct repeat upstream of *spc1* (*DR1*) on CRISPR protection (Fig. 3b, c). To avoid the expression of mutant crRNAs that would complicate our analyses, we generated pDR1(wt), a deletion variant of pCRISPR(wt) that lacks the 200 bp preceding the CRISPR locus. This plasmid is unable to produce functional crRNAs (data not shown), but is protected from interference in wild-type cells (Fig. 3b, c). In contrast, plasmid pDR1(AGAA→TCTG), which contains mutations at positions −5 to −2, was subject to CRISPR interference. This indicates that this region is critical for protection of the CRISPR locus. Mutagenesis of this and other regions of *spc1* upstream flanking sequences (Fig. 3b, c and Supplementary Fig. 5b) indicated that at least two consecutive mismatches from positions −4 to −2 are required to eliminate protection of the CRISPR locus (Fig. 4b).

CRISPR systems show a high degree of diversity in *cas* gene content[4,9,15,16], and therefore mechanistic differences between different CRISPR/*cas* subtypes are likely to exist. However, differential crRNA pairing potential with CRISPR loci and invasive targets outside of the spacer region (Fig. 4c) is intrinsic to all CRISPR systems, and therefore the mechanism of self/non-self discrimination that we have defined in *S. epidermidis* could apply broadly. The specific base pairs that are monitored could vary between CRISPR/*cas* systems, and we speculate that protein components of the system may 'proofread' paired versus unpaired structures at these sites in a manner that either aborts or enables later steps in interference, respectively. Our findings also highlight the importance of the previously noted 5′-terminal homogeneity of crRNAs[2,3,17], which consistently contain ~eight nucleotides of upstream repeat sequences. Finally, our results are inconsistent with a critical role for the CRISPR motif[12,25,27] during the interference phase

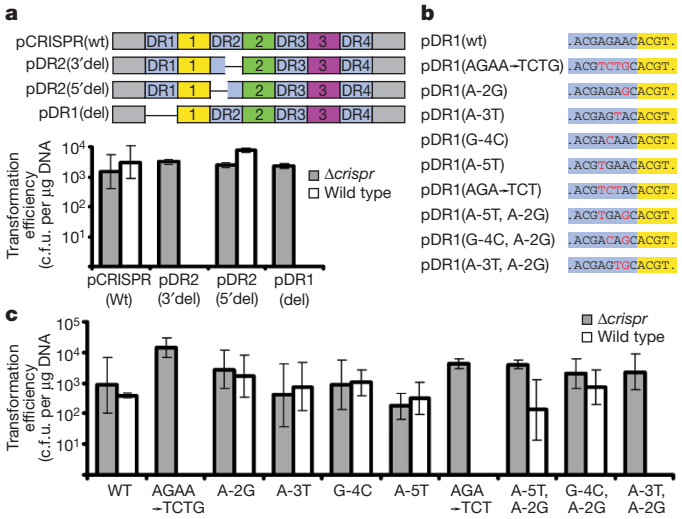


Figure 3 | Mutations in upstream flanking sequences of CRISPR spacers elicit autoimmunity. a, Deletions were performed in the flanking repeats of *spc1*: the 3′ half of DR2, the 5′ half of DR2, and all of DR1 were deleted from pCRISPR(wt) in pDR2(3′del), pDR2(5′del) and pDR1(del), respectively. **b**, **c**, Substitutions (red) were introduced in the 5′ flanking sequence (highlighted in purple) of *spc1* (highlighted in yellow), generating different pDR1 variants that were tested by transformation. All plasmids were transformed into *S. epidermidis* RP62a and Δ*crispr* strains. The average of at least three independent measures of the transformation efficiency (determined as c.f.u. per μg DNA) is reported and error bars indicate 1 s.d.

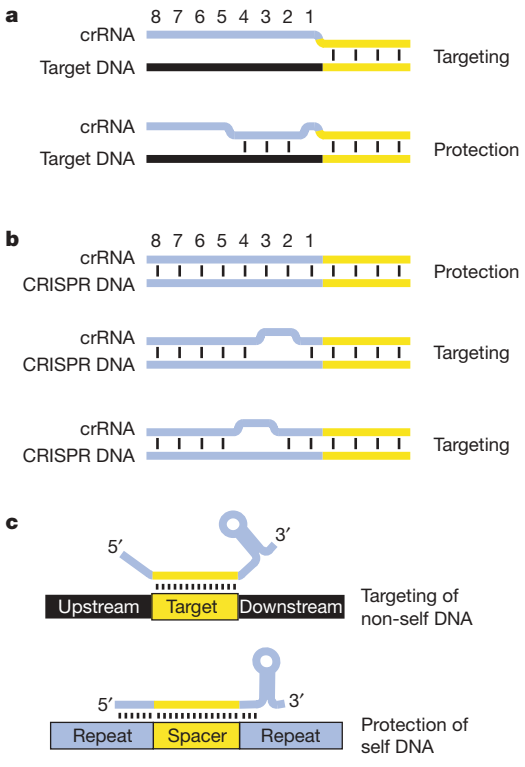


Figure 4 | Requirements for targeting and protection during CRISPR immunity. a, In *S. epidermidis*, CRISPR interference is enabled by mismatches between target DNA and crRNA sequences upstream of the spacer. Formation of at least three base pairs at positions −4, −3 and −2 eliminates targeting. **b**, Complementarity between the *S. epidermidis* CRISPR locus and the crRNA 5′ terminus protects it from interference. Disruption of base pairing at positions −4, −3 or −3, −2 eliminates protection. **c**, General model for the prevention of autoimmunity in CRISPR systems. The ability of crRNA termini (5′, 3′, or both) to base pair with potential targets enables discrimination between self and non-self DNA during CRISPR immunity.

of CRISPR immunity. No significant similarity exists between the two known targets of the *S. epidermidis* CRISPR locus (Supplementary Fig. 5c) and therefore we cannot determine if the nucleotides that are important for discrimination are also part of a CRISPR motif. However, our mutagenesis experiments indicate that no specific flanking nucleotides are required for interference: the decisive characteristic is non-complementarity with the crRNA rather than nucleotide identity *per se*. By extension, it is conceivable that CRISPR motif mutants previously shown to evade interference[12,25] could reflect a failure of self/non-self discrimination via crRNA/target base pairing, rather than a requirement for specific nucleotide identity within the CRISPR motif. We therefore suspect that CRISPR motifs are more important during the acquisition of new spacers[27]. In summary, our results reveal the mechanism of self/non-self discrimination in this recently discovered immune system and will facilitate efforts to exploit CRISPR interference for biotechnological applications.

METHODS SUMMARY

Bacterial strains and growth conditions. *S. epidermidis* wild-type (RP62a, ref. 20) and Δ*crispr* (LAM104, ref. 3) and *S. aureus* (RN4220, ref. 29) strains were grown in brain-heart infusion (BHI) and tryptic soy broth (TSB) media, respectively. When required, the medium was supplemented with antibiotics as follows: neomycin ($15\,\mu g\,ml^{-1}$) for selection of *S. epidermidis*; chloramphenicol ($10\,\mu g\,ml^{-1}$) for selection of pC194-based plasmids; and mupirocin ($5\,\mu g\,ml^{-1}$) for selection of pG0400-based plasmids. *Escherichia coli* DH5α cells were grown in LB medium, supplemented with ampicillin ($100\,\mu g\,ml^{-1}$) or kanamycin ($50\,\mu g\,ml^{-1}$) when necessary.

Conjugation and transformation. Conjugation and transformation were performed as described previously[3] with the following modification: transformations of *S. epidermidis* were recovered at 30 °C in 150 µl of BHI for 6 h. Corroboration of the presence of the desired plasmid in transconjugants or transformants was achieved by extracting DNA of at least two colonies or agar of empty plates, performing PCR with suitable primers and sequencing the resulting PCR product.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 August; accepted 25 November 2009.
Published online 13 January 2010.

1. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
2. Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321,** 960–964 (2008).
3. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322,** 1843–1845 (2008).
4. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR–a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Rev. Microbiol.* **6,** 181–186 (2008).
5. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).
6. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653–663 (2005).
7. Lillestøl, R. K., Redder, P., Garrett, R. A. & Brugger, K. A putative viral defence mechanism in archaeal cells. *Archaea* **2,** 59–72 (2006).
8. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).
9. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34,** 401–407 (2009).
10. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8,** 172 (2007).
11. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320,** 1047–1050 (2008).
12. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190,** 1390–1400 (2008).
13. van der Ploeg, J. R. Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* **155,** 1966–1976 (2009).
14. Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).
15. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLOS Comput. Biol.* **1,** e60 (2005).
16. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1,** 7 (2006).
17. Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14,** 2572–2579 (2008).
18. Tang, T. H. *et al.* Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA* **99,** 7536–7541 (2002).
19. Tang, T. H. *et al.* Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **55,** 469–481 (2005).
20. Gill, S. R. *et al.* Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J. Bacteriol.* **187,** 2426–2438 (2005).
21. Climo, M. W., Sharma, V. K. & Archer, G. L. Identification and characterization of the origin of conjugative transfer (*oriT*) and a gene (*nes*) encoding a single-stranded endonuclease on the staphylococcal plasmid pGO1. *J. Bacteriol.* **178,** 4975–4983 (1996).
22. Diep, B. A. *et al.* Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *Lancet* **367,** 731–739 (2006).
23. Weigel, L. M. *et al.* Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. *Science* **302,** 1569–1571 (2003).
24. Lillestøl, R. K. *et al.* CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* **72,** 259–272 (2009).
25. Semenova, E., Nagornykh, M., Pyatnitskiy, M., Artamonova, I. I. & Severinov, K. Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiol. Lett.* **296,** 110–116 (2009).
26. Shah, S. A., Hansen, N. R. & Garrett, R. A. Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem. Soc. Trans.* **37,** 23–28 (2009).
27. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).
28. Sugimoto, N., Nakano, M. & Nakano, S. Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes. *Biochemistry* **39,** 11270–11281 (2000).
29. Kreiswirth, B. N. *et al.* The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* **305,** 709–712 (1983).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Fang for cloning assistance and members of our laboratory for critical reading of the manuscript. We thank V. Gerbasi and J. Marques for experimental advice. L.A.M. is a Fellow of The Jane Coffin Childs Memorial Fund for Medical Research. This work was supported by a grant from the National Institutes of Health, USA, to E.J.S.

Author Contributions L.A.M. designed experiments with input from E.J.S.; L.A.M. conducted experiments. L.A.M. and E.J.S. analysed data, interpreted experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.A.M. (l-marraffini@northwestern.edu) and E.J.S. (erik@northwestern.edu).

METHODS

DNA cloning. All plasmids used in this study were constructed by cloning CRISPR or *nes* sequences into the HindIII site of pC194 (ref. 30). Inserts were generated by PCR using primers and templates described in Supplementary Table 2. Supplementary Table 3 contains primer sequences. To introduce mutations, internal primers carrying the desired substitution were used with external primers, thus generating two PCR products that were then joined by nested PCR using the external primers. Amplified DNA was cloned into pCR2.1 vector (Invitrogen), the resulting plasmid purified and the insert sequenced. Inserts were cut from pCR2.1 with HindIII, purified and ligated into pC194. Ligation products were transformed into *S. aureus* OS2 cells[31] and colonies containing the desired insert were identified by PCR using primers P86/P87.

pG0400 mutagenesis. Mutations were introduced by allelic exchange as described previously[3]. P15 and P18 primers were used with reverse and forward primers containing the desired mutation, respectively, to amplify 1 kb of *nes* sequence. PCR products were then joined by nested PCR using P15/P18 primers and recombined into pKOR1 (ref. 32). Recombination products were transformed into *E. coli* DH5α, and the resulting plasmids were purified, sequenced and transformed into *S. aureus* RN4220 (pG0400) (ref. 33) for allelic exchange. pG0400(G-2A), pG0400(G-2C) and pG0400(G-2T) mutants were generated using the knockout constructs pLM397, pLM406 and pLM405, respectively. Inserts contained in these plasmids were generated by joining the PCR products obtained with the following primer pairs: pLM397, P15/P190 and P191/P18; pLM406, P15/P208 and P209/P18; pLM405, P15/P206 and P207/P18. See Supplementary Table 3 for primer sequences.

Northern blot analysis. RNA extraction of *S. epidermidis* and oligonucleotide labelling was performed as indicated in ref. 3. Northern blot analysis was carried out according to ref. 34 using ultraviolet crosslinking and a hybridization temperature of 37 °C.

30. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* **150,** 815–825 (1982).
31. Schneewind, O., Model, P. & Fischetti, V. A. Sorting of protein A to the staphylococcal cell wall. *Cell* **70,** 267–281 (1992).
32. Bae, T. & Schneewind, O. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid* **55,** 58–63 (2006).
33. Morton, T. M., Johnston, J. L., Patterson, J. & Archer, G. L. Characterization of a conjugative staphylococcal mupirocin resistance plasmid. *Antimicrob. Agents Chemother.* **39,** 1272–1280 (1995).
34. Pall, G. S., Codony-Servat, C., Byrne, J., Ritchie, L. & Hamilton, A. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.* **35,** e60 (2007).

General Background Reading: The following books are an excellent source of information for fact checking or memory refreshing.

Undergraduate Texts

Stent, G.S. and R. Calendar. 1971. *Molecular Genetics: An Introductory Narrative*, (Second edition), W.H. Freeman, San Francisco, CA.

Watson, J.D., T. Baker, S. Bell, A. Gann, M. Levine, R. Losick. *Molecular Biology of the Gene*, Books a la Carte Edition (6th edition available in Crerar).

Maloy, S.R., J.E. Cronan, Jr., D. Freifelder. 1994. *Microbial Genetics* (Second edition), Jones and Bartlett Publishers, Inc., Boston, MA.

Lewin, B. 2008. *Genes IX*. Jones and Bartlett Publishers, Inc., Sudbury, MA.

Echols, H. (ed. C.A. Gross). 2001. *Operators and Promoters: The Story of Molecular Biology and Its Creators*. University of California Press, Berkeley, CA.

Snyder, L., W. Champness. 2007. *Molecular Genetics of Bacteria.* ASM Press, Washington, DC.

Comprehensive References

Neidhardt, F.C., ed. 1996. *Escherichia coli and Salmonella*: *Cellular and Molecular Biology*, Vol. I and II. (Second edition), ASM Press, Washington, DC.

Advanced Texts or Monographs

Hayes, W. 1964. *The Genetics of Bacteria and Their Viruses*, (Second edition). John Wiley, NY.

Beckwith, J.R. and D. Zipser, eds. 1970. *The Lactose Operon.* Cold Spring Harbor Laboratory, NY.

Hershey, A.D., ed. 1971. *The Bacteriophage Lambda.* Cold Spring Harbor Laboratory, NY.

Hendrix, R.W., J.W. Roberts, F.W. Stahl, and R.A. Weisberg, eds. 1983. *Lambda II.* Cold Spring Harbor, NY.

Scaife, J., D. Leach, and A. Galizzi, eds. 1985. *Genetics of Bacteria.* Academic Press, NY.

Ptashne, M. 1986. *A Genetic Switch.* Cell Press and Blackwell Scientific, Cambridge, MA.

Neidhardt, F.C., J.L. Ingraham and M. Schaechter. 1990. *Physiology of the Bacterial Cell: A Molecular Approach*. Sinauer Associates, Sunderland, MA.

Hoch, J.A. and T.J. Silhavy, eds. 1995. *Two-Component Signal Transduction*. ASM Press, Washington, DC.

Lin, E.C.C. and A.S. Lynch, eds. 1996. *Regulation of Gene Expression*. R.G. Landes Co., Austin, TX.

Storz, G. and R. Hengge-Aronis, eds. 2000. *Bacterial Stress Responses*. ASM Press, Washington, DC.

Ptashne, M. and A. Gann. 2002. *Genes & Signals*. Cold Spring Harbor Laboratory, NY.

Ptashne, M. 2004. *A Genetic Switch: Phage Lambda Revisited.* Cold Spring Harbor, NY.

Historical Works

Beckwith, J. 2002. *Making Genes, Making Waves*. Harvard University Press, Cambridge.

Brenner, S. 2001. *My Life in Science*. BioMed Central Limited, London, UK.

Brock, T.D. 1990. *The Emergence of Bacterial Genetics.* Cold Spring Harbor Laboratory, NY.

Cairns, J., G.S. Stent, and J.D. Watson, eds. 1966. *Phage and the Origins of Molecular Biology*. Cold Spring Harbor Laboratory, NY.

Friedberg, E. 2010. *Sydney Brenner: A Biography.* Cold Spring Harbor Laboratory, NY.

Jacob, F. 1988. *The Statue Within.* Basic Books, NY.

Joklik, W.K., L.G. Ljungdahl, A.D. O'Brien, A. von Graevenitz and C. Yanofsky. 1999. *Microbiology: A Centenary Perspective*. American Society for Microbiology, Washington, DC.

Judson, H.F. 1996. *The Eighth Day of Creation: The Makers of the Revolution in Biology*. Cold Spring Harbor Laboratory, NY.

Keller, E.F. 1983. *A Feeling for the Organism: The Life and Work of Barbara McClintock*, A.W.H. Freeman/OwlBook, Henry Holt and Company LLC, New York.

Lwoff, A. and A. Ullman, eds. 1979. *Origins of Molecular Biology: A Tribute to Jacques Monod*. Academic Press, NY.

Maddox, B. 2002. *Rosalind Franklin: the Dark Lady of DNA.* HarperCollins Publishers, UK.

Olby, R. 2009. *Francis Crick Hunter of Life's Secrets*. Cold Spring Harbor Laboratory, NY.

Weiner, J. 1999. *Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior.* Alfred A. Knopf, New York.