

# THE ART AND DESIGN OF GENETIC SCREENS: *ESCHERICHIA COLI*

Howard A. Shuman\* and Thomas J. Silhavy<sup>‡</sup>

This article summarizes the general principles of selections and screens in *Escherichia coli*. The focus is on the *lac* operon, owing to its inherent simplicity and versatility. Examples of different strategies for mutagenesis and mutant discovery are described. In particular, the usefulness and effectiveness of simple colour-based screens are illustrated. The power of *lac* genetics can be applied to almost any bacterial system with gene fusions that hook any gene of interest to *lacZ*, which is the structural gene that encodes  $\beta$ -galactosidase. The diversity of biological processes that can be studied with *lac* genetics is remarkable and includes DNA metabolism, gene regulation and signal transduction, protein localization and folding, and even electron transport.

## MICROBIAL GENETICS

The development of technologies such as large-scale genome sequencing and various types of microarray have added significantly to our understanding of bacteria. Indeed, these technologies have provided vast amounts of information about hundreds of microorganisms. Although this information provides new and challenging ways to think about biological problems, it cannot replace genetics. Closer examination graphically shows how much we have to learn about even the premier model bacterial species *Escherichia coli*. For example, in almost all genomes, including that of *E. coli*, approximately 30–40% of the identified open reading frames, or putative genes, have no known function or even a recognizable relative with a known function. Published tables of microarray data are replete with entries such as ‘hypothetical gene’. We are left with hundreds of genes without assigned functions. We think it unlikely that genomics, proteomics or any other ‘-omics’ alone can provide data that show function. Because forward genetic analysis — that is, starting with a biological function and proceeding to gene discovery — has been so successful over the past 60 years, it seems obvious that bacterial genetics will be needed now more than ever if we are to realize the full potential of this ‘-omic-derived’ information. Given the emphasis and appeal of the large-scale methods, much of the basis of genetic selections and screens is in danger of being lost

to those researchers who are trained only in these sophisticated ‘high-tech’ methods. We hope to illustrate here the extraordinary power of toothpicks and logic.

The origins of bacterial genetics derive from the insight of Beadle and Tatum<sup>1</sup> that the biochemical defects that are seen in complex organisms, such as man and flies, can also be studied in microorganisms. Their initial attempts to isolate biochemical mutants in *Neurospora crassa* were rapidly superseded by the popular use of *E. coli* and related bacteria for genetic studies. The ease of manipulating enormous clonal bacterial populations, and the discovery of recombination in bacteria by Lederberg<sup>2</sup>, paved the way for generations of geneticists to devise simple and effective means for the isolation and detailed characterization of a wide range of mutants.

The key to designing effective genetic screens lies in the understanding and use of basic bacterial physiology. The progress that has been made in defining pathways and metabolic intermediates in bacteria means that thousands of different phenotypic variants can be identified. These range from mutants that are unable to synthesize specific macromolecular precursors — such as amino acids or nucleosides — to mutants that are unable to catabolize a range of carbon or nitrogen sources. The concept of conditionally lethal phenotypes has allowed the isolation of mutants with defects in

\*Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York 10032, USA.

<sup>‡</sup>Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. Correspondence to T.J.S. e-mail: tsilhavy@molbio.princeton.edu doi:10.1038/nrg1087

Table 1 | **Common selections and phenotypic screens in bacteria**

Method	Result	Example
<b>Selection</b>		
Growth on various carbon sources (such as lactose, maltose and galactose)	Expression of wild-type genes that encode catabolic pathways	Isolation of Mel <sup>r</sup> revertants of lacO <sup>+</sup> (see main text)
Prototrophy, growth in the absence of further biosynthetic intermediates (such as amino acids, nucleotide precursors and vitamins)	Expression of wild-type genes that encode biosynthetic pathways	Reversion of <i>his</i> auxotroph
Resistance to phage	Loss of function or decreased expression of gene that encodes phage receptor	Identification of receptor for bacteriophage-λ
Resistance to antibiotic	Loss-of-function mutation in gene that encodes transporter; missense mutation in gene that encodes antibiotic target; up regulation of efflux system; gene that encodes inactivating enzyme	Decreased outer-membrane permeability for β-lactams; resistance to streptomycin alters ribosomal protein S12; plasmid encoded tetracycline exporter; plasmid encoded aminoglycoside phosphotransferases
Resistance to metabolic analogues	Loss of function in gene that encodes transporter; missense mutation in gene that encodes target enzyme	(See main text)
<b>Phenotypic screen</b>		
Ability to ferment sugars, (such as lactose, maltose and galactose), MacConkey, tetrazolium and so on; colour changes indicate the amount of acid production from catabolism	Various colour changes indicate the overall degree of fermentation, correlated to levels of gene expression	(See main text)
Chromogenic substrate (such as Xgal)	Formation of coloured product correlated to levels of hydrolytic enzyme (such as β-galactosidase)	(See main text)
Light production from luciferase	Formation of light correlated to levels of luciferase enzyme reporter	Identification of quorum-sensing systems using <i>lux</i> genes
Fluorescence from green fluorescent protein (GFP)	Intensity indicates level of <i>gfp</i> gene expression; intracellular location can indicate specific localization signals	Identification of promoters that are activated during intracellular growth of bacterial pathogens

essential functions such as DNA replication and protein synthesis. Following the discovery of antibiotics in the 1950s and 1960s, the ability to isolate specific antibiotic-resistant mutants yielded new insights into the drug targets and their mechanisms of action. TABLE 1 lists some general approaches to the isolation of bacterial mutants with a range of phenotypes.

Rather than presenting an exhaustive compendium of selections and screens, here we focus on one particular genetic system, the *lac* operon, owing to its inherent simplicity and wide applicability. Before introducing the *lac* system, we provide some essential concepts that illustrate the fundamental differences between selections and screens in bacteria. We then describe several screens and selections that are based on the ability of *E. coli* to metabolize different galactosides. Some of the galactosides, or their byproducts, result in visible colour changes in bacterial colonies. These visual cues, similar to those that were originally applied to screen fly populations for changes in eye colour, form the basis of many powerful screens.

Finally, we describe the ability to fuse the *lac* genes to any other gene, and show how the powerful screens and selections of the *lac* operon can then be applied to almost any bacterial gene. Several examples illustrate variations on the original approaches. These variations have provided extensive information about the molecular basis of basic cell functions such as DNA transposition, transcriptional and translational regulation, and protein secretion and folding.

#### **Selections versus screens**

One advantage of working with bacteria is that rare mutants can be identified among large populations (of the order of 10<sup>10</sup> individuals). All that is required is an environment in which only the rare mutant can grow; in other words, an appropriate selection. For example, base-pair changes at only a few positions in the *rpsL* gene can lead to resistance to streptomycin, but mutants that carry these changes can be easily isolated over and over again, because selection for streptomycin resistance is powerful.

Box 1 | **Mutagenesis**

As a general rule, spontaneous mutations are preferred over mutagen induced mutations. All possible base substitutions, frameshifts, deletions and rearrangements can occur spontaneously. If it is possible to design a selection for the desired mutant then a mutagen should not be used. Mutagens are useful for mutant screens, but all mutagens have disadvantages.

In general, there are three types of mutagen: base analogues, alkylating agents and radiation that damages DNA. One commonly used base analogue is 2-aminopurine. It is less dangerous than other mutagens; however, it is weak and causes a spectrum of mutations that is limited to transitions (a purine to a purine, or a pyrimidine to a pyrimidine, on the same DNA strand). If null mutations are of interest, then this limited spectrum is not a serious problem. However, it might be that the specific phenotype in question cannot be generated by a transition mutation.

Nitrosoguanidine is a mutagen that alkylates DNA. Because of its potency, it is commonly used. However, it also induces primarily transition mutations, and multiple mutations are not uncommon. Ultraviolet (UV) damages DNA by producing photoproducts such as thymine dimers. Mutagenesis occurs if cells repair this damage in an error-prone fashion. UV is a favourite of many geneticists because it causes a broad spectrum of different types of mutation: transition, transversion (a purine to a pyrimidine, or a pyrimidine to a purine, on the same DNA strand), frameshift, deletion and rearrangement. However, it is far more lethal to *Escherichia coli* than the other mutagens, is not potent and there are frequent hot spots.

Certain *E. coli* mutator strains have mutation frequencies that are increased by orders of magnitude. Perhaps the most common gene that is inactivated in such strains is *mutD*. Inactivation of this gene destroys the proofreading function of DNA polymerase III. The obvious advantage of a mutator strain is that no dangerous chemicals are required. Also, the spectrum of mutations that is obtained is broad. However, strain construction is required to introduce the mutator gene into the strain that is used for the selection or screen.

For an excellent review of mutagens and mutagenesis techniques see REF. 45; transposon mutagenesis is discussed in BOX 2.

Box 2 | **Use of transposons and genomics to search for mutants**

Transposon mutagenesis is commonly used for increasing the effectiveness of mutant screens. A wide range of transposon derivatives and delivery vectors are available (for example, see REF. 46). Similar to insertion sequence (IS) elements, transposons can insert in DNA in a near-random fashion. Indeed, many transposons are mobilized by IS elements. However, they are more useful because they can carry selectable markers, usually a gene that confers resistance to an antibiotic such as tetracycline. Transposons are introduced into bacterial strains on a vector that cannot replicate — for example, a mutant  $\lambda$ -phage — and selection is made for the marker that is carried on the transposon. The stable inheritance of the marker is only possible if the transposon has inserted into the bacterial genome. Accordingly, every cell that survives the selection is a mutant. This mutant population is then screened for variants that show the desired phenotype.

Transposon insertion in a gene usually inactivates that gene. So, transposon mutagenesis is not recommended for screens that target essential genes. Also, transposon insertions in operons can affect the expression of downstream genes owing to polarity. Except for these caveats, transposon mutagenesis is a powerful tool. It is highly efficient: as noted earlier, every cell that inherits the transposon is a mutant. However, mutagenized cells that carry more than one transposon are rare. Because the transposon carries a selectable marker, the insertion mutation that is caused by the transposon can be moved from one strain to another by transduction. Arguably, the most useful feature of transposon mutagenesis is that the site of insertion can be determined by PCR amplification of the insertion site directly from the mutant genome and subsequent DNA sequence analysis. PCR amplification from the genome uses an anchored random PCR strategy<sup>47</sup>, and DNA sequence analysis allows the precise identification of the insertion site because the complete genome sequence of *Escherichia coli* is known<sup>48</sup>.

The fact that transposon insertion sites can be readily identified at the level of the DNA sequence obviates the need for laborious genetic mapping to determine the genes that are affected by mutations identified in screens. This allows mutant screening on a scale that was previously unimaginable.

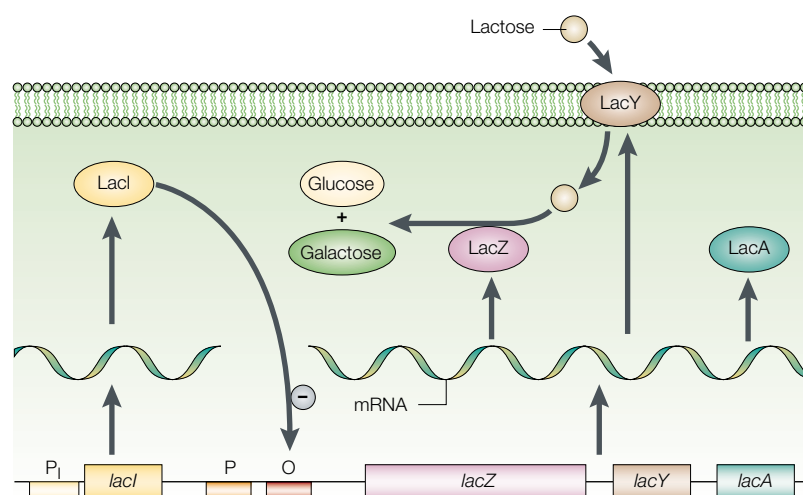
Geneticists use both forward and reverse selections. A forward selection allows the growth of mutants in which a particular gene function is lost. For example, the selection used by Luria and Delbrück<sup>3</sup> for resistance to phage T1 is a forward selection that allows the growth of bacteria that have lost the receptor for this BACTERIOPHAGE. A reverse selection allows the growth of revertants in which a particular gene function is restored. For example,  $\lambda$ -LYSOGENS of *E. coli* produce a protein called Rex that kills cells on infection by a T4 *rII*-mutant bacteriophage. This altruistic act prevents the spread of the phage infection through the bacterial population. Consequently, T4-*rII* mutants will not form PLAQUES on a lawn of an *E. coli* lysogen. Rare plaques that form in a lawn of bacteria that have been infected with the mutant phage are revertants of the *rII* mutation. These revertants, such as wild-type T4 phage, produce a functional rII protein that counteracts the lethal effects of Rex. Such selections are useful because they allow REVERSION ANALYSIS, which provides insights into the molecular nature of the mutation in question. Indeed, Crick *et al.*<sup>4</sup> used this reverse selection to determine the molecular nature of frameshift mutations and to uncover the triplet genetic code. Reverse selections also allow the identification of unlinked suppressor mutations, and suppressor analysis is one of the most versatile tools that is available to the geneticist.

In contrast to selections, screens provide a means to detect variants in a population. The main difference is that all of the members of a population must be examined. Accordingly, screens are not as effective as selections because fewer individuals can be examined. It would not be unreasonable to screen  $10^5$  colonies, but  $10^{10}$  is not possible as millions of petri dishes would be required. Screens can be made more effective by mutagenesis. For example, DeLucia and Cairns<sup>5</sup> mutagenized a population of *E. coli* and grew the mutagenized cells as single colonies on agar plates. Individual colonies were cultured in liquid media, and a mutant that lacks DNA polymerase I (*polA*) was identified by assaying enzyme activity in PERMEABILIZED CELLS. This required the screening of a few thousand individual colonies. The identification of this *polA* mutant provided strong evidence that DNA polymerase I was not the enzyme that was responsible for DNA replication. BOX 1 provides further information about the types of mutagenesis that are available to the bacterial geneticist.

Screens can be useful because they allow the detection of a broad dynamic range of gene activity. They can be designed to detect increases in activity (ups) or decreases in activity (downs). CHROMOGENIC substrates and indicator agars provide a means to detect these changes in activity.

**Red, white and blue: the tools of the *lac* system**

Hundreds of successful selections and screens have been applied to *E. coli*<sup>6</sup>. However, although they are useful for the specialist, long lists of selections and screens are of limited value to illustrate general principles.



**Figure 1 | The lactose operon of *Escherichia coli*.** The genes *lacZ*, *lacY* and *lacA* are transcribed from a single promoter (P) that produces a single mRNA from which the three proteins are translated. The operon is regulated by Lac repressor, the product of the *lacI* gene, which is transcribed from its own promoter (P<sub>I</sub>). The repressor inhibits transcription by binding to the *lac* operator (O). Repressor binding to the operator is prevented by the INDUCER.

#### BACTERIOPHAGE

A virus that infects and replicates in bacteria. These are either lytic viruses, which always kill the host, or temperate viruses, which can either lyse the host cell or establish a stable relationship in which the bacteriophage genome is stably maintained in the host genome (see also lysogen).

#### LYSOGEN

A bacterial strain that harbours the genome of a temperate bacteriophage.

#### PLAQUE

A zone of lysed bacterial lawn that is formed by the growth of a bacteriophage.

#### REVERSION ANALYSIS

The isolation and characterization of revertants, or secondary mutants, that regain a wild-type-like phenotype. This analysis can provide clues to the nature of the original mutation.

#### PERMEABILIZED CELLS

Cells that have disrupted membranes.

#### CHROMOGENIC

A colourless substrate that produces a coloured product following the action of one or more enzymes.

#### INDUCER

A substance that causes a gene, or set of genes, to be expressed at a higher level.

In TABLE 1 we give examples of the different types of system that are used for screening in *E. coli*. However, here we focus on one experimental system that can be used to study a wide range of biological processes. In so doing, our discussion encompasses logic and methods that are applicable to all genetic screens in *E. coli*.

**The *lac* system.** It can be argued that all of molecular biology stems from early detailed studies with three experimental systems: the *lac* operon of *E. coli*, and the two *E. coli* phages T4 and  $\lambda$  (see also the article by Campbell on p471 of this issue). Of these, the *lac* system remains in widespread use. Indeed, the *lacZ* gene has now been incorporated into the genomes of all genetically tractable experimental organisms. Accordingly, we use the *lac* system as an example. Note, however, that many of the principles illustrated here can be adapted to other systems for which similar experimental tools are available. The choice of the *lac* experimental system allows us to show one of the earliest and most successful applications of what is now termed ‘chemical genetics’ or ‘chemical biology’<sup>77</sup> (see also the article by Hughes on p432 of this issue).

The *lac* operon contains three genes that are transcribed from a single promoter in the order *lacZ*, *lacY* and *lacA* (FIG. 1). The *lacZ* gene specifies the cytoplasmic enzyme  $\beta$ -galactosidase, which hydrolyses lactose into glucose and galactose. The *lacY* gene encodes Lac permease, a cytoplasmic membrane protein that transports lactose into the cytoplasm. The *lacA* gene encodes a cytoplasmic galactoside transacetylase. Mutants that do not produce functional *LacZ* and *LacY* are phenotypically *Lac*<sup>−</sup> (that is, they cannot use lactose as a carbon or energy source, and cannot ferment lactose). Mutants that lack *LacA* have no obvious phenotype. Transcription of the *lac* operon is regulated by Lac repressor, the

product of the *lacI* gene, which is located immediately upstream of *lacZ*. The *lacI* gene is transcribed from its own promoter on the same DNA strand as the *lac* operon. Mutants that lack the repressor express the operon constitutively.

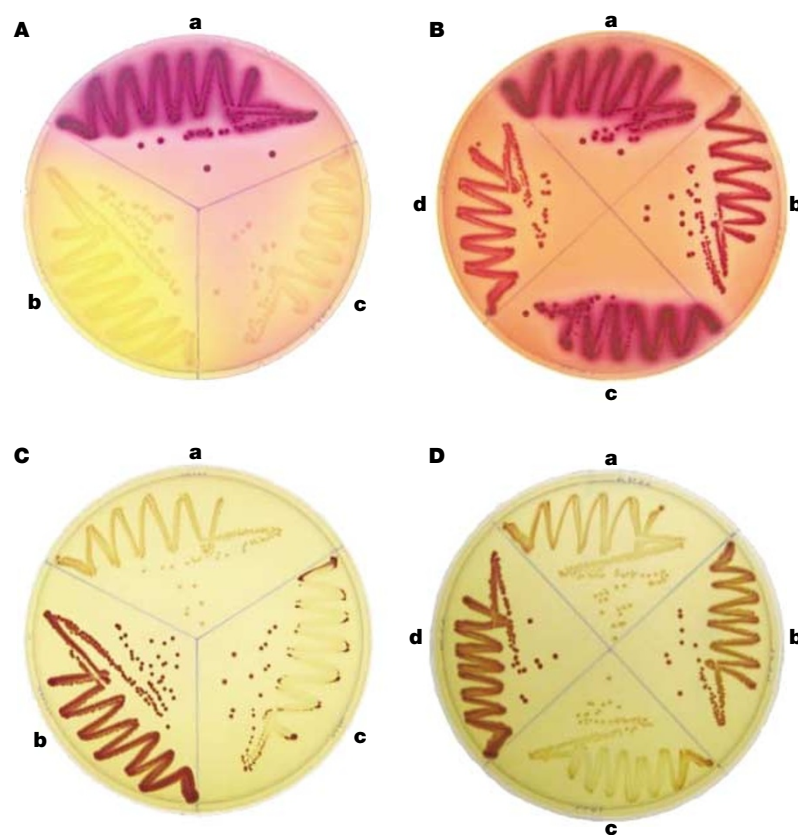
Carbohydrate chemists, such as B. Helferich and his student D. Teurk at the University of Bonn<sup>8</sup>, synthesized a range of galactosides that provided important tools for understanding the regulation of the *lac* operon. Using these reagents, Jacob and Monod discovered that certain galactosides were substrates but not inducers of the *lac* genes, for example, phenyl-galactoside (PG); others were inducers but not substrates, for example, isopropylthiogalactoside (IPTG). This established that regulation and enzyme activity were distinct properties<sup>9</sup>. Moreover, some of these galactosides, if added to growth media, altered either *LacZ* or *LacI* activity, which allowed the selection of mutants that would otherwise be extremely difficult to isolate (see later) — this, we suggest, is the essence of chemical genetics.

**Selections.** It is not possible to select directly for *Lac*<sup>−</sup> mutants. To select such mutants, situations must be designed in which the function of the *lac* gene(s) causes toxicity. Two methods are used: one relies on a metabolic defect, and the other uses a synthetic galactoside.

Mutations that disrupt a catabolic pathway can lead to the accumulation of metabolic intermediates that are toxic at high levels. In combination with certain mutations that block the metabolism of galactose (that is, *galE*) *LacZ* and *LacY* cause cell death if lactose is present. Malamy<sup>10</sup> used this observation to select for spontaneous *Lac*<sup>−</sup> mutants, some of which were caused by INSERTION SEQUENCES (IS elements).

Orthonitrophenylthiogalactoside (TONPG) is a substrate for *LacY*, but is not a substrate for either *LacZ* or *LacI*. High levels of permease cause the accumulation of this metabolic poison to toxic levels. For example, TONPG inhibits the growth of *lacI* mutants on glycerol minimal agar. Muller-Hill *et al.*<sup>11</sup> used TONPG to select PSEUDOREVERTANTS of a *lacI* thermosensitive (Ts) mutant at the non-permissive temperature. Among the mutants that were identified using this selection were strains that overproduced the defective repressor, thereby repressing the synthesis of *LacY*. These *lacI*<sup>q</sup> (quantity) mutants helped to facilitate the original biochemical purification and characterization of the repressor.

To select restoration of function for mutant *LacZ* or *LacY*, it is necessary to select for growth on lactose as the sole carbon source; this is done on lactose minimal agar. It is also possible to select for restoration of *LacY* activity in the absence of *LacZ*. This is done by selecting growth at high temperatures on an agar with melibiose, a different sugar. This works because *LacY* is promiscuous, in that it also transports the  $\alpha$ -galactoside melibiose. In *E. coli* K-12, the normal melibiose transport system, but not  $\alpha$ -galactosidase, is inactive at high temperatures.



**Figure 2 | MacConkey and tetrazolium agars.** The figure shows petri dishes containing either MacConkey agar with maltose (**A,B**) or tetrazolium agar with maltose (**C,D**). **A,C** | These agar plates were seeded with three bacterial strains and incubated for 24 hours: wild type (**a**), *mal* deletion (**b**), and *mal* point mutant (**c**). **B,D** | These agar plates were seeded with four strains with different abilities to use maltose: wild type (**a**), very leaky *mal* mutant (**b**), mutant with no detectable phenotype (**c**), and leaky *mal* mutant (**d**).

**Screens.** One of the reasons that the Lac system is so widely used for genetic screening is because a range of indicator media are available<sup>12</sup>. More importantly, these indicator media vary in sensitivity. This gives the ability to screen for up-regulated (ups) or down-regulated (downs) LacZ activity over a wide range of initial activity.

The most sensitive indicator of LacZ activity is the chromogenic substrate bromochloroindolylgalactoside (Xgal). This colourless compound is hydrolysed by LacZ to produce a compound that dimerizes to form the insoluble dye indigo — the dye that makes ‘blue jeans’ blue. Strains that express the *lac* operon at levels that are too low to enable growth on lactose minimal agar form pale blue colonies on media that contain Xgal. Scaife and Beckwith<sup>13</sup> used this to show that the low levels of LacZ in strains that carry *lac* promoter mutations were still subject to repression by LacI. Xgal can be added to any growth medium, but it should be noted that indigo will not form under anaerobic conditions or if the medium contains high concentrations of a reducing agent. Unlike the indicator agars that are described next, Xgal can be used to detect LacZ activity in the absence of LacY.

**INSERTION SEQUENCES**  
(IS elements). Sequences that can insert themselves into target sites on DNA.

**PSEUDOREVERTANT**  
A revertant that does not re-establish a wild-type genotype.

**ENTERIC BACTERIA**  
Bacteria that normally inhabit the intestines.

**LEAKY DOWN PHENOTYPE**  
A phenotype in which there is a moderate decrease in gene function.

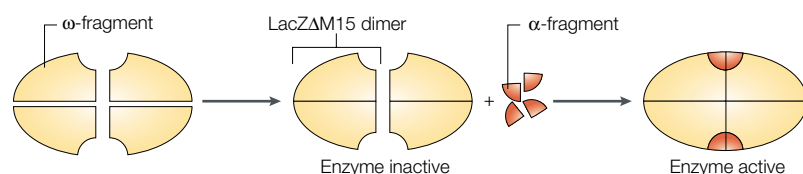
Lactose MacConkey agar (FIG. 2) was developed ~100 years ago for the study of ENTERIC BACTERIA. It is a rich medium that contains bile salts, which inhibit the growth of non-enterics, and the pH indicator phenol red. Enteric bacteria that can ferment the lactose produce acid and form red colonies. *E. coli* is the only enteric that forms red colonies. Lactose MacConkey agar is of intermediate sensitivity. Strains that produce low levels of Lac products, but can grow on minimal lactose agar, might form white colonies or pale pink colonies on lactose MacConkey agar. Strains that produce high levels of Lac products produce enough acid to form a hazy red precipitate of bile salts. Because it is easy to see a few red colonies on a background of white colonies, lactose MacConkey agar is useful for looking for Lac ups. Note that the pH indicator is soluble, so colonies slowly lose their colour over time as it diffuses into the agar.

MacConkey agar can be prepared with other sugars, such as galactose or maltose. Also, several sugars can be added, as can Xgal, which allows the screening of many catabolic pathways simultaneously. For example, Eschenlauer and Reznikoff<sup>14</sup> used galactose MacConkey agar that contained Xgal to distinguish Gal<sup>+</sup>–Lac<sup>+</sup> (purple), Gal<sup>+</sup>–Lac<sup>–</sup> (red), Gal<sup>–</sup>–Lac<sup>+</sup> (blue) and Gal<sup>–</sup>–Lac<sup>–</sup> (white).

Lactose tetrazolium agar (FIG. 2) is a rich medium that contains tetrazolium. All cells that grow on this medium reduce the tetrazolium to an insoluble formazan dye. However, robust acid production by lactose fermentation prevents formazan formation. So, strongly Lac<sup>+</sup> strains produce white colonies, whereas weakly Lac<sup>+</sup> or Lac<sup>–</sup> strains produce red colonies. This medium is particularly useful for detecting strains that reduce *lac* expression without abolishing it. Indeed, Scaife and Beckwith<sup>13</sup> used this medium to distinguish promoter mutations (LEAKY DOWN PHENOTYPE) from *lac*-null mutations. Again, because it is easy to see red colonies on a background of white colonies, lactose tetrazolium agar is useful for looking for Lac downs.

Tetrazolium agar can also be prepared with other sugars, and with more than one sugar. Schwartz and Beckwith<sup>15</sup> used tetrazolium agar that contained arabinose and maltose to identify mutants that were defective in both. These mutants identified the genes for catabolite activator protein (CAP, also known as CRP) and adenylate cyclase.

In summary, the three commonly used Lac indicator agars allow screening for ups and downs. For any given situation, it is best to try all three to see which shows the desired differences most distinctly. MacConkey and Xgal are better suited to detect low levels of lactose catabolism (MacConkey) or  $\beta$ -galactosidase (Xgal) — relatively low levels produce a reddish colour on MacConkey or a blue colour on Xgal. Above a certain level, the differences become less obvious. By contrast, relatively high levels of lactose catabolism are necessary to produce a white colony on tetrazolium agar. Even small decreases in lactose catabolism result in the formation of a pink to red colour. Because these differences are not based on general quantifiable differences, but, rather, on experience with a particular system, they illustrate the ‘art’ of screen design.



**Figure 3 |  $\alpha$ -complementation.** Active  $\beta$ -galactosidase is a tetramer. The *lacZΔM15* deletion removes amino-acid residues 11–41. This large internally deleted protein, often called the  $\omega$ -fragment, can dimerize, but it cannot tetramerize. Amino-terminal fragments of  $\beta$ -galactosidase as short as residues 3–92, the  $\alpha$ -fragments, restore enzymatic activity to LacZΔM15 by allowing tetramers to form.

The indicator agars are sensitive enough to detect sectored colonies; that is, colonies in which different sectors are different colours. As colonies develop from a single cell, sectoring must reflect genetic events that occurred after plating. For example, in the Scaife and Beckwith paper<sup>13</sup> mentioned earlier, the authors mutagenized cells and then plated the population on lactose tetrazolium agar. Because they did not allow the growth of the mutagenized cells before plating, most of the mutant colonies were sectored reflecting the time that was required to fix the mutations and segregate the mutant phenotype.

MacConkey and tetrazolium agars are rich media that support the growth of all cells, whether or not they can metabolize the added sugar. Nonetheless, each can be used as a selective medium. A lawn of Lac<sup>−</sup> bacteria on either medium is white. If reversion occurs during the growth of the lawn, the resulting Lac<sup>+</sup> bacteria can use the high concentrations of lactose that remain. However, the colour reaction on tetrazolium is reversed in a lawn of bacteria; eventually this Lac<sup>+</sup> revertant will overgrow the white lawn and appear as red papillae. It is unclear why Lac<sup>+</sup> revertants are red on a lawn of Lac<sup>−</sup> bacteria. Regardless, the examples cited later illustrate the advantage of using these media in a selective fashion.

It is important for us to point out that the colour phenotype on MacConkey or tetrazolium does not always reflect the *lac* genotype. For example, mutations that alter central metabolism can alter colony colour on these media. So, it is advisable to verify mutant colonies that are obtained from this indicator agar by some assay that measures LacZ directly, such as an Xgal screen or a  $\beta$ -galactosidase assay.

#### Benefits of a simple screen

The simplicity and sensitivity of the screening methods described have been used to study a wide range of biological processes other than lactose catabolism and its regulation. We discuss three examples to illustrate this point.

**The blue/white screen.** CLONING VECTORS that use a blue/white screen are now ubiquitous throughout molecular laboratories, but most who use this method are unaware of the origins of the technology. Behind this immensely useful molecular tool lies one of the classic wrinkles in *lac* genetics. This wrinkle came to light

when it was shown that certain pairs of *lacZ* alleles show intracistronic COMPLEMENTATION; that is, in rare cases, two different *lacZ* mutant genes can produce functional  $\beta$ -galactosidase if they are present in the same cell<sup>16</sup>. The large (1,021 amino acid) tetrameric structure of LacZ was the main feature that allowed intracistronic complementation. Biochemical analysis shows that amino-terminal fragments of LacZ as short as residues 3–92 restore activity to a mutant LacZ that is deleted for residues 11–41 (LacZΔM15)<sup>17</sup>. The amino-terminal fragment is often called the  $\alpha$ -fragment, and this intracistronic complementation is known as  $\alpha$ -complementation (FIG. 3).

Genetically, a wide range of *lacZ* nonsense mutations can provide  $\alpha$ . The large internally deleted protein, often called the  $\omega$ -fragment, is nearly always provided by *lacZΔM15*. Interestingly, the *lacZΔM15* deletion was isolated as a pseudorevertant of a *lacO<sup>0</sup>* mutation using the melibiose selection described earlier<sup>18</sup>. Because the *lacO<sup>0</sup>* mutations prevented the expression of both *lacZ* and *lacY*, it was thought that these mutations might define the *lac* promoter. Beckwith tested this idea by reversion analysis. He realized that by using the melibiose selection he could select for the restoration of LacY permease activity without simultaneously demanding the restoration of  $\beta$ -galactosidase activity. He anticipated that some Mel<sup>+</sup> revertants might also regain  $\beta$ -galactosidase activity (true revertants) and others might not (suppressors). Indeed, both types were found, and these results showed that the *lacO<sup>0</sup>* mutations were not promoter mutations but, rather, strongly POLAR mutations in *lacZ*. Some of the LacZ<sup>+</sup>–LacY<sup>+</sup> revertants resulted in the discovery of new suppressor tRNAs. Other LacZ<sup>+</sup>–Mel<sup>+</sup> revertants were found to eliminate the polar effects of the *lacZ* nonsense mutation and led to the discovery of the transcription anti-termination factor Rho. The *lacZΔM15* deletes the polar block on *lacY* expression and restores the ability to grow on melibiose at high temperature.

Gronenborn and Messing<sup>19</sup> used  $\alpha$ -complementation for the development of the M13 cloning vectors. DNA is cloned into sites in the  $\alpha$ -fragment that are carried on the vector. Vector molecules that have not received inserts retain  $\alpha$ -activity, as evidenced by the blue colonies that are produced on media that contained Xgal by strains that carry the vector and *lacZΔM15*. Colonies that carry vectors with inserts are white in these strains. This is commonly referred to as insertional inactivation. Today,  $\alpha$ -complementation forms the basis for every cloning vector that uses a blue/white screen.

**Mutators.** Mutator genes — genes that, if mutated, result in high rates of mutation — are obviously of great interest (BOX 1). Many mutator genes encode components of the DNA-repair apparatus of bacteria. So, as well as their practical applications, the study of mutator genes has helped to understand how mutational defects lead to human cancers. Nghiem *et al.*<sup>20</sup> used a clever PAPILLATION (FIG. 4) assay to find a mutator gene that generates transversions. They started with a *lacI* mutant strain that carried a *lacZ* nonsense mutation that could

**CLONING VECTORS**  
Engineered plasmids with convenient cloning sites and markers.

**COMPLEMENTATION**  
The use of partial diploids to determine whether two mutations affect the same or different genes. If the mutations are in the same gene, they generally fail to complement each other and the diploid retains the mutant phenotype. By contrast, mutations in different genes usually complement one another and restore a wild-type phenotype to the diploid. However, exceptions to both cases abound.

**POLAR**  
A phenomenon in bacterial transcriptional units in which the transcription of downstream genes is decreased by a translation stop in an upstream gene.

**PAPILLATION**  
The formation of secondary colonies that grow on top of a colony that has already formed on a nutrient agar surface.

## HETERODUPLEX

A DNA molecule that is formed by base pairing between strands that are derived from two DNA molecules that are not identical in sequence.

only revert by transversion mutation. The papillation assay works on single colonies on minimal agar that contains glucose, PG and Xgal. All colonies grow well on glucose, but rare *lacZ*<sup>+</sup> revertants can use PG (recall that PG is a noninducing substrate) as they express the *lac* operon constitutively (*lacI*). These revertants form papillae on the colonies, which the Xgal stains blue. So, a papillating colony is mostly white, with what looks like a group of small blue tumours growing out on its surface. Following mutagenesis, the authors screened for colonies that showed large numbers of blue papillae and discovered the mutator *mutY*, which showed a previously unknown mechanism of DNA repair.

**Transposition.** Transposons, such as *Tn10*, are genetic elements that can ‘jump’ (transpose) from one DNA molecule to another. Bender and Kleckner<sup>21</sup> made elegant use of sectoring on media that contained Xgal to show genetically that *Tn10* transposes by a mechanism that does not involve semiconservative DNA replication. *Tn10* that contained a HETERODUPLEX *lacZ* gene (*lacZ* wild type/*lacZ* nonsense) was allowed to transpose. If transposition occurred in the absence of DNA replication, then cells would inherit heteroduplex

DNA, and subsequent growth of this cell on media that contained Xgal would produce a blue/white sectorized colony. If DNA replication occurred during transposition, then cells would inherit only one of the donor strands, either *lacZ* wild type or *lacZ* nonsense, resulting in entirely blue or entirely white colonies. The fact that sectoring was observed provides genetic evidence that *Tn10* transposition occurs by a cut-and-paste mechanism rather than the replicative mechanism that is used by other elements, such as *Tn3*.

### LacZ fusions: making *lac* generally useful

LacZ fusions allow the tools of the Lac system to be applied to most genes, and have greatly expanded the range of biological processes that can be studied using the selection and screening methods described earlier. LacZ fusions come in two types (FIG. 5), which have important differences, and it is possible to construct both types using several methods<sup>22</sup>.

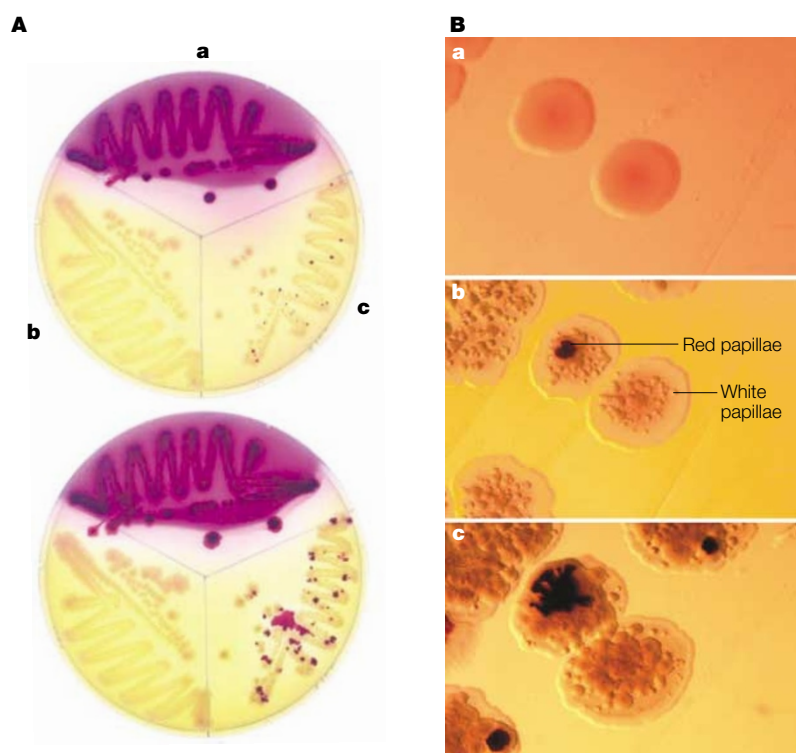
Transcriptional or operon fusions place the *lac* operon downstream from a different promoter. As such, these fusions are useful for monitoring transcription only. The mRNA that is produced by this type of fusion is wild type for *lacZ*, and all of the events that occur subsequent to transcription are the same as for wild-type *lacZ* mRNA. So, any post-transcriptional effects on *yfg* (‘your favourite gene’) expression — that is, the gene being studied — will not be reflected in LacZ production. We recommend designating transcriptional fusions as *yfg*’-*lacZ*<sup>+</sup>, in which the ‘prime’ symbol designates a deletion of the 3’ end of *yfg*, and the superscript ‘plus’ symbol indicates that the *lacZ* gene is wild type.

Translational or protein fusions are true gene fusions that result in the production of a hybrid protein. LacZ translational fusions produce a hybrid protein with amino-terminal sequences that are derived from *Yfg*, and have a large carboxy-terminal fragment of LacZ. In most cases, such fusions express an enzymatically active  $\beta$ -galactosidase molecule. So, in these cases, any post-transcriptional effects on *yfg* expression will be reflected in LacZ production. We recommend designating translational fusions as *yfg*’-*lacZ*. Note that ‘*lacZ*’ indicates a 5’ deletion of *lacZ* coding sequences, and the absence of the plus superscript denotes that the LacZ that is produced is not wild type: it is a hybrid protein.

As illustrated in the next section, these two types of LacZ fusion allow the use of all of the tools of Lac genetics for the study of *Yfg*. Specifically, as the later examples illustrate, it is possible to gain insights into the level and mechanism of gene regulation by comparing transcriptional and translational fusions.

**Regulation.** The cellular level of any given protein is determined by its rate of synthesis (transcription and translation) and its stability (degradation). Using the different types of *lacZ* fusion, each of these processes can be analysed.

Transcriptional control is the most common form of regulation, and *lacZ* transcriptional fusions provide the tools to examine transcription from a particular promoter whilst keeping other levels of control constant.



**Figure 4 | Development of *Mal*<sup>+</sup> papillae on MacConkey maltose agar. A** | The agar plate was seeded with three bacterial strains that were incubated for two days (top) and three days (bottom): wild type (a), *mal* deletion (b), and *mal* point mutant (c). Notice the appearance of red papillae growing up, and on top of, the white colonies in the part of the plate in which the point mutant is seeded. **B** | The formation of white and red papillae on colonies of the *mal* point mutant from section A. The growth of the *mal* point mutant is inhibited by maltose. The white papillae that form are the result of secondary mutations that eliminate maltose transport. The red papillae are the result of reversion to wild type. The images are of the same colonies taken after one (a), two (b) and three (c) days.

For example, *yfg* might be regulated at many levels, but a *yfg'*-*lacZ*<sup>+</sup> fusion will reflect transcription only.

Constitutive mutations in a regulatory gene are useful for determining whether the regulatory protein functions as a repressor or an activator on the promoter of *yfg*. The absence of a repressor results in constitutive expression. By contrast, activators are made to work in a constitutive fashion by gain-of-function mutations. In most cases these two possibilities can easily be distinguished in diploid analysis (FIG. 6).

To isolate constitutive mutations, a non-inducing substrate is necessary. In the case of the *lac* operon, the chemists provided such substrates (for example, PG). They have not always been so successful in designing substrates for other systems. However, transcriptional fusions solve this problem. For any *yfg'*-*lacZ*<sup>+</sup> fusion,

lactose is a non-inducing substrate. This was first shown in the maltose REGULON. *mal'*-*lacZ*<sup>+</sup> fusions are phenotypically Lac<sup>-</sup> in the absence of the inducer maltose. Constitutive mutations in the regulatory gene *malT* allow the growth of this fusion strain on lactose minimal agar. Because the *malT*<sup>c</sup> mutations are dominant in diploid analysis, this analysis established that *MalT* is an activator<sup>23</sup>.

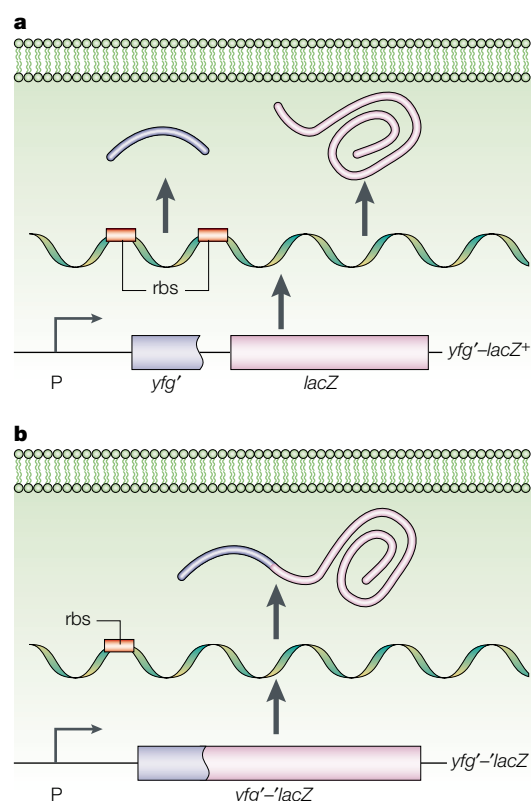
Similar logic can be used even if the regulatory gene is unknown. For example, this method was used to show that a stress-responsive two-component regulatory system that consists of the sensor histidine kinase *CpxA* and the response regulator *CpxR* regulates the gene for the periplasmic protease *DegP*. Dominant gain-of-function mutations that activate the *CpxA* kinase cause elevated expression of *degP'*-*lacZ*<sup>+</sup>, and confer a Lac-up phenotype that is easily detected on lactose MacConkey agar<sup>24</sup>. Such studies helped to show that *E. coli* has signal-transduction pathways that sense and respond to misfolded proteins in the cell envelope.

IS elements, similar to those present in *Tn10* (*IS10*), encode the transposase that allows these elements to move from one replicon to another. In the case of *IS10*, the expression of the transposase is regulated at the translational level by an antisense RNA. The antisense RNA hybridizes to the mRNA, thereby preventing ribosome binding at the translation initiation site. *LacZ* fusions were used to help establish this mechanism. Both transcriptional and translational fusions of *lacZ* to the gene for transposase were constructed. Expression of the antisense RNA reduced *lacZ* expression in the translational fusion more than tenfold. By contrast, expression of the antisense RNA had no effect on the expression of the transcriptional fusion<sup>25</sup>.

In recent years, it has become clear that SMALL RNAs have important roles in gene expression. Indeed, small RNAs were named 'Breakthrough of the Year' in 2002, by the journal *Science*<sup>26</sup>. Small RNAs are also common in bacteria. Indeed, *E. coli* might contain 50–100 of these molecules<sup>27</sup>. Many of these regulate gene expression by hybridizing with mRNAs, and transcriptional and translational *LacZ* fusions have been used extensively in their characterization.

Entry into STATIONARY PHASE in *E. coli* is accompanied by important changes in gene expression. *RpoS* is a  $\sigma$ -factor that mediates many of these changes. Starting with a transcriptional fusion to a gene that is repressed in stationary phase, a mutation that confers a Lac-up phenotype was identified after transposon mutagenesis (BOX 2) by screening on lactose MacConkey agar. This mutation identified a gene (*sprE* or *rssB*) that specifies a two-component response regulator. Interestingly, this mutation has no effect on the transcriptional *LacZ* fusion in log phase. The effect is only observed in stationary phase. Apparently this mutation was identified because many of the cells in a colony are in stationary phase.

To study the mechanism by which *SprE* regulates *RpoS*, three different *lacZ* fusions were used: one was an *rpoS'*-*lacZ*<sup>+</sup> transcriptional fusion, and the other two were *rpoS'*-*lacZ* translational fusions. The two translational fusions differed in the amount of *rpoS* sequence



**Figure 5 | Two types of *LacZ* fusion.** The target gene *yfg* ('your favourite gene') is transcribed from its promoter (P).

**a** | a 5' fragment of the target gene (*yfg'*) is fused to a wild-type *lacZ*<sup>+</sup> gene to make a transcriptional fusion. In this construct, the transcription of *lacZ*<sup>+</sup> is driven by the *yfg* promoter. The polycistronic mRNA that is produced is translated by ribosomes that bind independently to the ribosome-binding sites (rbs) that are located immediately upstream of each open reading frame to produce an amino-terminal fragment of *Yfg* and wild-type  $\beta$ -galactosidase. **b** | a 5' fragment of the target gene (*yfg'*) is fused in the correct translational reading frame to a large 3' fragment of *lacZ* (*lacZ*) to produce a translational fusion. Transcription of this hybrid gene is driven by the *yfg* promoter (P), and the mRNA that is produced is translated from the *yfg* rbs to produce a hybrid protein with amino-terminal sequences of *Yfg* and a large functional carboxy-terminal fragment of  $\beta$ -galactosidase.

#### REGULON

A group of transcriptional units or operons that are coordinately controlled by a regulator.

#### SMALL RNAs

Small untranslated RNAs that function in regulation.

#### STATIONARY PHASE

The phase of growth of a batch culture in which nutrients become limiting and growth ceases.

present: one (long) contained more *rpoS* sequence than the other (short). The results showed that *sprE* mutations had no effect on either the transcriptional or the short translational LacZ fusions. Only the long translational *rpoS'*-*lacZ* fusion was affected by the *sprE* mutations. Such results indicate that *sprE* does not affect *rpoS* transcription or translation, but, rather, that SprE operates post-translationally<sup>28,29</sup>. PULSE-CHASE EXPERIMENTS showed that SprE regulates the stability of RpoS and the long RpoS–LacZ hybrid protein<sup>28,30</sup>. The short RpoS–LacZ hybrid protein lacks a SprE recognition sequence<sup>31</sup>. Subsequent biochemical analysis shows that SprE delivers RpoS to the ATP-dependent ClpP/X protease for complete degradation<sup>32</sup>. These studies indicate that LacZ translational fusions, and the Lac phenotypes that they confer, can accurately report regulation by proteolysis.

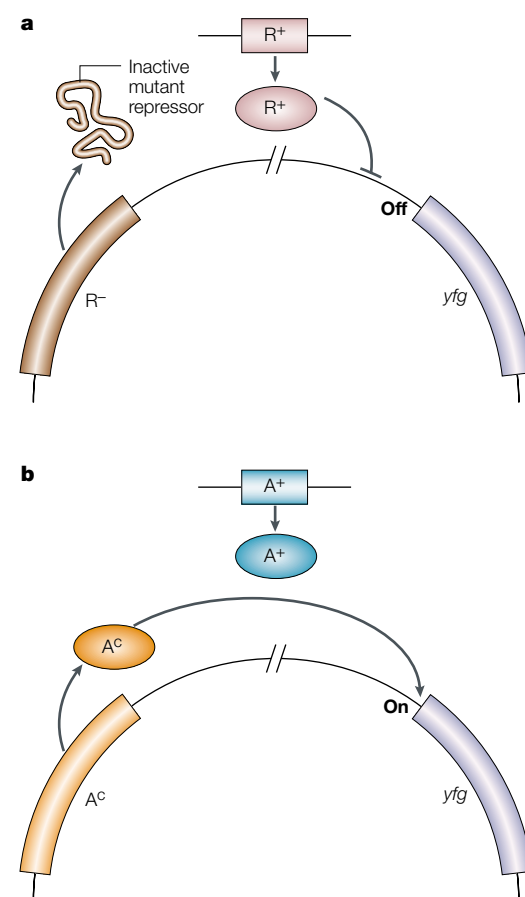
**Protein folding and secretion.** Genes that specify proteins that are destined for the periplasm or the outer membrane of *E. coli* have targeting signals at their 5' end. These proteins are made initially in precursor form with a SIGNAL SEQUENCE at the amino terminus. This sequence targets these precursors to the secretion machinery that is located in the inner membrane for translocation from the cytoplasm. The signal sequence is removed proteolytically during translocation. Consequently, *lacZ* translational fusions to these genes graft this targeting signal onto LacZ. The unexpected phenotypes that are conferred by these fusions opened the door for the genetic analysis of protein secretion (FIG. 7).

Initial studies of protein secretion in *E. coli* focused on genes in the maltose regulon that specify components of the maltose transport system. The *lamB* gene specifies an outer-membrane  $\beta$ -barrel protein that functions as a MALTOPORIN to allow substrates of the transport to cross the outer-membrane permeability barrier. The *malE* gene encodes a periplasmic maltose-binding protein that captures maltose in the periplasm and delivers it to the ABC transporter in the inner membrane. The ABC transporter is composed of two POLYTROPIC inner-membrane proteins, one of which is MalF. The ATPase MalK interacts with the cytoplasmic face of the inner-membrane proteins and drives the transport reaction<sup>33</sup>.

One of the new phenotypes that is conferred by *malE'*-*lacZ* or *lamB'*-*lacZ* translational fusions that contain a functional signal sequence is called lethal jamming (FIG. 7a). The signal sequence directs these proteins to the secretion machinery for translocation from the cytoplasm. However, LacZ sequences fold in the cytoplasm faster than they can be translocated. This plugs the secretion machinery, and the unfortunate constipated bacteria die. We know this because if we mutate the *lacZ* sequences so that folding is inhibited, the hybrid protein is secreted efficiently<sup>34</sup>.

Lethal jamming that is caused by these hybrid proteins is evidenced phenotypically as inducer sensitivity. Cells can tolerate low levels of the hybrid protein, but not the high levels that are produced by induction.

If mutants that are resistant to the inducer are selected for, the most common class fails to produce the hybrid protein. These are not interesting. However, if resistance to the inducer and Lac<sup>+</sup> are simultaneously demanded, by looking for blue colonies on maltose minimal agar that contains Xgal, then the only mutations that answer the selection are those that alter the signal sequence of the gene to which *lacZ* is fused<sup>35,36</sup> (FIG. 7A). These studies provided the first direct evidence to support the signal hypothesis of Blobel and Dobberstein<sup>37</sup>.



**Figure 6 | The use of diploid analysis to distinguish negative control by a repressor from positive control by an activator. a** | A portion of the chromosome that contains *yfg* ('your favourite gene') and an unlinked gene that specifies a repressor (*R*) is shown. Because of a mutation in the repressor gene (*R*<sup>-</sup>), an inactive mutant repressor is produced. Usually this mutant repressor does not function, and consequently, the *yfg* is on, even in the absence of inducer (constitutive). However, the introduction of a wild-type repressor gene *in trans* allows the production of a functional repressor (*R*<sup>+</sup>) that turns *yfg* expression off. This shows that the constitutive phenotype is caused by a recessive mutation (*R*<sup>-</sup>). **b** | A gain-of-function mutation in the gene for an activator (*A*<sup>-</sup>) turns on the expression of *yfg* even in the absence of inducer. The introduction of a wild-type activator gene *in trans* allows the production of wild-type activator (*A*<sup>+</sup>). However, in the absence of inducer, the wild-type activator is nonfunctional, and *yfg* expression is still activated by *A*<sup>-</sup>. This shows that the constitutive phenotype is caused by a dominant mutation (*A*<sup>-</sup>).

#### PULSE-CHASE EXPERIMENT

An experiment in which a radioactive form of a precursor is added for a brief period ('pulse') so that it is incorporated into a macromolecule, then a non-radioactive identical chemical is added to dilute the incorporation of the radioactive species ('chase').

#### SIGNAL SEQUENCE

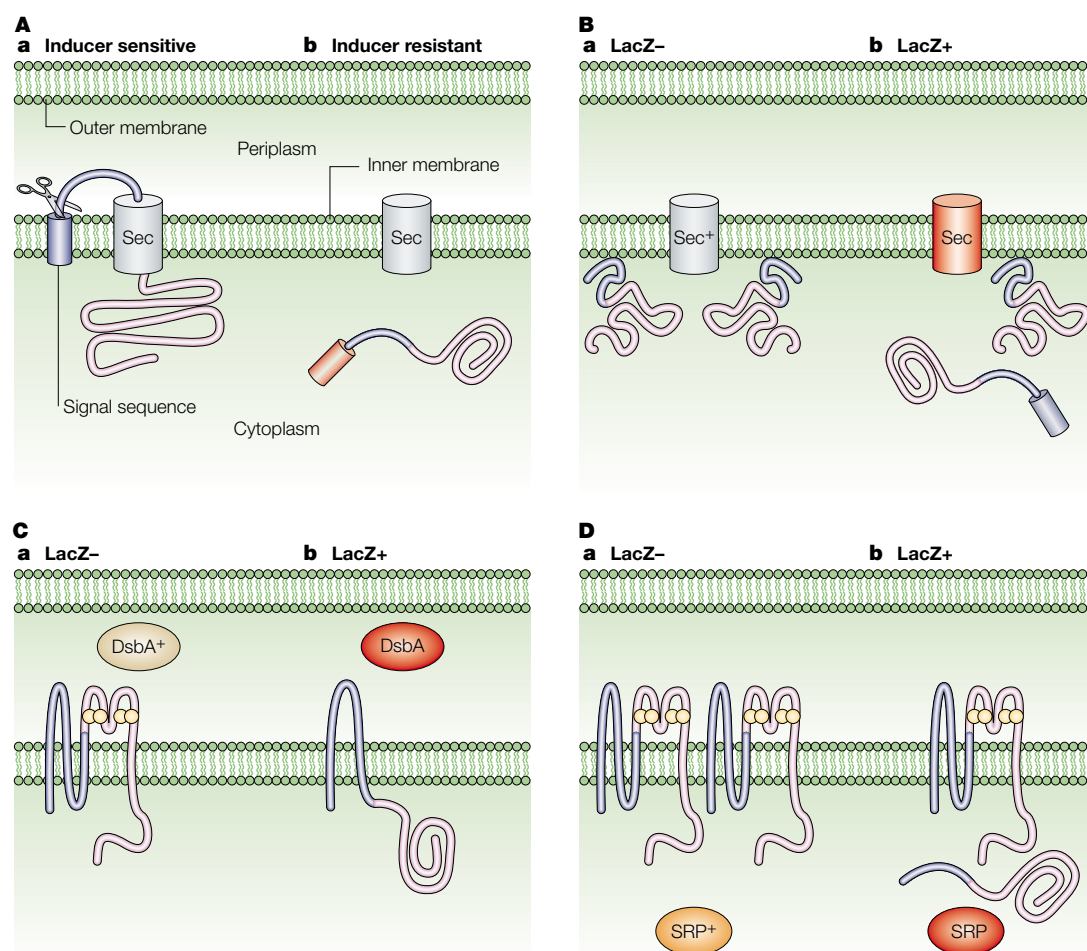
An amino-acid sequence at the amino terminus of a protein that directs the protein to the secretion system for translocation across the cell membrane.

#### MALTOPORIN

The *Escherichia coli* outer-membrane protein that functions as a diffusion channel for maltooligosaccharides.

#### POLYTROPIC

A membrane-spanning protein with more than one membrane-spanning segment.



**Figure 7 | The unexpected secretion-related phenotypes that are caused by *lacZ* translational fusions.** **Aa** | *LacZ*-hybrid proteins that have a functional signal sequence at the amino terminus are directed to the secretion machinery (Sec) in the cytoplasmic membrane for translocation from the cytoplasm. Amino-terminal sequences are translocated, and the signal sequence can be processed. However,  $\beta$ -galactosidase folds in the cytoplasm faster than it can be translocated and this results in a lethal jamming of the secretion machinery if hybrid-protein synthesis is induced to high levels. **Ab** | Signal-sequence mutations (mutant proteins are shown in red) relieve jamming because they prevent targeting of the hybrid protein to the Sec machinery. The  $\beta$ -galactosidase that accumulates in the cytoplasm folds normally and is fully active. **Ba** | Under non-inducing conditions, the low levels of *LacZ* hybrid protein that are produced are targeted to the cell envelope (designated here as hybrid protein that is associated with the membrane) in which  $\beta$ -galactosidase misfolds and is inactive (*LacZ*<sup>-</sup>). **Bb** | Mutations that compromise the Sec machinery interfere with hybrid-protein secretion, allowing a fraction of the molecules to remain in the cytoplasm where  $\beta$ -galactosidase folds properly and is fully active (*LacZ*<sup>+</sup>). **Ca** | Targeting of *LacZ*-hybrid proteins with transmembrane domains at the amino terminus can position *LacZ* sequences in the periplasm, in which they are anchored by disulphide bonds (double yellow circles). Under these circumstances,  $\beta$ -galactosidase misfolds and is inactive (*LacZ*<sup>-</sup>). **Cb** | Mutations that disrupt disulphide-bond formation (DsbA shown in red) allow  $\beta$ -galactosidase to slip back into the cytoplasm and fold properly (*LacZ*<sup>+</sup>). **Da** | Mutations that compromise the machinery that inserts proteins into the membrane (signal recognition particle (SRP) shown in red) interfere with the targeting of *LacZ*-hybrid proteins with transmembrane domains that allow a fraction of these molecules to remain in the cytoplasm, in which  $\beta$ -galactosidase folds properly (*LacZ*<sup>+</sup>).

The other new phenotype conferred by *malE'*-*lacZ* or *lamB'*-*lacZ* translational fusions relates to the amount of *LacZ* activity that is produced by cells in the absence of inducer (FIG. 7B). Under these conditions, the specific enzymatic activity of these hybrid proteins is abnormally low. This stems from the fact that *LacZ* is only functional in the cytoplasm, and under uninduced conditions in which jamming is not an issue; a significant fraction of the hybrid protein is directed from the cytoplasm. We know this because signal-sequence mutations markedly increase the specific activity of the hybrid proteins<sup>35,36,38</sup> (FIG. 7A).

As evidenced by the signal-sequence lesions, mutations that interfere with hybrid-protein secretion will increase the amount of hybrid protein in the cytoplasm, which results in a *Lac*-up phenotype. Such mutations can easily be identified as red papillae that grow out of the white lawns of the fusion strains on lactose tetrazolium agar. To avoid re-isolating signal-sequence mutations, *Lac* ups were selected at 30 °C, then mutant clones were screened for a temperature-sensitive (Ts) lethal phenotype at 42 °C. The reasoning was as follows. Leaky mutations that compromise the secretion machinery at 30 °C will cause some

## FLUOROPHORE

A small molecule, or a part of a larger molecule, that can be excited by light to emit fluorescence.

hybrid-protein accumulation in the cytoplasm and, therefore, a Lac-up phenotype (FIG. 7B). In some cases, at least, the secretion defect might be enhanced at elevated temperature, and as the secretion machinery is probably essential, this enhanced defect might cause temperature sensitivity. Although this strategy was implemented with fusion strains, the leaky Ts logic is commonly applied to identify mutations in other

types of essential gene. This strategy was first applied to identify the *secA* gene<sup>38</sup>, which we now know specifies the ATPase that drives protein secretion in *E. coli* and many other bacteria<sup>39</sup>.

Polytopic inner-membrane proteins, such as MalF, are characterized by stretches of hydrophobic amino acids that span the lipid bilayer in  $\alpha$ -helical conformation (transmembrane domains). It has long been thought that at least some of these hydrophobic sequences would function in a similar way to uncleavable signal sequences for membrane insertion<sup>40</sup>. To determine if special factors were required for inner-membrane protein targeting, *malF'*-*lacZ* fusions were used.

Certain *malF'*-*lacZ* fusions show low specific activities, and it was reasoned that this was caused by the targeting of LacZ sequences to the membrane (FIG. 7C). If true, it was thought that insights might be gained into the cellular factors that are required for membrane targeting by selecting Lac ups. Here again, defective targeting would allow cytoplasmic accumulation of the hybrid protein, and LacZ functions well in this environment. As the *malF'*-*lacZ* fusion strain does not grow on lactose minimal agar, Lac ups were selected initially by demanding growth on lactose.

Surprisingly, selection for Lac<sup>+</sup> using the *malF'*-*lacZ* fusion strain yielded mutations that disrupt disulphide-bond formation in the periplasm<sup>41</sup>. To explain this, it was proposed that the LacZ sequences of this hybrid protein were partially translocated to the periplasm, and in this environment the disulphide oxidase DsbA formed disulphide bonds between cysteine residues that are normally reduced in cytoplasmic LacZ. These disulphide bonds were proposed to 'lock' the partially translocated hybrid protein in place. Strains that lack DsbA formed disulphide bonds poorly. So, LacZ could slip back into the cytoplasm and regain activity (Lac<sup>+</sup>) (FIG. 7C). Although these studies did not yield the expected mutants, they launched a new and active research field<sup>42</sup>. The genes that were identified as a result of this work, and the subsequent follow-up, are involved in electron transfer between the cytosol and the periplasm. These genes, and their products, carry out previously unrecognized reactions that cells use to control the redox potential in different cellular compartments.

Reasoning that the Lac<sup>+</sup> selection with the *malF'*-*lacZ* fusion was too demanding, Tian *et al.*<sup>43</sup> revisited the original idea and devised a new and more sensitive screen. This was done by plating mutagenized cultures on maltose minimal agar (to induce the fusion) that was supplemented with Xgal to detect even low levels of LacZ activity. To further increase the sensitivity, plates were incubated for two days at 30 °C, then for several more days at 4 °C. As a result, among the faintly blue colonies, mutants were found that carried lesions in genes for components of the bacterial signal-recognition particle (SRP), a factor that is known to be required for the targeting of inner-membrane proteins (FIG. 7D).

## Box 3 | 'High-tech' genetics

One of the areas in microbiology that has benefited most from the huge amount of genome sequence information is bacterial pathogenesis. Many important human diseases are caused by bacteria that are less easy to study genetically than *Escherichia coli*. To capitalize on the availability of the genomic information, several groups have devised clever selection-based techniques that combine the power of genetics with large-scale methods. In many cases these approaches can be used with any type of selection, even those that are more complicated than the selections described for the *lac* system. Indeed, some of the methods use selections that are carried out in animals and are based on the ability of mutants to either survive or proliferate in animal hosts.

The first of these methods is referred to as 'genetic footprinting'<sup>49,50</sup>. In this method, large pools of transposon-induced mutants are subjected to selection in either cell culture or in animals, and pools of mutants that survive selection are recovered. The density of transposon insertions in a given location of the bacterial genome among the members of the pool can be determined by the analysis of PCR products using specific primers for the desired region. Transposon insertions that interfere with the ability of a given mutant to survive the selection will be lost from the pool. Comparison of the PCR products from the unselected and selected pools by gel electrophoresis allows the detection of those regions that are missing (hence the 'footprint'). Although this method can be used for any organism that can be mutagenized with a transposon and for which a genome sequence is available, it is laborious owing to the fact that only relatively small regions of the genome (<10 kb) can be examined in each experiment.

A second approach to detecting mutants that are lost during selection is called 'signature-tagged mutagenesis' or STM<sup>51</sup>. In this method, the transposon is modified by the insertion of a random 20-bp signature sequence so that each transposon-induced mutant carries a unique tag. Transposon-induced mutants are arrayed in 96-well format and each 96-member pool is used to infect an animal. At a particular time following infection, bacteria are harvested from the desired organ (for example, the spleen) and the surviving mutants are recovered on nutrient agar. PCR is used to amplify the transposon ends that contain the signature tags from the input and selected pools. The PCR products are then labelled with <sup>32</sup>P. Filters that contain the arrayed colonies used for infection are hybridized with <sup>32</sup>P-labelled probes that are derived from either the input pool or the pool of survivors. All of the signatures are present in the probe that is made from the input pool. By contrast, any signatures that are present in mutants that did not survive during the infection will be absent and will not hybridize to corresponding mutants on the filter that contains the arrayed colonies. So, by comparing the two filters it is possible to detect the specific mutants that do not survive in the animal. The location of the transposon in the original mutant can then be determined by standard cloning and sequencing methods. The power of this approach is limited by the fact that relatively small pools of mutants (~96) are examined in each experiment.

The third genome-based selection method is called transposon-site hybridization (TraSH)<sup>52</sup>. This method takes advantage of the availability of genome microarrays. In this method, a transposon is used to generate a large mutant library that is subjected to growth under two different conditions (for example, permissive and selective). Following growth under the different conditions, a combination of PCR and *in vitro* transcription is used to amplify sequences that are adjacent to the transposon and to label the products with a FLUOROPHORE. Different fluorophores are used to label the pooled sequences that are derived from the two different growth conditions. The fluorescent products are then used to probe a microarray that represents all of the open reading frames of an organism. From the relative intensities of the hybridization signals, it is possible to infer whether transposon mutations in a given location are under-represented in one of the growth conditions.

# Conclusion

We have presented several experimental approaches that take advantage of the simplicity and elegance of bacterial genetic systems. The *lac* operon has proved to be an experimental system of unparalleled versatility. Not only have the tools of the bacterial geneticist been used to understand the *lac* system, but these same simple tools can be applied to almost any biological system that can be studied in bacteria. We have made the point that visual screens and selections that are based on a mixture of knowledge and calculated guesswork can yield new insights into poorly understood phenomena.

Shuman and colleagues applied this principle to the study of a Gram-negative bacterial species that is metabolically different from *E. coli*. The Legionnaires disease bacterium *Legionella pneumophila* is able to kill many phagocytic cells, including human macrophages. To study the genetic basis of macrophage killing, it was easy to visually screen for transposon-induced mutants that had lost the ability to lyse a monolayer of macrophages on the surface of tissue-culture wells. The use of this simple visual screen facilitated the identification of ~60 mutants from a total of ~5,000 individual

colonies<sup>44</sup>. The subsequent analysis of these mutants by complementation, and eventual DNA sequencing, led to the discovery of 23 genes that encode a new Type IV secretion system, now known to be present in other pathogens and used to translocate virulence determinants to host cells. Others have used light production from bacterial colonies that express bacterial luciferase as an effective visual screen. The vast diversity of bacterial species and their endless metabolic variation leaves plenty of room for new unanticipated approaches to be developed.

We feel that there is merit in relying on simple experimental systems and focusing on biology. The availability of newer technologies, such as PCR, nucleotide sequencing and microarrays, enhances the simple methods, rather than replacing them. BOX 3 describes three ingenious approaches that combine powerful selections and modern 'high-tech' methods. These methods have already started to yield important new information about the biology of host-pathogen interactions. This is noteworthy for bacterial pathogens that are not as easily manipulated as *E. coli*. This combination of approaches will undoubtedly lead to rapid progress in uncovering new areas of prokaryotic biology.

1. Beadle, G. W. & Tatum, E. L. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl Acad. Sci. USA* **27**, 499–506 (1941).
2. Lederberg, J. Genetic recombination and linked segregation in *Escherichia coli*. *Genetics* **32**, 505–525 (1947).
3. Luria, S. E. & Delbruck, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
4. Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961).
5. De Lucia, P. & Cairns, J. Isolation of an *E. coli* strain with a mutation affecting DNA polymerase. *Nature* **224**, 1164–1166 (1969).
6. LaRossa, R. A. in *Escherichia coli and Salmonella* (ed. Neidhardt, F. C.) 2527–2587 (American Society for Microbiology, Washington, 1996).
7. Schreiber, S. L. The small-molecule approach to biology. *Chem. Eng. News* **81**, 51–61 (2003).
8. Rickenberg, H. V., Cohen, G. N., Buttin, G. & Monod, J. La galactoside-permease of *Escherichia coli*. *Ann. Inst. Pasteur* **91**, 829–857 (1956).
9. Cohn, M. in *The Operon* (eds Miller, J. H. & Reznikoff, W. S.) 1–9 (Cold Spring Harbor Laboratory Press, New York, 1978).
10. Malamy, M. H. in *The Lactose Operon* (eds Beckwith, J. R. & Zipser, D.) 359–373 (Cold Spring Harbor Laboratory Press, New York, 1970).
11. Muller-Hill, B., Crapo, L. & Gilbert, W. Mutants that make more *lac* repressor. *Proc. Natl Acad. Sci. USA* **59**, 1259–1264 (1968).
12. Silhavy, T. J., Berman, M. L. & Enquist, L. W. Experiments with gene fusions (Cold Spring Harbor Laboratory, New York, 1984).
13. Scaife, J. & Beckwith, J. R. Mutational alteration of the maximal level of *Lac* operon expression. *Cold Spring Harb. Symp. Quant. Biol.* **31**, 403–408 (1966).
14. Eschenlauer, A. C. & Reznikoff, W. S. *Escherichia coli* catabolite gene activator protein mutants defective in positive control of *lac* operon transcription. *J. Bacteriol.* **173**, 5024–5029 (1991).
15. Schwartz, D. & Beckwith, J. R. in *The Lactose Operon* (eds Beckwith, J. R. & Zipser, D.) 417–422 (Cold Spring Harbor Laboratory Press, New York, 1970).
16. Ullmann, A. & Perrin, D. in *The Lactose Operon* (eds Beckwith, J. R. & Zipser, D.) 143–172 (Cold Spring Harbor Laboratory Press, New York, 1970).
17. Langley, K. E., Villarejo, M. R., Fowler, A. V., Zamenhof, P. J. & Zabin, I. Molecular basis of  $\beta$ -galactosidase  $\alpha$ -complementation. *Proc. Natl Acad. Sci. USA* **72**, 1254–1257 (1975).
18. Beckwith, J. Restoration of operon activity by suppressors. *Biochim. Biophys. Acta* **76**, 162–164 (1963).
19. Gronenborn, B. & Messing, J. Methylation of single-stranded DNA *in vitro* introduces new restriction endonuclease cleavage sites. *Nature* **272**, 375–377 (1978).
20. Nghiem, Y., Cabrera, M., Cupples, C. G. & Miller, J. H. The *mutY* gene: a mutator locus in *Escherichia coli* that generates GC-TA transversions. *Proc. Natl Acad. Sci. USA* **85**, 2709–2713 (1988).
21. Bender, J. & Kleckner, N. Genetic evidence that *Tn10* transposes by a nonreplicative mechanism. *Cell* **45**, 801–815 (1986).
22. Hand, N. J. & Silhavy, T. J. A practical guide to the construction and use of *lac* fusions in *Escherichia coli*. *Methods Enzymol.* **326**, 11–35 (2000).
23. Debarbouille, M., Shuman, H. A., Silhavy, T. J. & Schwartz, M. Dominant constitutive mutations in *malT*, the positive regulator gene of the maltose regulon in *Escherichia coli*. *J. Mol. Biol.* **124**, 359–371 (1978).
24. Danese, P. N., Snyder, W. B., Cosma, C. L., Davis, L. J. & Silhavy, T. J. The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. *Genes Dev.* **9**, 387–398 (1995).
25. Simons, R. W. & Kleckner, N. Translational control of IS10 transposition. *Cell* **34**, 683–691 (1983).
26. Kennedy, D. Breakthrough of the year. *Science* **289**, 2283 (2002).
27. Gottesman, S. Stealth regulation: biological circuits with small RNA switches. *Genes Dev.* **16**, 2829–2842 (2002).
28. Pratt, L. A. & Silhavy, T. J. The response regulator SprE controls the stability of RpoS. *Proc. Natl Acad. Sci. USA* **93**, 2488–2492 (1996).
29. Pratt, L. A. *Osmotic and Growth-Phase Control of the Porin Regulon in Escherichia coli* (Princeton Univ. Press, Princeton, 1996).
30. Muffler, A., Fischer, D., Altuvia, S., Storz, G. & Hengge-Aronis, R. The response regulator RssB controls stability of the  $\sigma^{24}$  subunit of RNA polymerase in *Escherichia coli*. *EMBO J.* **15**, 1333–1339 (1996).
31. Becker, G., Klauck, E. & Hengge-Aronis, R. Regulation of RpoS proteolysis in *Escherichia coli*: the response regulator RssB is a recognition factor that interacts with the turnover element in RpoS. *Proc. Natl Acad. Sci. USA* **96**, 6439–6444 (1999).
32. Zhou, Y., Gottesman, S., Hoskins, J. R., Maurizi, M. R. & Wickner, S. The RssB response regulator directly targets  $\sigma^{24}$  for degradation by ClpXP. *Genes Dev.* **15**, 627–637 (2001).
33. Boos, W. & Shuman, H. Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiol. Mol. Biol. Rev.* **62**, 204–229 (1998).
34. Snyder, W. B. & Silhavy, T. J.  $\beta$ -galactosidase is inactivated by intermolecular disulfide bonds and is toxic when secreted to the periplasm of *Escherichia coli*. *J. Bacteriol.* **177**, 953–963 (1995).
35. Emr, S. D., Schwartz, M. & Silhavy, T. J. Mutations altering the cellular localization of the phage  $\lambda$ -receptor, an *Escherichia coli* outer membrane protein. *Proc. Natl Acad. Sci. USA* **75**, 5802–5806 (1978).
36. Bassford, P. & Beckwith, J. *Escherichia coli* mutants accumulating the precursor of a secreted protein in the cytoplasm. *Nature* **277**, 538–541 (1979).
37. Blobel, G. & Dobberstein, B. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* **67**, 835–851 (1975).
38. Oliver, D. B. & Beckwith, J. *E. coli* mutant pleiotropically defective in the export of secreted proteins. *Cell* **25**, 765–772 (1981).
39. Mori, H. & Ito, K. The Sec protein-translocation pathway. *Trends Microbiol.* **9**, 494–500 (2001).
40. Blobel, G. Intracellular protein topogenesis. *Proc. Natl Acad. Sci. USA* **77**, 1496–1500 (1980).
41. Bardwell, J. C., McGovern, K. & Beckwith, J. Identification of a protein required for disulfide bond formation *in vivo*. *Cell* **67**, 581–589 (1991).
42. Debarbieux, L. & Beckwith, J. Electron avenue: pathways of disulfide bond formation and isomerization. *Cell* **99**, 117–119 (1999).
43. Tian, H., Boyd, D. & Beckwith, J. A mutant hunt for defects in membrane protein assembly yields mutations affecting the bacterial signal recognition particle and Sec machinery. *Proc. Natl Acad. Sci. USA* **97**, 4730–4735 (2000).

**This paper describes the convoluted, but ultimately fruitful, use of genetic selection to identify mutations that affect protein secretion.**

44. Sadosky, A. B., Wiater, L. A. & Shuman, H. A. Identification of *Legionella pneumophila* genes required for growth within and killing of human macrophages. *Infect. Immun.* **61**, 5361–5373 (1993).
45. Miller, J. H. *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria* (Cold Spring Harbor Laboratory Press, New York, 1992).
46. Kleckner, N., Bender, J. & Gottesman, S. Uses of transposons with emphasis on Tn10. *Methods Enzymol.* **204**, 139–180 (1991).
47. Gibson, K. E. & Silhavy, T. J. The LysR homolog LrhA promotes RpoS degradation by modulating activity of the response regulator *sprE*. *J. Bacteriol.* **181**, 563–571 (1999).
48. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
49. Smith, V., Botstein, D. & Brown, P. O. Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc. Natl Acad. Sci. USA* **92**, 6479–6483 (1995).
50. Akerley, B. J. *et al.* Systematic identification of essential genes by *in vitro* mariner mutagenesis. *Proc. Natl Acad. Sci. USA* **95**, 8927–8932 (1998).
51. Hensel, M. *et al.* Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400–403 (1995).
52. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl Acad. Sci. USA* **98**, 12712–12717 (2001).

#### Acknowledgments

Work in our laboratories has been supported by the National Institutes of Allergy and Infectious Disease and the National Science Foundation (H.A.S.), and by the National Institute of General Medical Sciences (T.J.S.). Thanks to R. Freudenberg for help with the photographs of the agar plates and N. Ruiz for help with figures.

#### Online Links

##### DATABASES

**The following terms in this article are linked online to:**

**BioCyc:** <http://biocyc.org/ecocyc>

CpxA | CpxR | CRP | DegP | *degP* | DsbA | *galE* | LacA | *lacA* | *lacI* | *lacY* | *LacY* | *lacZ* | *LacZ* | *lamB* | *malE* | MalF | MalK | *malT* | *MalT* | *mutD* | *mutY* | *polA* | RpoS | *rpoS* | *rpsL* | *rssB* | *secA*

**Access to this interactive links box is free online.**