Introduction

While there are many approaches to developing new antibacterial drugs, all share a common property. Drugs must target processes that are critical for bacterial growth or survival during infection. Thus, the genes that encode these functions, so-called essential genes, represent the complement of possible antibiotic targets and identifying these genes and their encoded pathways can help lead drug discovery efforts. In the case of tuberculosis (TB), a disease for which current treatment is suboptimal, starting with target identification could lead to drugs with different mechanisms of action and, possibly, improved efficacy.

What is meant by “essential genes?” This term suggests genes that are absolutely required for bacterial survival. In fact, most of the studies described rely on bacterial growth to assay for essentiality. Thus, they actually define genes required for growth, not survival. Many of these approaches cannot distinguish between bacteria that grow very slowly and those that either die or cannot grow at all. Therefore, the function of identified genes can vary considerably depending on the method used for their identification. Of course, gene products are often only required under specific conditions. For example, in the case of Mycobacterium tuberculosis, several genes must be intact so that bacteria can survive during infection in a mouse but not for growth in defined medium. Others are required for growth under all assayed conditions, a group of genes that is far more technically difficult to study. Here we will review the strategies that have been employed to identify and validate essential genes in mycobacteria and look forward to possible future advances.

Choosing candidate essential genes

Most of the approaches described below are designed to study targeted genes. Several strategies for selecting which genes to target have been employed. Principal among these is an understanding of bacterial physiology and structure. For example, several genes are known to play important parts in cell division while others are critical for obtaining and metabolizing nutrients. On the other hand, gene products might be attractive drug targets because of what is known about them in other systems. For example, protein kinases are particularly druggable targets so it is valuable to understand their role in bacterial growth. Finally, screens, described below,
have identified a number of potential drug targets. Hasan et al.² have collated these results and assigned weighted scores for the potential of gene products to serve as targets for antibiotic development.

### Attempted knockouts

The first evidence for gene essentiality generally arises from negative data — the inability to knockout a targeted gene. Generating knockout mutants is usually the first step in determining the function of a gene. Several methods have been employed to produce such knockouts, including the use of non-replicating vectors,¹ long linear DNA fragments² and incompatible plasmids.³ Gene replacement can be performed using plasmids that require two recombination steps⁴ or as a single step using mycobacteriophage-mediated delivery.⁷ Unfortunately, the rate of illegitimate recombination is often high, particularly when attempting to introduce an unfavorable mutation. The introduction of mycobacteriophage proteins Che9c gp60 and gp61 — the homologs of RecE and RecT, respectively — can both increase the rate of true homologous recombination allowing the insertion of single nucleotide changes.⁸,⁹ A failure to obtain appropriate recombinants is often initially attributed to technique and later to biology. But how can a failed experiment be used to determine the probability of a gene actually being required for optimal growth?

The simplest method is to use homologous recombination rates to estimate the probability of gene essentiality.⁶,¹⁰⁻¹² This is performed as shown in figure 1A. In this strategy, homologous recombination is used to insert a vector sequence into the appropriate chromosomal location by a single crossover event. A second crossover event (resolution) can either restore the original structure of the chromosome or result in a deletion of the targeted gene. In the absence of selection, either second crossover event should occur at equal frequency. However, if the targeted gene is essential, the only allowable crossover is the one that preserves the gene. If there are a sufficiently large number of resolution reactions that fail to remove the gene, it is very likely that the gene is essential. This can be further controlled by adding a second copy of the targeted gene, either on an episome or integrated at a remote site, a process that can be facilitated with the use of vectors that are temperature sensitive for replication.¹⁴ If the resolution reactions occur only in the presence of the functional second copy, this strongly suggests gene essentiality and crossover frequency can be used to estimate the probability that a gene is truly essential.¹⁵

While simple, this approach is often not persuasive. The readout can be biased by the specific attributes of the constructs that are employed. For example, inserts might create toxicity independent of the effect on the targeted gene (e.g., by producing transcripational effects on adjacent regions of the chromosome). Of course, failure to disrupt a gene might be evidence of essentiality but does not provide any evidence for the functional role of the gene product.

### Using site-specific recombination system

One of the major limitations of using homologous recombination is the low rate at which it occurs. One way to overcome this is to use a higher frequency, site-specific recombination system. To do this, Pashley and Parish took advantage of the mycobacteriophage L5 excisionase¹⁶ (Fig. 1B). This enzyme efficiently removes DNA integrated into the ΦL5 attB site. These researchers constructed a merodiploid strain in which one copy of the gene was integrated at the L5 attB site then deleted the native copy of the gene. Introduction of another vector carrying the L5 excisionase gene will remove the integrated copy. For non-essential genes, transformation efficiency of excisionase carrying plasmid should be the same for all 3 strains. On the other hand, if the target gene is essential, no transformants will be obtained in strain B. Another modification uses the L5 integrase¹⁰,¹⁷,¹⁸ (Fig. 1C). This method employs the same recombination constructs; however, instead of the excisionase, another vector carrying L5 integrase is used to promote high frequency “switching” of introduced genes at the attB site. The rate of switching can be easily monitored using a different antibiotic marker. If the targeted gene is essential, no switching will occur in strain B.

These approaches have two advantages. The relatively high recombination rates make ambiguous results less likely. In addition, it is quite simple to integrate variants of the complementing gene (e.g., alleles from other species and mutant alleles) and test for complementation. Once again, however, results can be influenced by local effects at the site of the deletion of the native chromosomal gene. In addition, this method cannot be used to study the behavior of these genes during infection.

### Antisense oligonucleotides

Inhibiting gene expression using antisense has been a very successful strategy for identifying important genes in other pathogen such as *Staphylococcus aureus*.¹⁹ Since antisense RNA can be supplied in *trans*, this approach is very attractive as it does not require that site-directed mutations be introduced into the chromosome.

In fact, the Horwitz lab has taken this one step further.²⁰,²² As shown in a series of publications, they are able to partially silence genes using exogenous oligonucleotides. These phosphorothioate-modified nucleic acids are stable and can be taken up into target cells resulting in decreased synthesis of several targeted proteins. Although individual oligonucleotides had only limited effects, mixtures of different targeting antisense molecules could result in significant inhibition. They have used this method to show the potential of various proteins as targets for antibiotic development.

Despite the ease of this approach, it has not seen widespread application. Perhaps this is because it is not yet clear what rules govern successful design of these oligonucleotides. In fact, we do not yet know if these molecules target transcription, RNA stability or
translation. Thus far, oligonucleotide design has been largely empiric. In addition, the use of these tools is limited to in vitro growth conditions and they cannot be applied in vivo. However, as described below, antisense RNA produced by the targeted organism using conditional promoters might overcome some of these obstacles.

**Regulated promoters**

Studying genes required in vitro is particularly challenging as mutations that block their function do not result in viable colonies. An important strategy to avoid this problem is to use promoters that are only expressed conditionally. Practically, these include...
promoters that are regulated by ligands that are not ordinarily encountered by mycobacteria, allowing expression to be artificially regulated.

Thus far, three different regulated promoters have been used in mycobacteria. The acetamidase promoter, which can be induced by acetamide, is derived from Mycobacterium smegmatis. This promoter is most useful in M. smegmatis but has also been employed in M. tuberculosis. Several variations of tetracycline-regulated promoters have also been described. Most of these are induced using tetracycline and tetracycline analogs. However, a recent study reports a reverse promoter in which the addition of inducer results in inhibition of transcription. Each of these promoters has its own distinctive characteristics — the strength of the promoter, degree of inducibility and background level of transcription in the absence of the regulator — each of which determine how appropriate they are for studying individual genes. Notably, of the inducers described, only tetracycline and its analogs are known to have good bioavailability so that tetracycline-regulated promoters might be the sole class appropriate for studying genes during infection.

Regulated promoters have been used in two ways. First, these promoters provide an alternate method for generating antisense RNA. Instead of adding exogenous oligonucleotides, these promoters can be fused to antisense genes that are only expressed in the presence of inducer. As of now, there is only one published example of regulated antisense expression targeting the hisD gene. While other examples have been presented at meetings, this strategy has proven difficult to apply in a number of instances.

Others have replaced native promoters with regulated promoters, producing strains in which targeted genes are produced only in the presence (or, in the case of the reverse tetracycline promoter, in the absence) of inducer. There are now several examples where this strategy has been used successfully and this appears to be a very powerful method. Not only can this approach be used to validate gene essentiality but it also results in strains that, when depleted of the target, are useful for studying gene function. Unfortunately, not every gene can be approached using any given promoter probably because many genes need to be expressed at appropriate levels. In addition, in non-dividing cells, perhaps the majority of bacteria during chronic infection, rates of transcription are likely to be low and protein half-lives correspondingly extended. Thus, altering transcription might have very delayed effects.

Screens for essential genes

All of the above-described methods evaluate genes individually. Because many of these methods require the construction of strains with specific promoter fusions, it is quite difficult to perform them at large scale. Querying large numbers of genes requires methods that do not rely on homologous recombination events.

The simplest approach to constructing large numbers of tagged mutants is to use transposon mutagenesis. Although transposons insert randomly, they do so at sufficient frequency that the chromosome can be saturated with insertions. The methods that identify essential genes rely on the fact that such genes cannot tolerate insertions. Thus, these approaches rely on analyzing large numbers of diverse mutants, the types that can be generated using transposon mutagenesis.

Two methods have been used to detect genes that are not disrupted in a large pool of insertion mutants. Sequencing the insertion sites of individual mutants identifies disrupted genes unambiguously. Sequencing many mutants produces a compendium of insertion sites representing nonessential genes. Genes that are not disrupted are potentially essential. By sequencing a very large number of mutants and performing statistical analyses, Lamichhane et al. were able to identify likely essential genes and estimate the probability of essentiality. This approach has excellent resolution but, given the large number of sequences required, is difficult to reproduce under multiple conditions using current sequencing technology.

An alternative way to map insertions is to use DNA microarrays. We used a method we termed transposon site hybridization (TraSH) to generate probes from a pool of transposon mutants that would hybridize to chromosomal regions adjacent to the site of transposon insertion on microarrays (Fig. 1C). This allowed us to determine the positions of pooled probes in the chromosome and rapidly identify genes that lacked insertions. Because the entire pool is analyzed simultaneously, we could repeat this experiment under several different conditions (growth on different defined media, in macrophages with different activation states and at various points during murine infection) to define the various sets of required genes. TraSH has much less resolution than sequencing and is more susceptible to false positive and negative results. However, it is far easier to perform multiple experiments. An alternative method, designer arrays for defined mutant analysis (DeADMan) similarly uses microarrays but, in this case, with known mutations instead of random insertions. DeADMan has been used to analyze the survival of large numbers of mutants simultaneously. This is a very useful method for studying conditionally essential genes but cannot be used to identify genes required for in vitro growth.

Conclusion

Several important technical advances have helped us to identify genes that are required for bacterial growth and that could serve as potential antibiotic targets. These have been very useful for studying the roles of genes during in vitro growth. However, current drugs, which target gene products that are essential in vitro, have proven to be less than optimal for tuberculosis treatment. To find genes that, when inhibited, could
result in more rapid and more complete clearance of infection will require robust tools that can be used in animal models. The first generation of these is now becoming available and, with refinements, might help considerably in the search for new anti-tubercular agents.

Acknowledgements

We would like to thank Sarah Fortune, Jeff Murry, and Simon Dillon for their helpful comments.

Funding: Bill & Melinda Gates Foundation, Taiwan Merit Scholarship NSC-095-SAF-1-564-610-TMS (J-R.W) and Heiser Grant of The New York Community Trust (J-R.W).

Competing Interests: None declared.

References