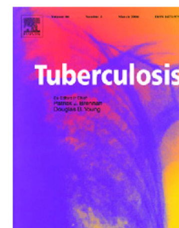
Tuberculosis (2008) **88** Suppl. 1, S19-S24

Available at www.sciencedirect.com

journal homepage: http://intl.elsevierhealth.com/journals/tube



The many roads to essential genes

Jun-Rong Wei, Eric J. Rubin*

Department of Immunology and Infectious Diseases, Harvard School of Public Health, 200 Longwood Ave, Boston, MA 02115, USA

KEYWORDS
Mycobacterium tuberculosis;
Essential genes

Summary
Antibiotics target functions that are required for bacterial growth and survival. As genetic tools for studying *Mycobacterium tuberculosis* continue to improve we are increasingly able to identify genes that encode these important effectors. Here we review the strategies that have been used to identify and validate essential genes in mycobacteria and look forward to possible future advances.
© 2008 Elsevier Ltd. All rights reserved

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.[1] 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 such essential genes in mycobacteria and indicate their advantages and disadvantages. These methods produce tradeoffs. In general, methods that analyze single genes have very low throughput but have far greater specificity. Screening methods can identify large numbers of candidates but are far more likely to make incorrect predictions.

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,

*Corresponding author: Tel: +1 617 432 3335;
Fax: +1 617 432 3354
E-mail address: erubin@hsph.harvard.edu (E.J. Rubin)

1472-9792/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.

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,[3] long linear DNA fragments[4] and incompatible plasmids.[5] Gene replacement can be performed using plasmids that require two recombination steps[6] 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.[6,10–13] 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 transcriptional 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 excisionase[16] (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 integrase[10,17,18] (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 pathogens 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

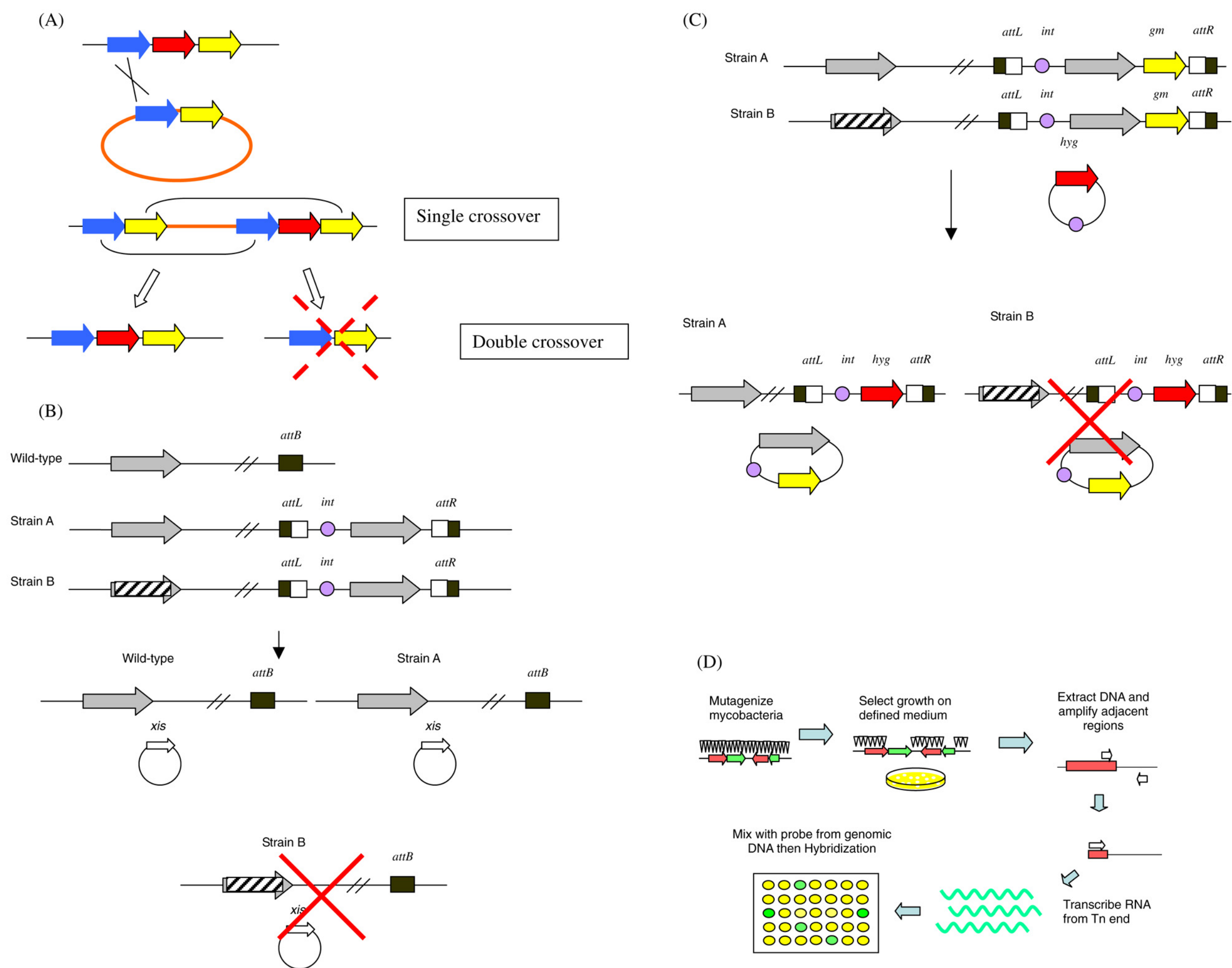


Figure 1 Methods used to define and study essential genes. (A) Testing for essentiality by measuring homologous recombination rates. A suicide vector (circular structure), which cannot replicate in mycobacteria, contains sequences (blue and yellow) flanking a targeted gene (red). A recombination event in the blue region results in integration of the plasmid into the chromosome. The integrated structure can be removed by a second recombination reaction, either in the blue region, resulting in reconstruction of the original chromosomal structure or in the yellow region, resulting in deletion of the targeted gene. If the targeted gene is essential, the only permissible second recombination event maintains the presence of the gene. However, in the presence of a second, complementing copy, both recombinations should occur with equal frequency. (B) Use of ΦL5 excisionase to determine gene essentiality. This figure is adapted from Parish et al.[16] A merodiploid strain (strain A) is constructed with a second copy of the targeted gene inserted into the ΦL5 *attB* site. Next, homologous recombination is used to disrupt the copy of the gene in the native chromosomal location (strain B). Introducing a plasmid that encodes the ΦL5 excisionase (xis) will remove the integrated copy of a non-essential gene. However, if the gene is essential, the ΦL5 excisionase plasmid will not be successfully transformed. (C) Vector/gene switching. Strains A and B are similar to those described in (B) with an added gentamicin resistant cassette shown in yellow (gm). Introduction of a plasmid containing a hygromycin resistance gene (hyg, in red) results in high frequency switching if there is no selective pressure. However, if the targeted gene is essential, switching will not occur. (D) TraSH (transposon site hybridization). A library of transposon mutants, each containing a single insertion, is grown under defined conditions. Only mutants that carry an insertion in non-essential genes grow. Insertion sites are mapped by hybridizing probes to adjacent chromosomal regions to a DNA microarray.

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*.[23] This promoter is most useful in *M. smegmatis* but has also been employed in *M. tuberculosis*.[24,25] The *nitA* promoter, from *Rhodococcus rhodochorus*, has been employed to study the role of protein kinases in *M. tuberculosis* using the inducer ε-caprolactam.[26] Several variations of tetracycline-regulated promoters have also been described.[27–30] 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.[31] 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 thus far, 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.[32] 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.[27] 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. Seqeuncing 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.[33] 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[34] (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,[35] in macrophages with different activation states[36] and at various points during murine infection[1]) 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.[37] 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.[38] 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-tuberculous 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-I-564-610-TMS (J-R.W) and Heiser Grant of The New York Community Trust (J-R.W).

Competing Interests: None declared.

References

1. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* 2003;**100**:12989–12994.
2. Hasan S, Daugelat S, Rao PS, Schreiber M. Prioritizing genomic drug targets in pathogens: application to *Mycobacterium tuberculosis*. *PLoS Comput Biol* 2006;**2**:e61.
3. Husson RN, James BE, Young RA. Gene replacement and expression of foreign DNA in *mycobacteria*. *J Bacteriol* 1990;**172**:519–524.
4. Balasubramanian V, Pavelka MS, Jr., Bardarov SS, Martin J, Weisbrod TR, McAdam RA, Bloom BR, Jacobs WR, Jr. Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. *J Bacteriol* 1996;**178**: 273–279.
5. Pashley CA, Parish T, McAdam RA, Duncan K, Stoker NG. Gene replacement in mycobacteria by using incompatible plasmids. *Appl Environ Microbiol* 2003;**69**:517–523.
6. Parish T, Stoker NG. *glnE* is an essential gene in *Mycobacterium tuberculosis*. *J Bacteriol* 2000;**182**:5715–5720.
7. Bardarov S, Bardarov Jr S Jr, Pavelka JM, Jr., Sambandamurthy V, Larsen M, Tufariello J, Chan J, Hatfull G, Jacobs JW, Jr. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in *Mycobacterium tuberculosis, M. bovis BCG* and *M. smegmatis*. *Microbiology* 2002;**148**:3007–3017.
8. van Kessel JC, Hatfull GF. Recombineering in *Mycobacterium tuberculosis*. *Nat Methods* 2007;**4**:147–152.
9. van Kessel JC, Hatfull GF. Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol Microbiol* 2008;**67**:1094–1107.
10. Parish T, Schaeffer M, Roberts G, Duncan K. HemZ is essential for heme biosynthesis in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 2005;**85**:197–204.
11. Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I. *ideR*, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* 2002;**70**:3371–3381.
12. Gomez M, Doukhan L, Nair G, Smith I. *sigA* is an essential gene in *Mycobacterium smegmatis*. *Mol Microbiol* 1998; **29**:617–628.
13. Gomez JE, Bishai WR. *whmD* is an essential mycobacterial gene required for proper septation and cell division. *Proc Natl Acad Sci U S A* 2000;**97**:8554–8559.
14. Zahrt TC, Deretic V. An essential two-component signal transduction system in *Mycobacterium tuberculosis*. *J Bacteriol* 2000;**182**:3832–3838.
15. Robertson D, Carroll P, Parish T. Rapid recombination screening to test gene essentiality demonstrates that *pyrH* is essential in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 2007;**87**:450–458.
16. Parish T, Lewis J, Stoker NG. Use of the mycobacteriophage L5 excisionase in *Mycobacterium tuberculosis* to demonstrate gene essentiality. *Tuberculosis (Edinb)* 2001; **81**:359–364.
17. Parish T, Roberts G, Laval F, Schaeffer M, Daffe M, Duncan K. Functional complementation of the essential gene *fabG1* of *Mycobacterium tuberculosis* by *Mycobacterium smegmatis fabG* but not *Escherichia coli fabG*. *J Bacteriol* 2007;**189**:3721–3728.
18. Pashley CA, Parish T. Efficient switching of mycobacteriophage L5-based integrating plasmids in *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* 2003;**229**: 211–215.
19. Ji Y, Zhang B, Van SF, Horn, Warren P, Woodnutt G, Burnham MK, Rosenberg M. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 2001;**293**:2266–2269.
20. Harth G, Zamecnik PC, Tang JY, Tabatadze D, Horwitz MA. Treatment of *Mycobacterium tuberculosis* with antisense oligonucleotides to glutamine synthetase mRNA inhibits glutamine synthetase activity, formation of the poly-L-glutamate/glutamine cell wall structure, and bacterial replication. *Proc Natl Acad Sci U S A* 2000;**97**:418–423.
21. Harth G, Horwitz MA, Tabatadze D, Zamecnik PC. Targeting the *Mycobacterium tuberculosis* 30/32-kDa mycolyl transferase complex as a therapeutic strategy against tuberculosis: Proof of principle by using antisense technology. *Proc Natl Acad Sci U S A* 2002;**99**:15614–15619.
22. Harth G, Zamecnik PC, Tabatadze D, Pierson K, Horwitz MA. Hairpin extensions enhance the efficacy of mycolyl transferase-specific antisense oligonucleotides targeting *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 2007;**104**:7199–204.
23. Parish T, Mahenthiralingam E, Draper P, Davis EO, Colston MJ. Regulation of the inducible acetamidase gene of *Mycobacterium smegmatis*. *Microbiology* 1997;**143 (Pt 7)**:2267–2276.
24. Manabe YC, Chen JM, Ko CG, Chen P, Bishai WR. Conditional sigma factor expression, using the inducible acetamidase promoter, reveals that the *Mycobacterium tuberculosis sigF* gene modulates expression of the 16-kilodalton alpha-crystallin homologue. *J Bacteriol* 1999; **181**:7629–7633.
25. Cho SH, Warit S, Wan B, Hwang CH, Pauli GF, Franzblau SG. Low-oxygen-recovery assay for high-throughput screening of compounds against nonreplicating *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2007;**51**:1380–1385.
26. Kang CM, Abbott DW, Park ST, Dascher CC, Cantley LC, Husson RN. The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. *Genes Dev* 2005;**19**:1692–1704.
27. Ehrt S, Guo XV, Hickey CM, Ryou M, Monteleone M, Riley LW, Schnappinger D. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res* 2005;**33**:e21.
28. Blokpoel MCJ, Murphy HN, O'Toole R, Wiles S, Runn ESC, Stewart GR, Young DB, Robertson BD. Tetracycline-inducible gene regulation in mycobacteria. *Nucl Acids Res* 2005;**33**:e22.

29. Carroll P, Muttucumaru DG, Parish T. Use of a tetracycline-inducible system for conditional expression in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Appl Environ Microbiol* 2005;**71**:3077–3084.
30. Gandotra S, Schnappinger D, Monteleone M, Hillen W, Ehrt S. *In vivo* gene silencing identifies the *Mycobacterium tuberculosis* proteasome as essential for the bacteria to persist in mice. *Nat Med* 2007;**13**:1515–1520.
31. Guo XV, Monteleone M, Klotzsche M, Kamionka A, Hillen W, Braunstein M, Ehrt S, Schnappinger D. Silencing *Mycobacterium smegmatis* by using tetracycline repressors. *J Bacteriol* 2007;**189**:4614–4623.
32. Parish T, Stoker NG. Development and use of a conditional antisense mutagenesis system in mycobacteria. *FEMS Microbiol Lett* 1997;**154**:151–157.
33. Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR. A post-genomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 2003; **100**:7213–7218.
34. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A* 2001;**98**:12712–12717.
35. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 2003;**48**:77–84.
36. Rengarajan J, Bloom BR, Rubin EJ. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* 2005;**102**:8327–8332.
37. Lamichhane G, Tyagi S, Bishai WR. Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. *Infect Immun* 2005;**73**:2533–2540.
38. Xie Z, Siddiqi N, Rubin EJ. Differential antibiotic susceptibilities of starved *Mycobacterium tuberculosis* isolates. *Antimicrob Agents Chemother* 2005;**49**:4778–4780.