Review

Medical Principles and Practice

Med Princ Pract 2004;13:177–184
DOI: 10.1159/000078312

Received: September 15, 2003
Revised: February 7, 2004

Defining the Mandate of Tuberculosis Research in a Postgenomic Era

Prachee Chakhaiyar Seyed E. Hasnain

Laboratory of Molecular and Cellular Biology, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India

Key Words
Tuberculosis · Mycobacterium · In silico

Abstract
The identification of *Mycobacterium tuberculosis* by Robert Koch in 1882 as the causative agent of tuberculosis, the release of the drug rifampicin in 1970 and the sequencing of the *M. tuberculosis* genome in 1998 are three major events that have revolutionized tuberculosis research. In spite of these breakthroughs, the continued status of tuberculosis as the largest killer amongst infectious diseases is an issue of major concern. Although directly observed short course chemotherapy exists to treat the disease, the emergence of drug-resistant strains has severely threatened the efficacy of the treatment. The recent sequencing of the *M. tuberculosis* genome holds promise for the development of new vaccines and the design of new drugs. This is all the more possible when the information from the genome sequence is combined with proteomics and structural and functional genomics. Such an integrated approach has led to the birth of a new field of research christened 'postgenomics' that holds substantial promise for the identification of novel drug targets and the potential to aid the development of new chemotherapeutic compounds to treat tuberculosis. The challenge before the scientific community therefore lies in elucidation of the wealth of information provided by the genome sequence and its translation into the design of novel therapies for the disease. All the major developments in the field of tuberculosis research after the sequencing of the *M. tuberculosis* genome will be discussed in this review.

Introduction

The sequencing of genomes of various pathogenic microbes has ushered in a most exciting era of medicine. Recent studies have generated hopes that decoding the genome of microbial pathogens will assist in the design of successful intervention strategies that would provide the host an edge over the pathogen. The genome sequence of *Mycobacterium tuberculosis* has provided new dimensions to tuberculosis researchers around the world. Finding a new drug against tuberculosis (TB) will now be largely dependent on our ability to amalgamate postgenomic information with the newly developed fields of transcriptomics and proteomics. The availability of the mycobacterial genome sequence has set up a platform for provision of answers to questions such as the functioning of the organism as an integrated system and activity in conjunction with the host. There are state-of-the-art technologies that have attempted to provide answers to these questions. Though genome sequencing definitely reveals more than 'the tip of an iceberg', it still needs to be supplemented with additional information such as gene function, regulation and interactions. A combination of all of this knowledge holds promise for the development of new

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1011–7571/04/0134–0177$21.00/0

Accessible online at:
www.karger.com/mpp

Dr. Seyed E. Hasnain
Laboratory of Molecular and Cellular Biology
CDFD, Nacharam
Hyderabad, 500 076 (India)
Tel. +91 40 27155604-05, Fax +91 40 7155479, 7155610, E-Mail hasnain@cdfd.org.in

intervention regimes against TB. This review will attempt to address some of these issues in a holistic fashion.

Genome Analysis in silico: Potentials of Bioinformatics

The availability of the genome sequence of various organisms, prokaryotic as well as eukaryotic, and even different strains of the same species, as in the case of *M. tuberculosis*, *Helicobacter pylori* etc., has provided immense opportunity for bioinformaticists to play around with the enormous data generated. With the development of more efficient bioinformatic tools like Position-Specific Iterated Basic Local Alignment Search Tool (PSI BLAST), Cluster of Orthologous Groups (COG), many genes with unassigned functions are now being annotated quickly and possibly correctly [1]. The COG database (http://www.ncbi.nlm.nih.gov/COG/), which is based upon phylogenetic classification of proteins encoded in complete genomes, has been used to reannotate the genomes of various organisms including *M. tuberculosis*. Data mining has also been used for accurate prediction of protein function [2]. When the *M. tuberculosis* genome was published [3], around 60% of the genes were labeled as 'unknowns' or 'hypothetical'. Today, thanks to new bioinformatic and experimental approaches, this figure has been reduced to about 48%. In a major effort to reannotate the genome, 22 new protein-coding genes, most of which are shorter than the 100-codon cutoff used initially, have been identified [4]. However, there are still about 376 proteins that do not have any homologues and are possibly unique to *M. tuberculosis*.

The bioinformatic approach has also helped in the functional classification of protein families. Computational methods for classification of cyclic nucleotide (cNMP)-binding protein and nucleotide cyclase superfamilies have shown that *M. tuberculosis* encodes several more putative cNMP-binding proteins than other prokaryotes; the functions of many of these are unknown [5].

Identification of secreted proteins of *M. tuberculosis* by in silico analysis for the presence of N-terminal signal peptide has also resulted in the identification of about 52 proteins, with experimental proof for about 17 of them [6]. Functional prediction of these secreted proteins would be important since it could also lead to the identification of proteins that could be important antigens. Regions of the genome that elicit a strong T cell response can also be exploited for development of a positive disease outcome. Associated studies include use of advanced and accurate programs that can predict T cell epitopes in the context of mycobacterial genome. Development of such programs along with in vitro methods for screening and confirming epitopes would increase the pace of development of new generation T cell epitope-based vaccines [7].

Bioinformatics has also aided in our understanding of the possible mechanisms of transcription termination in *M. tuberculosis*. Use of specific programs for deciphering hairpin formation at the 3′ end of mRNAs has suggested that the bacterium does not depend much on hairpin formation for transcription termination [8], hence rho-dependent transcription termination assumes obvious significance. Nevertheless, the probable hairpin forming regions have been evaluated using specific algorithms and experimental verification of the role of these sequences as transcription terminators suggests that there is a more or less relaxed requirement for the U trail in the mycobacterial genome [9]. While the in silico approach does help in prediction of gene function and in the shortlisting of genes for functional analysis, all in silico predictions require an in vivo verification and validation.

Comparative Genomics

With the advancement of sequencing technology, the number of organisms being sequenced is increasing at a rapid pace. Amongst the *M. tuberculosis* complex, *M. tuberculosis* was the first organism to be sequenced followed by *Mycobacterium leprae* and *Mycobacterium bovis*. Sequencing *of M. bovis BCG*, *Mycobacterium marinum* and the nonpathogenic bacterium *Mycobacterium smegmatis* are in various stages of completion and annotation. The availability of the genome sequence of related organisms has dramatically enhanced the study of comparative genomics [10]. Comparison at the genetic level provides clues to the possible loss or gain of virulence of one species as compared to another.

The genome of the leprosy bacillus and the massive deletions, pseudogenes etc. associated with it makes us ponder the reasons behind its reductive evolution [11]. Following suggestions that the leprosy bacillus is slowly heading towards its own death, its genome is being exploited to understand the complex implications of the loss of several important genes. It has been suggested that the loss of sigma factors could also lead to the accumulation of pseudogenes [12].

Chromosomal polymorphism between *M. tuberculosis*, *M. bovis*, and *M. bovis* BCG Pasteur have also been stud-

ied using the BAC library of all three [13]. The study detected a novel 12.7-kb segment present in *M. tuberculosis* but absent from *M. bovis* and *M. bovis* BCG Pasteur. Analysis of this segment showed that this region contains genes whose probable products show low similarity to proteins involved in polysaccharide biosynthesis.

Studies on comparative genomics of various mycobacterial species have revealed several deletions that exist in both virulent and avirulent species. Amongst these deletions, the most widely discussed is the RD1 region that has been deleted from *M. bovis* BCG and *Mycobacterium microti* [14]. Restoration of RD1 region in BCG by gene knock-in experiments has shown a clear transformation of the morphology to virulent type. Upon analysis of this deleted region, it was found that a few RD1 encoded proteins are localized in the cell wall while others code for immunodominant T cell antigens: ESAT-6 and CFP-10. Immunogenicity of the RD1 region has also been established using a synthetic peptide approach [15]. While immunocompetent mice could clear wild-type BCG, BCG RD1 knock-in showed increased persistence and partial reversal of attenuation. Five other RD loci were also tested for their effect on the virulence of BCG. However, none of them showed a response similar to RD1 knock-ins. A recent report has shown that recombinant BCG exporting ESAT-6 confers enhanced protection against TB in mice [16]. In a reverse approach, knocking out the RD1 region from *M. tuberculosis* led to reduced ability of the resultant bacterium to infect human and mouse macrophage cell lines [17]. These studies point towards the possibility of development of novel recombinant BCG vaccines.

Comparative genomic analysis of *M. tuberculosis* and *M. bovis* BCG has revealed that there are two major rearrangements in the genome of *M. bovis* BCG Pasteur [18]. These rearrangements correspond to deletion and duplication of certain regions. Though the significance of these regions in terms of attenuation of virulence or decreased immunogenicity of BCG is not known, it nonetheless reinforces the need to explore their function.

Epidemiological studies for identification of strain-specific signatures also form an important part of comparative genomics as they guide the understanding of evolution, and to an extent the pathogenesis, of the organism [19–21]. Our laboratory has also taken an initiative to assign specific signatures to strains from various geographic regions of the world utilizing different techniques like fluorescent amplified fragment length polymorphism, MIRU VNTR, spoligotyping, IS1081/6110 typing, etc. [22, 23]. The availability of the genome sequence has additionally helped us fine map the unique signatures associated with virulent strains. One of our studies also suggests that *M. tuberculosis* strains from HIV-seropositive and -seronegative patients are very distinct [24]. Availability of a database on strain-specific signatures would also help in identification of strains responsible for outbreaks [25].

Whole Genome Microarray: A Clue to Gene Expression under Different Conditions

Gene regulation in *M. tuberculosis* is tightly controlled by environmental signals and in this context microarray has become a very important tool for identification of genes expressed under a variety of growth conditions as well as for elucidation of the function of regulatory proteins by mutation analysis. Drug-induced alterations in gene expression have been studied by DNA microarray technology [26]. Isoniazid-induced genes have been studied using microarrays. It was found that isoniazid exposure leads to the expression of proteins physiologically relevant to the drug's mode of action, including an operonic cluster of five genes encoding type II fatty acid synthase enzymes and *fbp*C, which encodes trehalose dimycolyl transferase. Other genes like *efp*A, *fad*E23, *fad*E24, and *ahp*C, which are likely to mediate processes that are linked to the toxic consequences of the drug, were also expressed. Identification of these genes may define new drug targets.

Mycobacterium faces a microaerophilic environment within the host. Therefore, the hypoxic response genes become very important for the pathogen. Whole genome microarray studies identified >100 genes whose expression is rapidly altered by defined hypoxic conditions [27]. Many of the genes involved in aerobic metabolism were found to be repressed. A putative two-component response regulator pair Rv3133c/Rv3132c was also found to be induced in hypoxic conditions. Later studies showed that Rv3133c is a very important regulator of hypoxic response [28].

Microarray along with proteomics has been used to investigate the response of *M. tuberculosis* to nutrient starvation [29]. This has provided evidence for slowdown of the transcription apparatus, energy metabolism, lipid biosynthesis and cell division in addition to induction of the stringent response and of several other genes that may play a role in maintaining long-term survival within the host. This work has therefore generated a model with which one can search for agents active against persistent

M. tuberculosis and has revealed a number of potential targets expressed under these conditions.

The transcriptional response to low pH that mimics the environment to which the mycobacteria are exposed following phagocytosis has been studied by microarray technology [30]. Eighty-one genes were differentially expressed >1.5-fold, including many involved in fatty acid metabolism. The most highly induced genes showed homologies with those of nonribosomal peptide synthetases/polyketide synthetases.

Microarray-based information is available on regulatory proteins and several transcription factors of *M. tuberculosis*. The *M. tuberculosis* genome has 13 annotated sigma factors of which the ECF sigma factor, SigE, has been shown to contribute to pathogen survival following several distinct stresses [31]. An alternative sigma factor, SigH, has also been shown to play a role in mycobacterial stress response [32]. Survival of an *M. smegmatis sig*H *sig*E double mutant was found to be markedly decreased following 53°C heat shock and exposure to cumene hydroperoxide. The expression of the second gene in the *sig*H operon is required for complementation of the *sig*H stress phenotypes [33]. Role of these two sigma factors, SigH and SigE, in global gene expression profiles of *M. tuberculosis* has been studied using microarray technology. The SigH regulon and its role in global regulation of heat shock response and the thiol-specific oxidizing agent diamide have been defined using both quantitative RT-PCR and microarray technology [34]. In this study, 48 genes were identified whose expression increased after exposure of *M. tuberculosis* to diamide. Of these, 39 were not induced in the *sig*H mutant, suggesting their direct or indirect regulation by SigH. Many of these genes include thioredoxin, thioredoxin reductase and enzymes involved in cysteine and molybdopterine biosynthesis, which are related to thiol metabolism. SigH was also shown to be the regulator of other sigma factors such as SigB, SigE, and SigH itself. A related study has also shown that the *M. tuberculosis sig*H gene is dispensable for bacterial growth and survival within the host, but is required for the production of immunopathology and lethality [35].

The role of sigma factors in stress response suggests that these genes could be involved in pathogenicity. Microarray studies were carried out with *sig*E mutant strains and the SigE regulon was defined after confirmation with RT-PCR analysis [31]. It was shown that a functional *sig*E is required for full expression of *sig*B and for its induction after SDS exposure but not after heat shock. Genes no longer induced in the absence of *sig*E were also identified. These genes encode proteins belonging to different classes including transcription regulators, enzymes involved in fatty acid degradation and classical heat shock proteins. Apart from identifying the genes under the regulatory control of SigE, it was also shown that the *sig*E mutant is defective in its ability to grow inside both human and murine inactivated macrophages and is more sensitive, than the wild-type strain, to the killing activity of activated murine macrophages. Thus sigma factors appear to be attractive drug targets.

Another well-characterized global transcription regulator of *M. tuberculosis* is the iron-dependent repressor (IdeR) protein. Postgenomics, in conjunction with microarray technology, has defined the IdeR regulon [36]. These genes encode a variety of proteins, including putative transporters, proteins involved in siderophore synthesis and iron storage, members of the PE/PPE family, a membrane protein involved in virulence, transcription regulators, and enzymes involved in lipid metabolism. A PPE member, Rv2430c, that is overexpressed in an IdeR mutant has recently been characterized in our laboratory as an immunodominant antigen [37]. IdeR itself is an attractive drug target as IdeR deletion is lethal in *M. tuberculosis.* A high-resolution structure of IdeR has been solved [38].

The Phenomenon of Mycobacterial Latency

A major bottleneck in the development of a TB drug or vaccine has been the ability of *Mycobacterium* to enter a phase of dormancy inside the host cells. Postgenomic studies have shown that the glyoxylate shunt enzyme isocitrate lyase is very important for mycobacterial persistence in macrophages and mice [39]. Lipids are the chief carbon source during the dormant phase of infection and perhaps that is the reason behind the existence of a large number of proteins involved with lipid metabolism. The structure of isocitrate lyase has already been solved [40], but it has not been feasible to design an inhibitor for this target as the amount of inhibitor required to block this actively produced enzyme would be so high that it might even prove toxic to the host.

Mycobacterial latency has often been linked to hypoxic conditions within the host. However, response of *M. tuberculosis* to hypoxic conditions has remained poorly understood. Though microarray analysis has given considerable clues, a recent report represents another breakthrough in the understanding of the factors that control hypoxic response in *M. tuberculosis.* The transcription factor that largely regulates hypoxic response has been

identified [27, 28]. Verification of the role of this putative transcription factor (Rv3133c/DosR) was carried out by targeted disruption of the gene followed by analysis of the transcriptome of the mutant, as well as the wild-type bacterium. Induction of the hypoxia-responsive genes was not observed in the case of mutated *dos*R.

Conserved DosR binding site was also observed in the upstream sequences of all the genes under its regulatory control. The host response in the latent phase has also been studied in some detail. It has been shown that a CD8+ subset of T cells is active and produces IFN-γ only in the latent phase of infection [41].

Our understanding of the physiology of the bacterium in the latent phase has also been increased by a proteomics approach (identification of the highly up-regulated genes under hypoxic conditions) using techniques such as two-dimensional electrophoresis, and protein signature peptide analysis by liquid chromatography-mass spectrometry [42]. Targeting latent bacteria is a challenge. Using models depicting latency, the genome sequence has helped in the identification of some genes that allow the bacteria to control this phase. Targeting these genes would help in combating the latent phase of infection.

Genomics versus Proteomics: Answers Given by Proteomics

While the end of the 20th century saw the birth of genomics, the beginning of the 21st century is witnessing the birth and a rapid development of proteomics. Proteomics actually complements genomics in demonstrating which genes are really expressed and translated into a functional protein. Transcriptomics and proteomics are two extremely important tools for the identification of novel drug targets and vaccine candidates. These twin technologies are dramatically influencing the study of mycobacterial pathogenesis and virulence.

The power of proteomics suggests that subtle differences in bacterial culture conditions may have important implications for gene expression and protein production in mycobacteria. Study of nutrient starvation in *M. smegmatis* using a proteomics approach led to the identification of a nonspecific DNA binding protein, Dps, that has a role in DNA repair by nonspecific binding [43]. The *M. tuberculosis* homologue of Dps could be performing a similar role. Proteomics study of *M. tuberculosis* using two-dimensional electrophoresis and matrix-assisted laser desorption ionization and nano-electrospray mass spectrometry has also been used to identify 6 new proteins of *M. tuberculosis* not predicted by genomics methods [44]. A proteomics approach has even been used to identify the T cell antigens of *M. tuberculosis* [45]. With increased emphasis on structural proteomics, it is becoming important to quickly shortlist protein targets for structural studies that will appropriately take into consideration protein sequences, functions and genomic distribution [46]. Solving the three-dimensional structure of important proteins could lead to the design of novel inhibitors and hence the development of novel drugs.

Bioinformatics has also aided the field of proteomics in the context of development of methods for clustering protein sequences and building families of potentially orthologous sequences. In silico analysis of protein similarity and genetic neighborhood has also been used to understand functional relationships between proteins [47]. This study observed linkage between genes coding for membrane proteins and those involved in fatty acid metabolism. ESAT6-like transport proteins were also found to occur with the mce operon. The analysis could also indicate that very few transport proteins are present in the bacterium, which could be a reflection of its intracellular lifestyle.

Ensuring a Better Diagnosis for TB

The most popular diagnostic test for TB, the tuberculin skin test, has limitations on account of background obtained in BCG-vaccinated individuals. Development of new PCR-based methods for TB diagnosis is gaining increased popularity on account of their higher specificity and sensitivity [48]. Comparative genome sequence analysis has given some clue to the regions that are unique to *M. tuberculosis* and which are deleted in the case of BCG. These regions, which also constitute prominent T cell antigens like ESAT6, are being exploited to develop a specific antigen-based diagnostic assay for TB [49].

Novel Drugs and Vaccine Development: Is It a Near Possibility?

The genomics approach has led to the identification of genes/proteins important to the pathogen and is progressing fast towards development of a viable vaccine for TB. However, the challenge lies in selecting the most probable candidate immunogens or virulence genes from amongst approximately 4,000 genes of *Mycobacterium*. The availability of the genome sequence of pathogens in combina-

tion with a better understanding of protective pathogen-specific immune responses has laid the foundations for the development of new-generation recombinant vaccines. In the case of TB, recombinant BCG vaccines have gained much publicity owing to their observed protective efficacy in model infections. BCG vaccine (rBCG30) expressing and secreting the 30-kD major secretory protein of *M. tuberculosis,* has been shown to be more potent in increasing survival after challenge than parental BCG in the guinea pig model of pulmonary TB [50, 51]. Recombinant BCG expressing ESAT6 has also been shown to provide mild protection against TB infection [52]. In a very promising recent report, recombinant BCG bearing the entire RD1 deletion region along with the flanking region, which is predicted to contain the secretory signal, has shown an enhanced protection in mice [16]. This suggests that the loss of the ability of BCG to elicit a protective immune response is due to the loss of important 'chunks' of the genome and that restoration of the same indicates that there is hope for a better vaccine than BCG.

DNA vaccine for TB is in a likely possibility considering the report of hsp65-based DNA vaccine providing protection in mice against intravenous infection with TB [53]. However, in another study where the major antigens ESAT6 and Hsp60 were injected into mice as DNA vaccines, protection was observed only in the case of intravenous bacterial infection and not in the case of the more general aerosol mode of infection [54].

While development of a new vaccine for TB is the prime target of TB research, control of TB by chemotherapy is also the focus of many laboratories. The increasing information on the biology of *M. tuberculosis* and the availability of the genome sequence have provided a list of novel drug targets to kill the pathogen. Identification of new drug targets is important as most of the existing targets have been altered or mutated by the bacterium [55, 56]. Genomics has given us a brief idea about the complete sets of genes, enzymes and proteins possessed by the bacterium; subsequent research has already uncovered several leads for new drug targets.

Genomics has the potential to identify new drug targets as well as to understand why some drugs are without effect whereas others are fully active. The *M. tuberculosis* genome has been shown to possess a large number of putative drug efflux pumps, which might be playing an important role in imparting resistance to the bacterium. Work performed in our laboratory has shown that one of the putative drug efflux pumps is overexpressed in a drug-resistant clinical isolate of *M. tuberculosis* [57]. These pumps could themselves serve as drug targets. Other important potential drug targets in microbial genomes could be outer-membrane proteins, host interaction factors, permeases, enzymes of intermediary metabolism, systems for DNA replication, transcription, and repair, translation apparatus, etc. The probable role and essentiality of all these genes should be studied in detail to develop comprehensive 'multi-pronged' combinations of antibiotics that would fight bacterial pathogens effectively [58]. Virulence genes, uncharacterized essential genes, species-specific genes, unique enzymes and drug transporters are also a current focus of research for identification of new drug targets [59].

It is paradoxical that while there has been no dearth of identification of drug targets, no new drug has been introduced into the market even in the postgenomic era. However, the postgenomic era has suggested a few promising drugs for the disease such as the nitroimidazopyrans, a class of drugs that inhibit the biosynthesis of protein and cell wall lipids [60]. These drugs could target replicating as well as nondividing bacteria and have potent bactericidal activity against multidrug-resistant bacteria.

It is presumed that the application of functional genomic tools, such as microarray and proteomics, in combination with modern approaches, such as structure-based drug design and combinatorial chemistry, will lead to the development of new drugs that are not only active against drug-resistant TB but can also shorten the course of TB therapy.

Conclusion

Sequencing alone will not bring about a paradigm shift in medical research but a thorough understanding of its implications will definitely ensure that such a change will take place. The initial outcome of any genome sequencing project is just an unpunctuated string of 4 letters (A, T, G, C). Genomics involves compiling this information and postgenomics involves mining useful information from the genome. The genome approach has been applied to TB and the targets set by commercial companies and research institutions support optimism that a permanent cure for the disease is not a distant vision.

Acknowledgments

The research in Dr. Hasnain's laboratory cited in this review was supported by grants from the Department of Biotechnology and Council of Scientific and Industrial Research (CSIR), Government of India. P.C. thanks the CSIR for a Senior Research Fellowship.

References

1 Muller A, MacCallum RM, Sternberg MJ: Benchmarking PSI-BLAST in genome annotation. J Mol Biol 1999;293:1257–1271.
2 King RD, Karwath A, Clare A, Dehaspe L: Accurate prediction of protein functional class from sequence in the *Mycobacterium tuberculosis* and *Escherichia coli* genomes using data mining. Yeast 2000;17:283–293.
3 Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, Mclean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajani MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;11:537–544. Erratum in: Nature 1998;396:190.
4 Camus JC, Pryor MJ, Medigue C, Cole ST: Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. Microbiology 2002;148:2967–2973.
5 McCue LA, McDonough KA, Lawrence CE: Functional classification of cNMP-binding proteins and nucleotide cyclases with implications for novel regulatory pathways in *Mycobacterium tuberculosis*. Genome Res 2000;10: 204–219.
6 Gomez M, Johnson S, Gennaro ML: Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. Infect Immun 2000;68:2323–2327.
7 Sbai H, Mehta A, DeGroot AS: Use of T cell epitopes for vaccine development. Curr Drug Targets Infect Disord 2001;1:303–313.
8 Washio T, Sasayama J, Tomita M: Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. Nucleic Acids Res 1998; 26:5456–5463.
9 Unniraman S, Prakash R, Nagaraja V: Alternate paradigm for intrinsic transcription termination in eubacteria. J Biol Chem 2001;276: 41850–41855.
10 Brosch R, Gordon SV, Pym A, Eiglmeier K, Garnier T, Cole ST: Comparative genomics of the mycobacteria. Int J Med Microbiol 2000; 290:143–152.
11 Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG: Massive gene decay in the leprosy bacillus. Nature 2001;409:1007–1011.
12 Madan Babu M: Did the loss of sigma factors initiate pseudogene accumulation in *M. leprae*? Trends Microbiol 2003;11:59–61.
13 Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C, Barrell BG, Cole ST: Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect Immun 1998;66:2221–2229.
14 Pym AS, Brodin P, Brosch R, Huerre M, Cole ST: Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. Mol Microbiol 2002;46:709–717.
15 Mustafa AS, Cockle PJ, Shaban F, Hewinson RG, Vordermeier HM: Immunogenicity of *Mycobacterium tuberculosis* RD1 region gene products in infected cattle. Clin Exp Immunol 2002;130:37–42.
16 Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, Griffiths KE, Marchal G, Leclerc C, Cole ST: Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat Med 2003;9:533–539.
17 Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR: Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation. J Infect Dis 2003;187:117–123.
18 Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST: Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. Yeast 2000;17:111–123.
19 Tazi L, Kreiswirth B, Carriere C, Tibayrenc M: Molecular epidemiology of *Mycobacterium tuberculosis* and its relevance to the surveillance and control of TB: An e-debate. Infect Genet Evol 2002;2:153–158.
20 Hasnain SE: Molecular epidemiology of infectious diseases: A case for increased surveillance. Bull World Health Organ 2003;81:474.
21 Alland D, Whittam TS, Murray MB, Cave MD, Hazbon MH, Dix K, Kokoris M, Duesterhoeft A, Eisen JA, Fraser CM, Fleischmann RD: Modeling bacterial evolution with comparative-genome-based marker systems: Application to *Mycobacterium tuberculosis* evolution and pathogenesis. J Bacteriol 2003;185: 3392–3399.
22 Siddiqi N, Shamim M, Amin A, Chauhan DS, Das R, Srivastava K, Singh D, Sharma VD, Katoch VM, Sharma SK, Hanief M, Hasnain SE: Typing of drug resistant isolates of *Mycobacterium tuberculosis* from India using the IS6110 element reveals substantive polymorphism. Infect Genet Evol 2001;1:109–116.
23 Ahmed N, Alam M, Abdul Majeed A, Asad Rahman S, Cataldi A, Cousins D, Hasnain SE: Genome sequence based, comparative analysis of the fluorescent amplified fragment length polymorphisms (FAFLP) of tubercle bacilli from seals provides molecular evidence for a new species within the *Mycobacterium tuberculosis* complex. Infect Genet Evol 2003;2:193–199.
24 Ahmed N, Caviedes L, Alam M, Rao KR, Sangal V, Sheen P, Gilman RH, Hasnain SE: Distinctiveness of *Mycobacterium tuberculosis* genotypes from human immunodeficiency virus type 1-seropositive and -seronegative patients in Lima, Peru. J Clin Microbiol 2003;41:1712–1716.
25 Raman LA, Siddiqi N, Shamim M, Deb M, Mehta G, Hasnain SE: Molecular characterization of *Mycobacterium abscessus* strains isolated from a hospital outbreak. Emerg Infect Dis 2000;6:561–562.
26 Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, Brown PO, Schoolnik GK: Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc Natl Acad Sci USA 1999; 96:12833–12838.
27 Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK: Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. Proc Natl Acad Sci USA 2001;98:7534–7539.
28 Park HD, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, Schoolnik GK, Sherman DR: Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. Mol Microbiol 2003; 48:833–843.
29 Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K: Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Mol Microbiol 2002;43:717–731.
30 Fisher MA, Plikaytis BB, Shinnick TM: Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. J Bacteriol 2002;184:4025–4032.
31 Manganelli R, Voskuil MI, Schoolnik GK, Smith I: The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: Role in global gene expression and survival in macrophages. Mol Microbiol 2001;41:423–437.
32 Raman S, Song T, Puyang X, Bardarov S, Jacobs WR Jr, Husson RN: The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. J Bacteriol 2001;183: 6119–6125.
33 Fernandes ND, Wu QL, Kong D, Puyang X, Garg S, Husson RN: A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. J Bacteriol 1999;181:4266–4274.

34 Manganelli R, Voskuil MI, Schoolnik GK, Dubnau E, Gomez M, Smith I: Role of the extracytoplasmic-function sigma factor sigma(H) in *Mycobacterium tuberculosis* global gene expression. Mol Microbiol 2002;45:365–374.
35 Kaushal D, Schroeder BG, Tyagi S, Yoshimatsu T, Scott C, Ko C, Carpenter L, Mehrotra J, Manabe YC, Fleischmann RD, Bishai WR: Reduced immunopathology and mortality despite tissue persistence in a *Mycobacterium tuberculosis* mutant lacking alternative sigma factor, SigH. Proc Natl Acad Sci USA 2002;99:8330–8335.
36 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.
37 Choudhary RK, Mukhopadhyay S, Chakhaiyar P, Sharma N, Murthy KJR, Katoch VM, Hasnain SE: PPE antigen Rv2430c of *Mycobacterium tuberculosis* induces a strong B-cell response. Infect Immun 2003;71:6338–6343.
38 Feese MD, Ingason BP, Goranson-Siekierke J, Holmes RK, Hol WG: Crystal structure of the iron-dependent regulator from *Mycobacterium tuberculosis* at 2.0-A resolution reveals the Src homology domain 3-like fold and metal binding function of the third domain. J Biol Chem 2001;276:5959–5966.
39 McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan WT, Swenson D, Sacchettini JC, Jacobs WR Jr, Russell DG: Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature 2000;406:735–738.
40 Sharma V, Sharma S, Hoener zu Bentrup K, McKinney JD, Russell DG, Jacobs WR Jr, Sacchettini JC: Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. Nat Struct Biol 2000;7:663–668.
41 Van Pinxteren LA, Cassidy JP, Smedegaard BH, Agger EM, Andersen P: Control of latent *Mycobacterium tuberculosis* infection is dependent on CD8 T cells. Eur J Immunol 2000;30:3689–3698.
42 Rosenkrands I, Slayden RA, Crawford J, Aagaard C, Barry CE 3rd, Andersen P: Hypoxic response of *Mycobacterium tuberculosis* studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. J Bacteriol 2002;184:3485–3491.
43 Gupta S, Pandit SB, Srinivasan N, Chatterji D: Proteomics analysis of carbon-starved *Mycobacterium smegmatis*: Induction of Dps-like protein. Protein Eng 2002;15:503–512.
44 Jungblut PR, Muller EC, Mattow J, Kaufmann SH: Proteomics reveals open reading frames in *Mycobacterium tuberculosis* H37Rv not predicted by genomics. Infect Immun 2001;69:5905–5907.
45 Covert BA, Spencer JS, Orme IM, Belisle JT: The application of proteomics in defining the T cell antigens of *Mycobacterium tuberculosis*. Proteomics 2001;1:574–586.
46 Abascal F, Valencia A: Clustering of proximal sequence space for the identification of protein families. Bioinformatics 2002;18:908–921.
47 Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST: Analysis of the proteome of *Mycobacterium tuberculosis* in-silico. Tuber Lung Dis 1999;79:329–342.
48 Su WJ: Recent advances in the molecular diagnosis of tuberculosis. J Microbiol Immunol Infect 2002;35:209–214.
49 Gennaro ML: Immunologic diagnosis of tuberculosis. Clin Infect Dis 2000;30(suppl 3):S243–246.
50 Horwitz MA, Harth G, Dillon BJ, Maslesa-Galic' S: Recombinant bacillus Calmette-Guérin (BCG) vaccines expressing the *Mycobacterium tuberculosis* 30-kDa major secretory protein induce greater protective immunity against tuberculosis than conventional BCG vaccines in a highly susceptible animal model. Proc Natl Acad Sci USA 2000;97:13853–13858.
51 Horwitz MA, Harth G: A new vaccine against tuberculosis affords greater survival after challenge than the current vaccine in the guinea pig model of pulmonary tuberculosis. Infect Immun 2003;71:1672–1679.
52 Bao L, Chen W, Zhang H, Wang X: Virulence, immunogenicity, and protective efficacy of two recombinant *Mycobacterium bovis* bacillus Calmette-Guérin strains expressing the antigen ESAT-6 from *Mycobacterium tuberculosis*. Infect Immun 2003;71:1656–1661.
53 Lowrie DB, Tascon RE, Bonato VLD, Lima Valeria MF, Faccioli LH, Stavropoulos E, Colston MJ, Hewinson RG, Moelling K, Silva CL: Therapy of tuberculosis in mice by DNA vaccination. Nature 1999;400:269–271.
54 Taylor JL, Turner OC, Basaraba RJ, Belisle JT, Huygen K, Orme IM: Pulmonary necrosis resulting from DNA vaccination against tuberculosis. Infect Immun 2003;71:2192–2198.
55 Siddiqi N, Shamim M, Jain NK, Rattan A, Amin A, Katoch VM, Sharma SK, Hasnain SE: Molecular genetic analysis of multi-drug resistance in Indian isolates of *Mycobacterium tuberculosis*. Mem Inst Oswaldo Cruz 1998;93:589–594.
56 Siddiqi N, Shamim M, Hussain S, Choudhary RK, Ahmed N, Prachee C, Banerjee S, Savithri GR, Alam M, Pathak N, Amin A, Hanief M, Katoch VM, Sharma SK, Hasnain SE: Molecular characterization of multidrug-resistant isolates of *Mycobacterium tuberculosis* from patients in North India. Antimicrob Agents Chemother 2002;46:443–450.
57 Siddiqi N, Das R, Pathak N, Banerjee S, Ahmed N, Katoch VM, Hasnain SE: *Mycobacterium tuberculosis* isolate with a distinct genomic identity overexpresses a TAP like efflux pump. Infection 2004; in press.
58 Galperin MY, Koonin E: Searching for drug targets in microbial genomes. Curr Opin Biotechnol 1999;10:571–578.
59 Smith I: *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. Clin Microbiol Rev 2003;16:463–496.
60 Kendall SC, Warrener P, Vandevanter DR, Sherman DR, Arain TM, Langhorne MH, Anderson SW, Towell JA, Yuan Y, McMurray DN, Kreiswirth BN, Barry CE, Baker WR: A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. Nature 2000;405:962–966.