BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **253,** 70–79 (1998)
ARTICLE NO. RC989709

Comparison of Predicted and Observed Properties of Proteins Encoded in the Genome of *Mycobacterium Tuberculosis* H37Rv

Brooke L. Urquhart,[1] Stuart J. Cordwell, and Ian Humphery-Smith
Centre for Proteome Research and Gene-Product Mapping, The University of Sydney, National Innovation Centre, Australian Technology Park, Eveleigh, Australia, 1430

Received October 20, 1998

Proteome studies complement current molecular approaches through analysis of the actively translated portion of the genome (the "functional proteome"). Two-dimensional gel electrophoresis (2-DGE) utilising immobilized pH gradients of pH 2.3–5.0 and pH 6.0–11.0, developed with predetermined regions of overlap compatible with commercially available pH 4.0–7.0 gradients, permitted the display of a significant portion of the proteome of *Mycobacterium tuberculosis* H37Rv. A significant portion of the *M. tuberculosis* proteome, in the molecular mass (M_r) window 5 kDa to 200 kDa and with isoelectric point (p*I*) between pH 2.3 and 11.0, was visualised for the first time. A total of 493 protein spots were effectively resolved, including 126 spots that could not be seen using standard pH 4.0–7.0 gradients. These results were used to compare the physical properties of the observed proteins to the theoretical predictions of the recently completed *M. tuberculosis* H37Rv genome. Most proteins were found in the p*I* and mass window of pH 4.0–7.0 and 10–100 kDa. Analysis of the predicted proteome revealed a bimodal p*I* distribution, with substantial numbers of proteins in the p*I* regions 4.0–7.0 and 9.0–12.0 as has been seen for the majority of completed genomes. Such data may reveal current limitations in experimental extraction and separation of extremely basic, high M_r and hydrophobic proteins via 2-DGE. Conversely, 13 acidic proteins were observed with p*I* less than the lowest value predicted by the genome. In addition, a subset of small proteins (<10 kDa) were observed within the p*I* region of pH 5.0–8.0 that were not predicted by the complete genomic sequence, reflecting the current inability to distinguish small genes from within DNA sequence. This work represents the foundation for comparing the protein expression patterns of different pathogenic and nonpathogenic *M. tuberculosis* strains. The characterization of *M. tuberculosis* protein expression, further facilitated by the recent completion of the genome sequence, could aid in developing more effective diagnostic or therapeutic reagents. © 1998 Academic Press

Key Words: **functional genomics; immobilised pH gradient; *Mycobacterium tuberculosis;* proteome; two-dimensional gel electrophoresis.**

Tuberculosis remains a major health problem, both in industrialised countries and in the developing world. Currently one third of the human population is infected with *Mycobacterium tuberculosis*, resulting in 8 million cases of clinical tuberculosis each year and a reported 2.9 million deaths [1–3], a greater toll of human life than that caused by any other single pathogen. Multiple factors, acting independently and in combination, have contributed to the increase in tuberculosis infection world-wide. These include the concurrent epidemic of Human Immunodeficiency Syndrome infection and a lethal synergy with the AIDS virus [4–5], increased immigration from countries where tuberculosis is endemic, and primary transmission among populations such as the homeless, children, non-nationals and certain ethnic minorities. This epidemic is further compounded by the evolution of multi-drug-resistant strains of *M. tuberculosis* that have continued to weather the 'antibiotic revolution' [6–8]. The WHO estimates that the mortality rate due to tuberculosis infection will increase by 38.7% from 1990 to 2000, and those tuberculosis deaths related to HIV will rise by 331% [9]. The efficacy of the widely used vaccine strain *Mycobacterium bovis* (BCG) varies considerably, affording protection for approximately 50% of vaccinated subjects, with wider variability of 0-80% depend-

[1] To whom correspondence should be addressed. B.L. Urquhart, Centre for Proteome Research and Gene-Product Mapping, National Innovation Centre, Australian Technology Park, Eveleigh, Australia, 1430. Fax: 61-2-9319-1081. E-mail: b.urquhart@microbio.usyd.edu.au.

Abbreviations used: 2-DGE, two-dimensional gel electrophoresis; ADC, albumin dextrose catalase; BCG, Bacille Calmette—Guérin; IPG, Immobilised pH gradient; PGRS, polymorphic glycine-rich sequence.

0006-291X/98 $25.00

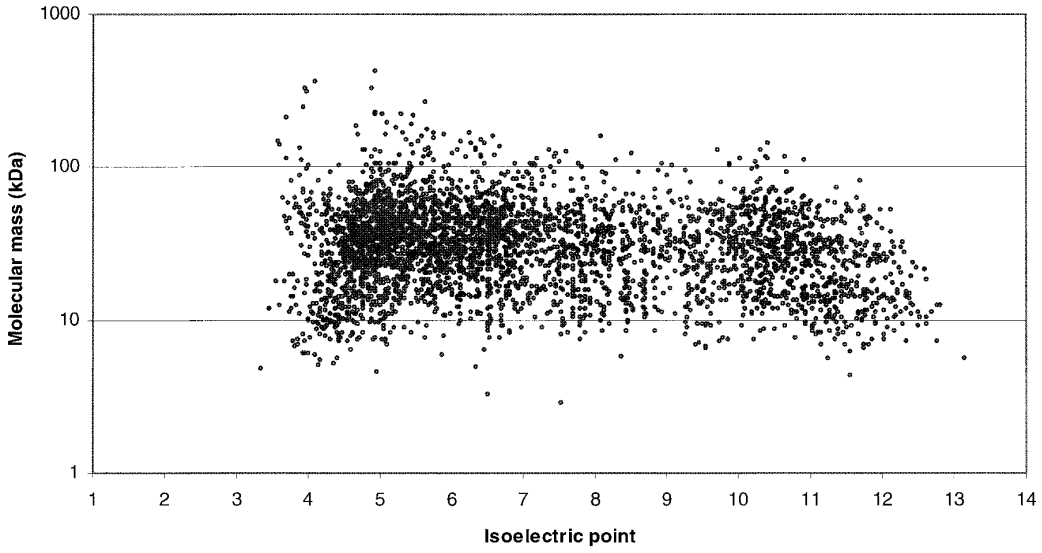

FIG. 1. Two-dimensional plot of calculated p*I* and M_r of the 3, 924 predicted proteins encoded by the complete *M. tuberculosis* genome sequence.

ing on the population studied, reinforcing the need for alternative methods of immunoprophylaxis against tuberculosis [10–12].

There is a crucial demand for answers to the many questions surrounding mycobacterial virulence, survival in macrophages and those proteins that stimulate

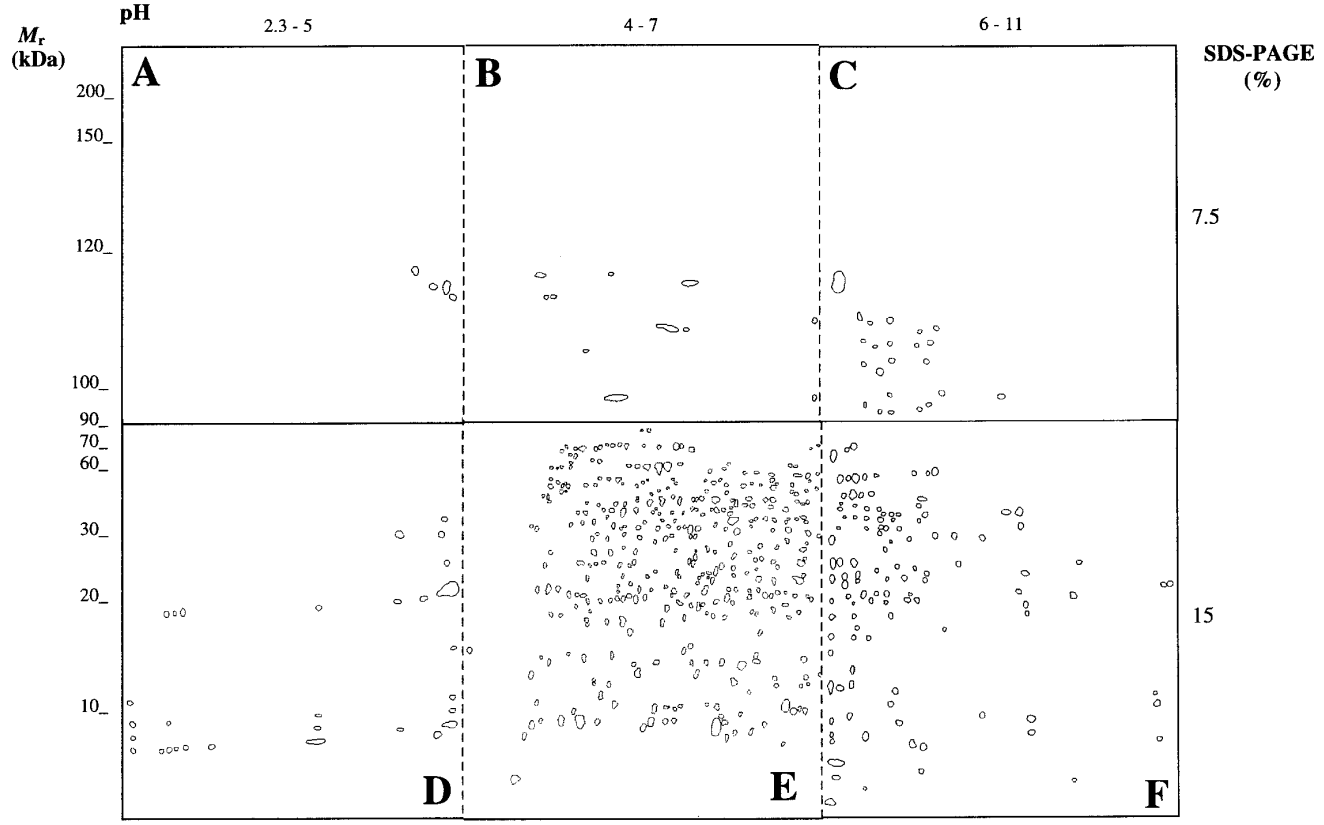

FIG. 2. Six computer generated protein 'windows' illustrating the silver-stained protein output of *M. tuberculosis* H37Rv at 37°C across the p*I* gradient pH 2.3–11. (A) pH 2.3–5.0, 7.5%T, (B) 4.0–7.0, 7.5%T, (C) 6.0–11.0, 7.5%T, (D) pH 2.3–5.0, 15%T, (E) 4.0–7.0, 15%T, (F) 6.0–11.0, 15%T.

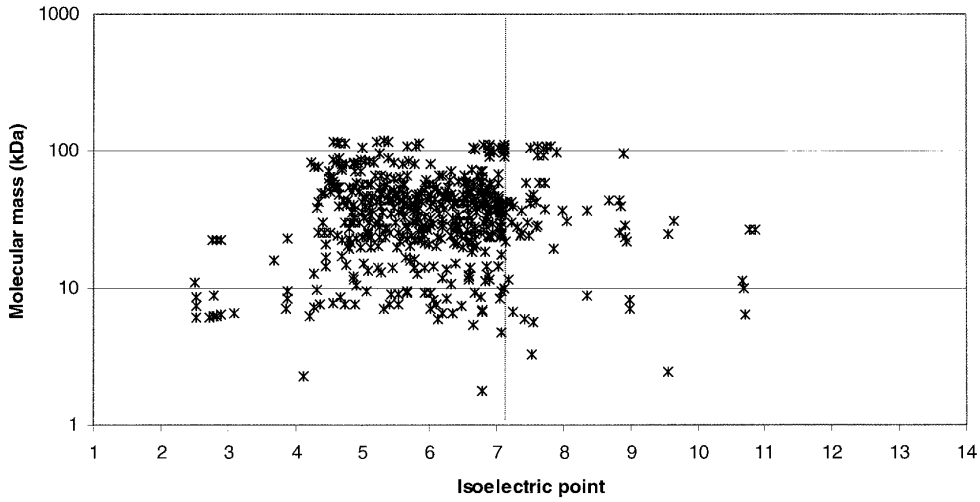

FIG. 3. Two-dimensional plot of the observed p*I* and M_r of the 493 *M. tuberculosis* proteins visualised within a series of silver-stained 2D gels. Dashed line represents p*I* value at which protein expression is greatly reduced.

protective immunity. Such answers will serve the development of anti-mycobacterial drugs, more effective vaccines and diagnostic reagents, since discerning between infection with *M. tuberculosis*, other non-pathogenic mycobacteria and prior immunisation with BCG continues to confound the medical community.

A variety of genetic methods have been employed to explore the mechanisms of mycobacterial virulence [13–16] most notably in the reconstruction of BCG as a new immunotherapeutic tool [17]. However, mycobacteria remain difficult to domesticate in the laboratory. Their slow generation time, resistance to many standard antimicrobial reagents, and thick lipid-rich cell wall inhibit the development of single colonies on growth media and hinder genetic manipulation. Furthermore, many of the prefabricated tools used in bacterial molecular biology are also unsuitable for use because of the requirement for *Mycobacterium*-specific promoters, origins of replication and signal sequences, confounded further by high G+C content DNA [17].

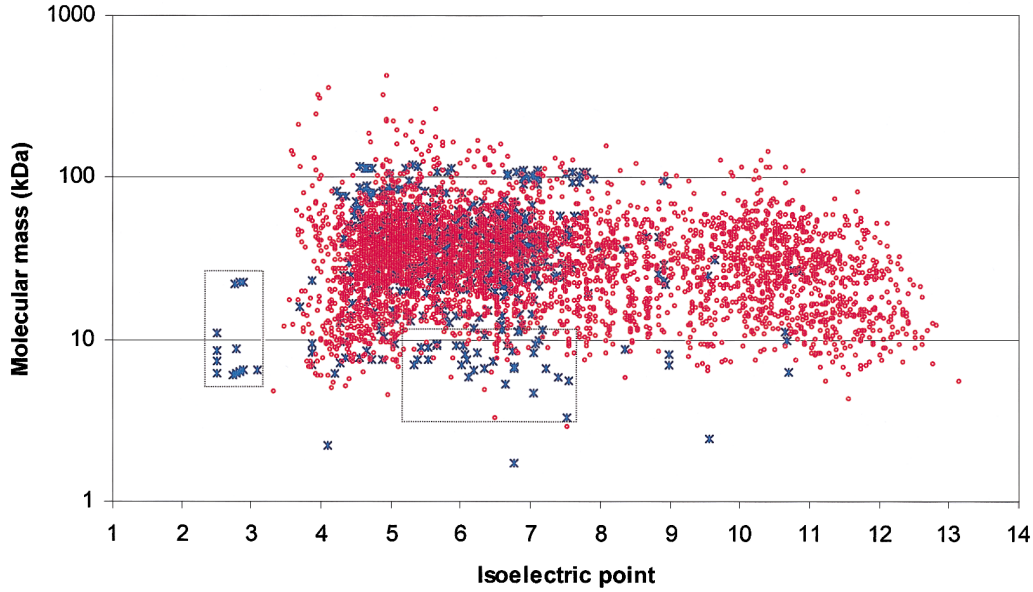

FIG. 4. Two-dimensional plot of the observed p*I* and M_r of 493 *M. tuberculosis* proteins visualised, and superimposed on the predicted values from 3, 924 conceptually translated ORFs from the complete *M. tuberculosis* genome. The blue crosses represent the p*I* and M_r of the silver-stained 2-DGE spots and the red circles represent the p*I* and M_r of the conceptually translated ORFs. Dashed boxes serve to highlight the presence of proteins of (i) extremely acidic p*I*, and (ii) $M_r \leq 10$ kDa, observed experimentally but not represented following predicted proteome plot.

TABLE 1

Published Analyses of *Mycobacterium* Species Using 2-D Gel Electrophoresis

Strain	Genotype	Growth medium	2-DGE Staining method	No. of proteins observed	Reference
M. bovis BCG Pasteur	Avirulent	Middlebrook 7H9/ADC	Silver nitrate	772	[50]
M. tuberculosis H37Rv	Wild type Virulent	GAS	Silver nitrate	205	[51]
M. bovis BCG Pasteur	Avirulent	Middlebrook 7H9/OADC	[^{35}S] methionine	≈170	[52]
M. tuberculosis Erdman	Wild type Virulent	+ 0.04% Tween 80 ”	”	”	
M. tuberculosis H37Rv	Wild type Virulent	Proskauer-Beck	Gold	≈130	[53]
M. tuberculosis Erdman	Wild type Virulent	Middlebrook 7H11; 7H9	[^{35}S] methionine	≈120	[54]
M. tuberculosis H37Rv	Wild type Virulent	Dubos	Coomassie brilliant blue, silver	≈75	[55]
M. bovis BCG–CSL (Commonwealth Serum Laboratories)	Avirulent	Dubos + albumin	^{125}I	≈50	[56]

Source. Modified from [50].
Note. GAS, glycerol-alanine-salts; OADC, oleic acid-albumin-dextrose-catalase.

The somewhat limited impact of genetics in analytical medicine applied to tuberculosis and other diseases [18] may be explained by the small percentage of diseases (2%) related to monogenic causality [19]. The one-gene-one-effect theory is no longer tenable. The dogma of DNA to RNA to protein to phenotype and disease rarely applies to the assembly of human disease states. Genetic analysis in itself cannot predict or diagnose multigenic diseases. Mycobacterial virulence is most certainly polygenic, and it is unlikely that a lone gene is solely responsible for the clinical manifestations of this disease. In combination with genetic elements, we must also consider the complex gene interactions and cellular events, which in an intricate epigenetic network influence and modify gene expression and/or post-translational modifications [19].

Total genomic sequence is currently available for a single unicellular eukaryote and 15 bacteria [20–35], and work has commenced on a further 61 genomes [36, 37]. More than a century after Robert Koch's postulates [38], the Wellcome Trust Pathogen Genome Unit at the Sanger Centre (Hinxton, Cambridge, UK) in collaboration with Cole and colleagues at the Pasteur Institute (Paris, France) have recently presented the complete genomic blueprint for *M. tuberculosis* [33, 39–40]. The genome consists of 4, 441, 529 nucleotide base pairs encoding 3, 924 genes. The sequence of every potential antigen and drug target is now available for effective vaccine construction. However, the quest to make biology finite has revealed little about genome structure and function. Approximately 45% of the genes revealed in the 15 other fully sequenced microbial genomes encode 'hypothetical' or 'unknown' proteins. Furthermore, the ability of *M. tuberculosis* to cause disease has not been ascribed to any well-defined virulence factors and the genetic information within must be fully deciphered before an attempt can be made to rewrite the course of this disease. Scientists at the Institute for Genome Research (TIGR; Gaithersburg, MD, USA) are currently working on other mycobacterial genomes, including a second isolate of *M. tuberculosis*. However, the major task following completion will be testing the predictions from the genomic sequence. Open Reading Frames (ORFs) must be authenticated. Alone, sequence is unable to predict: (i) if and when gene-products are translated, (ii) the relative concentrations of gene-products *in vivo,* (iii) the cellular or subcellular distribution of gene-products, (iv) the extent of post-translational modifications, (v) occurrence of small genes/ORFs of ≤300 bp left undetected following genomic annotation and (vi) the effects of gene 'knock-out' and over expression.

Proteome (the protein complement of a genome, [41]) studies are currently based on the production of two-dimensional electrophoresis gel 'maps', and the post-separational characterisation of the proteins detailed within. Proteomics complements whole genome sequencing by determining which genes are active at any given point in time. This has been termed the "functional proteome" [42], namely those proteins carrying out life processes within cells and tissues under varying physiological conditions. Large-scale proteome studies are currently in progress for the fully sequenced bacteria *E. coli* [43–46], *Haemophilus influenzae* [47], and the yeast *Saccharomyces cerevisiae* [48–49]. Previous reports detailing the application of 2-DGE to mycobacterial proteins are summarised in Table 1.

Although the majority of proteins are found close to biological pH and within a 30-50 kDa mass range, highly acidic (<pH 4.0), highly basic (>pH 9.0), low molecular mass (<10 kDa), and high molecular mass (>100 kDa) proteins cannot be simultaneously resolved on a single 2-D gel. Novel immobilised pH gradients of pH 2.3–5.0 and pH 6.0–11.0 used for isoelectric focussing in the first dimension were designed to maximise the number of protein spots resolved during proteome mapping in conjunction with 15%T and 7.5%T second dimension acrylamide concentrations. These pH gradients have been developed with predetermined regions of overlap in the pH zones 4.0–5.0 and 6.0–7.0 compatible with that of commercially available pH 4.0–7.0 IPG gel strips to permit cross-referencing between each gel image and their adjacent pH 'windows'. Recent reports detailing their use with *Spiroplasma melliferum* [42], *Ochrobactrum anthropi* [57] and *Mycobacterium bovis* BCG [50] vindicate the use of these novel pH gradients for the analysis of extreme pH regions of *M. tuberculosis* H37Rv. These results were subsequently utilised to compare the observed and theoretical properties of proteins encoded by the complete *M. tuberculosis* genome.

MATERIALS AND METHODS

Bacterial strains and media. *M. tuberculosis* H37Rv (ATCC 272940) cultures were maintained in Middlebrook 7H9 medium supplemented with albumin-dextrose-catalase (ADC; Difco Labs, Detroit, MI, USA). Batch cultures of *M. tuberculosis* were stored in 1, 2 and 5 mL aliquots at −80°C in 7H9/ADC broth.

Sample preparation. Whole cell lysates of *M. tuberculosis* were prepared by diluting a late-log phase culture 1:10 in 800 mL of Middlebrook 7H9/ADC broth. Bacteria were cultured at 37°C for 21 days with continuous stirring. Morphology and acid fastness were verified by Ziehl-Neelsen staining. Cells were harvested as previously described [50]. Design, construction and quality control of prefabricated IPG pH 2.3–5.0 and pH 6.0–11.0 gel strips are detailed elsewhere [57]. 2-DGE and silver-staining were performed as described in [50].

Image analysis. Silver-stained gels were scanned using a Molecular Dynamics (Sunnyvale, CA, USA) personal densitometer. The image was then transferred as an 8-bit gel image file into PHORETIX 2-D gel analysis software (NonLinear Dynamics, Newcastle-upon-Tyne, UK) where it was converted to an 8-bit .tif file. Multiple high quality silver-stained gel images were used to create a composite Phoretix reference map of *M. tuberculosis* based upon a representative gel image to which were added spots unique to the remaining gels in each series of p*I* and M_r 'windows'. PHORETIX software was used to automate the removal of streaks and gel-related artifacts, spot detection, spot measurement, and spot matching between image files. Spots were characterised on the basis of high circularity (perimeter/area, in pixels) and spot intensity. In addition, the centroid within each protein spot boundary was used to determine corresponding (*x, y*) coordinates. In total, 15 gels covering the entire gradient were analysed.

Determination of observed pI and M_r. Both M_r (GIBCO BRL, Gaithersburg, MD, USA) and p*I* (Bio-Rad, Hercules, CA, USA) markers were run independently and in tandem with *M. tuberculosis* samples to act as internal calibrants for determining the observed values. The M_r values were plotted versus "*y*" coordinates from the 2-D gels, and a logarithmic equation was fitted by linear least-squares regression. Using this equation, the "*y*" coordinates of each spot was converted to observed M_r. A similar procedure was used for calculating the observed p*I* except the predicted p*I* and "*x*" coordinates were fitted by linear least-squares regression to a linear equation.

Theoretical pI and M_r derived from genomic sequence. A list of ORFs, encoding predicted proteins, detected in the complete *M. tuberculosis* H37Rv genome sequence were obtained by ftp from ftp://sanger.ac.uk/pub.tb.sequences/ and analysed using the Genetics Computer Group (GCG) sequence analysis software package version 8.0 (GCG, University of Wisconsin Biotechnology Centre, Madison, WI, USA) to predict p*I* and M_r. The number of proteins were assembled in 5 kDa increments to examine the M_r distribution, while p*I* was assembled in 0.1 pH unit increments. M_r and p*I* distribution corresponding to experimental data was determined from the computer generated reference maps in similar increments.

RESULTS

Conceptual M. tuberculosis H37Rv Proteome

The p*I* and M_r of 3, 924 predicted proteins from the complete *M. tuberculosis* genomic sequence were calculated and plotted to show the distribution of the theoretical proteome on a 2-D gel (Figure 1). To simulate protein mobility during 2-DGE, the *x*-axis was drawn on a linear scale to represent protein mobility during isoelectric focusing along the linear pH gradients, while the *y*-axis was drawn on a logarithmic scale to represent migration during second-dimension SDS-PAGE. *M. tuberculosis* predicted proteins ranged from p*I* 3.33 (ORF number rv3652; PGRS-family protein fragment) to 13.15 (rv3924; 50S ribosomal protein L34) and molecular mass from 2.9 kDa (rv3599c; hypothetical 2.9 kDa protein) to 431.6 kDa (rv2048c; polyketide synthase). Following database comparisons, many of the 3, 924 proteins (~40%) have been attributed precise functions, while 44% show some degree of similarity, and the remaining 16% resemble no known proteins. Based on the conceptual translation of novel ORFs in the genomic sequence, it is currently unknown which of the hypothetical ORFs are even expressed *in vivo* and what role they may undertake. In addition, the predicted p*I*, and M_r do not take into account the effects of protein processing such as post-translational modifications.

Observed M. tuberculosis H37Rv Proteome

Six 'windows' of protein expression [50] were created with a horizontal overlap of one p*I* unit (Figure 2). A total of 541 protein spots were visualised for silver-stained gels of *M. tuberculosis* following 21 days growth at 37°C. When redundancy between the regions of p*I* overlap was taken into account, a total of 493 proteins spots were effectively resolved (Table 2). The positions of these spots with respect to p*I* and M_r is illustrated in Figure 3. Some loss of resolution was observed in the regions of overlap following the construction of 5 pH units within one 18cm IPG for the pH

TABLE 2

Number of Protein Spots Observed for *M. tuberculosis* H37Rv from Six 'Windows' of Protein Expression

Window of protein expression	Total number of spots	Number of nonredundant spots
Window A	4	1
Window B	11	11
Window C	23	20
Window D	30	17
Reference pH 4–7 Window[a]	356	356
Window F	117	88
Total	541	493

[a] Spots visualised in each 'window' in addition to those seen in the reference pH 4–7 Window E, i.e., excluding redundancy vertically and horizontally.

6.0–11.0 region, in contrast to 3 pH units for pH 4–7, identical in IPG strip length. This resulted in a 40% reduction in available IPG migration length in the basic region. Subsequently, proteins visualised in the pH region 6-7 may appear to comigrate in this condensed basic zone. However, approximately 88 and 18 novel protein spots respectively, were seen using pH 6.0–11.0 and pH 2.3–5.0 immobilised pH gradients. At this stage, the new pH 6.0–11.0 gradient is incapable of allowing the focusing of larger protein loads and some horizontal streaking in the first dimension of very basic proteins remains a problem [42]. Observed p*I* ranged from pH 2.5 to pH 10.9, and M_r ranged from approximately 1.7 kDa to 117 kDa.

Comparison of Conceptual and Experimental Proteomes

The distribution of the expressed and theoretical *M. tuberculosis* proteome was compared by plotting the p*I* and M_r of all experimentally observed proteins with the p*I* and M_r of the predicted proteins (Figure 4). In both instances, the abundant proteins were found to cluster in the p*I* and M_r range pH 4.0–7.0 (Figure 5a and 5b) and 10–100 kDa, respectively. However, for the predicted proteome, a bimodal p*I* distribution was observed, with substantial numbers of proteins in the additional p*I* region 9–12 (Figure 4). The near absence of proteins observed experimentally in this p*I* region suggests a need for specifically optimised protocols in association with current IPG technology for resolving basic proteins. Görg and colleagues have recently described construction of an immobilised pH 4–12 gradient, whereby horizontal streaking at the basic end, often described for linear IPGs up to pH 10, was not observed [58]. Furthermore, the use of a trichloroacetic acid/acetone extraction procedure for proteins in the region pH 9–12 resulted in ideal separation of very basic proteins otherwise not observed on the pH 4–12 gradient following extraction with classical 'lysis buffer' [50]. As no known sample preparation protocol allows ideal solubilisation of both extreme acidic and basic proteins, different solubilisation procedures may effect prefractionation of cellular extracts prior to electrophoretic separation, therefore reducing the complexity of these translation profiles. Conversely, there were 13 proteins observed in the acidic region of the gels, with p*I* less than 3.33, the lowest p*I* value predicted from the genome sequence. These may represent small, unannotated genes, proteins having undergone post-translational modification, or breakdown peptides from larger proteins either with unknown or without function. Further characterisation by mass-spectrometry and N-terminal sequencing may reveal their identity and putative function. Since only 493 (~13%) proteins were observed experimentally in comparison to almost 4,000 predicted from the genome sequence, this reflects a far from optimal appraisal of currently available separation and visualisation techniques. However, this figure is comparable to the percentages visualised using 2-DGE for *E.coli* proteins [45–46] and is significantly greater than the percentages of the human genome that have been visualised [59–60]. The difference may be accounted for by those proteins for which the methods of extraction were not ideal, for example, membrane proteins, which account for 20-30% of all predicted ORFs, hydrophobic proteins, misinterpretations of the genomic sequence and low copy number molecules that cannot be visualized with currently available methods. In addition, post-translational modifications such as glycosylation, acetylation and/or phosphorylation facilitate the generation of a diversity of proteins from any one gene, resulting in variable p*I* and/or M_r values in comparison to those predicted from the DNA sequence.

The theoretical proteome of *M. tuberculosis* showed a decrease in the number of proteins in the range between pH 7.3–7.4 and more noticeably at pH 8.3–8.4 (Figure 1 and 5a). The cytoplasmic pH of most cells is between pH 7.2–7.4 [23], and it may be difficult to maintain protein structure, rigidity and solubility when p*I* is near cytoplasmic pH. This theory is supported by the fact that the observed proteome similarly has few proteins in this range. With 24 proteins in the p*I* interval 6.9–7.0, a substantial decrease to no more than 5 proteins can be seen for the following p*I* intervals from pH 7.0 through to pH 11.0 (Figure 5b). Similar results were recently observed for the p*I* of proteins from *M. bovis* BCG [50], *E. coli* [45–46] and *H. influenzae* [47].

Of considerable interest was the presence of a subset of proteins with observed $M_r \leq 10$ kDa and p*I* between pH 5.0–8.0 for which there were few corresponding proteins encoded by the genome (Figure 4). This could be the result of protein truncation during translation

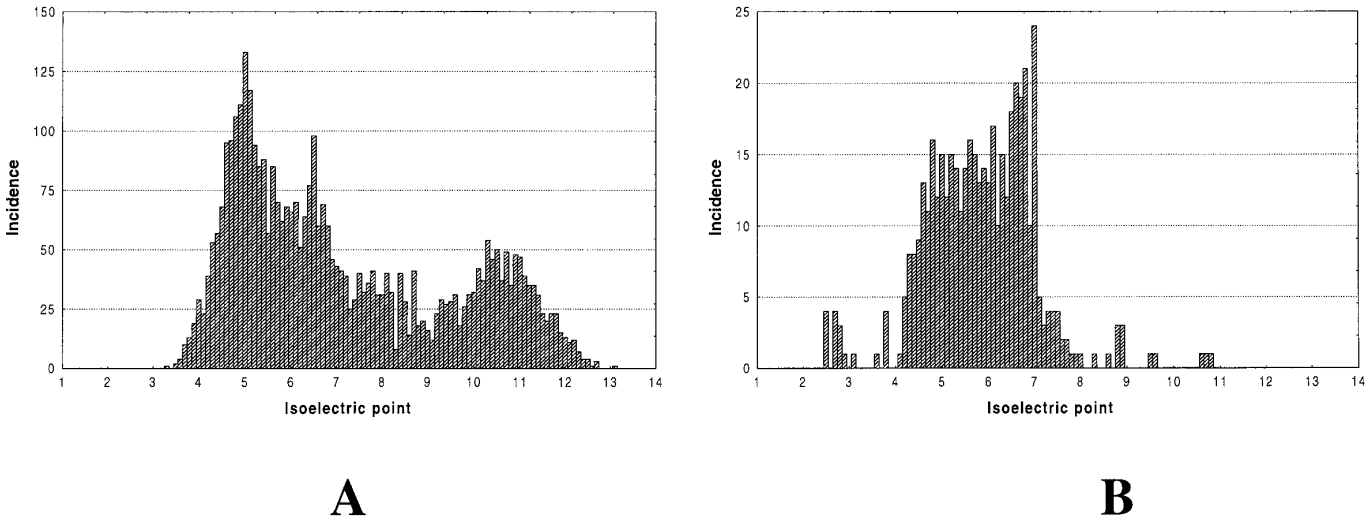

FIG. 5. (A) Distribution of proteins according to p*I* for the predicted *M. tuberculosis* proteome. (B) Distribution of proteins according to p*I* for the observed *M. tuberculosis* H37Rv proteome.

or misinterpretation of genomic sequence resulting in small genes that remain indistinguishable from random stretches of DNA without defined start/stop codons. These small M_r proteins will now be characterised by mass-spectrometry and N-terminal sequence determination for comparison with complete *M. tuberculosis* H37Rv sequence data. Other studies have reported similar findings, namely during the yeast genome project where researchers applied a minimum size cutoff for potential coding regions of 300 base pairs, or a protein of around 100 residues. While successfully excluding many artificial candidates from the list of potential proteins, it excluded at least 65 proteins (representing more than 1% of all yeast proteins) that had previously been biochemically or genetically characterised [61]. Similarly, characterisation of proteins with $M_r < 10$kDa in *E.coli* revealed that only 13 of 42 (31%) *N*-terminally sequenced proteins matched the sequence predicted from the original genomic annotation [62]. Many of these proteins were breakdown peptides of unknown function from larger proteins. However, such data also leads to the re-annotation of small gene sequences through large-scale genome research.

DISCUSSION

Initial efforts to resolve *Mycobacterium* whole cell lysates using commercially available pH 4.0–7.0 immobilised pH gradients and 2-DGE resulted in clustering of proteins in the acidic region and an inability to visualise basic proteins (data not shown). The construction of novel pH 2.3–5.0 and pH 6.0–11.0 gradients has resulted in the visualisation of a significant number of proteins in the extreme acidic and basic regions. Here, through the co-occurrence of protein spots within zones of overlapping pI and M_r present horizontally and vertically we were able to construct the most comprehensive documentation of the actively translated proteome of *M. tuberculosis* H37Rv to date. Upon comparison with previously published analyses, the authors have observed at least a 3-fold increase in the numbers of proteins observed (Table 2). Several authors have similarly demonstrated a relationship between increased gel size and improved resolution of protein spots [63–68]. However, upon comparison with the predicted *M. tuberculosis* proteome, the illustrated under-representation of extremely basic and high M_r proteins highlight current limitations with 2-DGE technology. There remains a need for specifically optimised protocols for those proteins for which the methods of extraction or separation were not ideal, for example, membrane-associated and hydrophobic proteins, or larger proteins which undergo adsorption to charged binding sites within the IPG, and are thus prevented entry into the second dimension [69–70]. Much of the proteome may also be present in the form of low copy number proteins [71–72], for which current detection methods are inadequate. Efforts to synthesise narrow range immobilised pH gradients in regions where increased resolution is required will also aid separation and characterisation of protein gene-products. The surprising expression of a number of small gene-products ($\leq$10 kDa) not mirrored by genomic predictions, may be explained by misinterpretations of the genomic sequence not accounting for those small genes ($<$100 codons) which remain difficult to distinguish during genomic annotation [73–74]. In addition, highly processed proteins resulting in variable pI and/or M_r values will migrate on the 2-D gel

contrary to sequence predictions. These results further reinforce the need to authenticate annotated ORFs.

A total of 493 proteins were observed over 6 proteome 'windows', a fraction of the total number predicted from the complete genomic sequence. As cells possess vast differences in solubility, M_r and p*I*, multiple protein extractions and gel conditions will be required to assure that the proteome is faithfully displayed. Proteome studies for the fully sequenced bacterium *Mycoplasma genitalium* [41] report visualisation of approximately 49% of 470 predicted proteins. However, bacteria possessing more complex genomes such as *E. coli* (4, 269 genes) and *H. influenzae* (1, 743 genes) report observed numbers of proteins in the order of 36% [45] and 22% [47], respectively. Unlike the somewhat static nature of the genome, an organism will have many proteomes dictated by environmental variability. Furthermore, as the evolutionary complexity of an organism increases, it appears that the percentage of total proteins expressed at a given instant are reduced [75]. This is since smaller genome organisms must maintain a greater percentage of proteins with "housekeeping" type functions, while higher organisms maintain complex systems specific for many different environmental challenges.

We now have access to complete genomic sequence for several micro-organisms. However, the DNA code must be supplemented by corresponding information at the level of transcription and translation to reveal the mechanisms of genome function. Significant amounts of data are already available for levels of mRNA transcripts within cells detected by microarray technology [76–77]. However, while such technology has found a niche in the era of functional genomics, recent reports [78] suggest that mRNA abundance does not necessarily correspond to cellular levels of protein expression. This is thought to be due to factors such as variable mRNA half-life. Therefore, ORF expression must be verified by identification of expressed proteins and assessment of protein sequence, versus that predicted by genomic annotation. Extending the p*I* and M_r range over which proteins can be visualised has been shown to significantly increase the potential of 2-D electrophoresis to resolve complex cellular extracts. The technology developed for this study in association with post-separational analysis techniques such as mass spectrometry, will help rapidly match sequence data with such maps of protein expression, thereby facilitating the identification of differences in protein expression between pathogenic and nonpathogenic *M. tuberculosis* strains under a variety of environmental conditions. Once identified, those genes/gene-products essential for virulence may then become targets for the production of new diagnostic and therapeutic reagents.

ACKNOWLEDGMENTS

The authors thank David Basseal for his assistance with Phoretix manipulations, Peter Maxwell for downloading and manipulating DNA sequence data, and Daniel Roach for initial cultures. BLU is the recipient of an Australian Postgraduate Award. This work was supported in part by funding from The University of Sydney, Glaxo-Wellcome Australia, and the National Health and Medical Research Council of Australia.

REFERENCES

1. Smith, P. G., Moss, A. R. (1994) *in* Tuberculosis (Bloom, B. R., Ed.), pp. 47–60, ASN Press, Washington, D.C.
2. Dolin, P. J., Ravihlione, M. C., Kochi, A. (1994) *WHO Bulletin* **72,** 213–220.
3. WHO, Geneva (1998) *WHO The World Health Report.*
4. Sepkowitz, K. A., Raffalli, J., Riley, L., Kiehn, T. E., Armstrong, D. (1995) *Clin. Microbiol. Revs.* **8,** 180–199.
5. Drobniewski, F. A., Pozniak, A. L., Uttley, A. H. C., (1995) *J. Med. Microbiol.* **42,** 85–91.
6. Morris, S., Bai, G. H., Suffys, P., Portillo-Gomez, L., Fairchok, M., Rouse, D. (1995) *J. Infect. Dis.* **171,** 954–960.
7. Heym, B., Alzari, P. M., Honore, N., Cole, S. T. (1995) *Mol. Microbiol.* **15,** 235–245.
8. Cole, S. T., Telneti, A. (1995) *Euro. Resp. J.* **8,** 710–713.
9. Centers for Disease Control and Prevention (1993) *Morbid. Mortal. Week. Rep.* **42,** 961–964.
10. Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., Mosteller F. (1994) *J. Amer. Med. Assoc.* **271,** 698–702.
11. Wilson, M. E., Fineberg, H. V., Colditz, G. A. (1995) *Clin. Infect. Dis.* **20,** 982–991.
12. Anderson, P. (1994) *Infect. & Immun.* **62,** 2536–2544.
13. Pascopella, L., Collins, F. M., Martin, J. M., Lee, M. H., Hatfull, G. F., Stover, C. K., Bloom, B. R., Jacobs, W. R., Jr. (1994) *Infect. Immun.* **62,** 1313–1319.
14. Plum, G., Clark-Curtiss, J. (1994) *Infect. Immun.* **62,** 476–483.
15. Shinnick, T. M., King, C. H., Quinn, F. D. (1995) *Amer. J. Med. Sci.* **309,** 92–98.
16. Dale, J. W. (1995) *Euro. Respir. J.* **8,** 633–648.
17. O'Donell, M. A. (1997) *Trends Biotechnol.* **15,** 512–516.
18. Finlay, B. B., Falkow, S. (1989) *Microbiol. Revs.* **53,** 210–230.
19. Strohman, R. (1994) *Bio/technology* **12,** 156–164.
20. Fleischmann, R. D., Adams, M. D., White, G., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., Venter, J. C. (1995) *Science* **269,** 496–512.
21. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Wediman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterbcak, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P. -C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchinson, C. A. III, Venter, J. C. (1995) *Science* **270,** 397–403.

22. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H. P., Fraser, C. M., Smith, H. O., Woese, C. R., Venter, J. C. (1996) *Science* **273,** 1058–1073.

23. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Tabata, S. (1996) *DNA Res.* **3,** 109–136.

24. Kaneko, T., Sato, S., Kotani, H., Tanak, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Tabata, S. (1996) *DNA Res.* **3,** 185–209.

25. Himmelreich, R., Hilbert, H., Plagens, H., Pirki, E., Li, B. C., Herrmann, R. (1996) *Nucleic Acids Res.* **24,** 4420–4449.

26. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S. G. (1996) *Science* **274,** 546–567.

27. Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E.,Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M., Venter, J. C. (1997) *Nature* **388,** 539–547.

28. Blattner, F. R., Plinkett, III, G., Bloch. C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goedeon, M. A., Rose, D. J., Mau, B., Shao, Y. (1997) *Science* **277,** 1453–1474.

29. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Chirch, G. M., Daniels, C. J., Mao, J.-I., Rice, P., Nöllong, J., Reeve, J. N. (1997) *J. Bacteriol.* **179,** 7135–7155.

30. Kunst, F., Ogasawara, N., Moszer, I., Albetini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S.-K., Codani, J.-J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Düsterhoft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.-Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.-F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaer-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S. Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.-M., Levine, A., Lui, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moesti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M. O., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.-H., Parro, V., Pohl, T. M., Portelle, D., Porwollok, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, E., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B.-S., Soldo, B., Sorokin, A., Taconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikaa, H.-F., Zumstein, E., Yoshikawa, H., Danchin, A. (1997) *Nature* **390,** 249–256.

31. Klenk, H.-P., Clayton, R., Tomb, J.-F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Krypides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenny, K., Adams, M. D., Loftus, B., Peterson, S., Reich, C. I., McNeil, L. K., Badger, J. H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J. D., Weidman, J. F., McDonald, L., Utterback, T., Cotton, M. D., Spriggs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., D'Andrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olsen, G. J., Fraser, C. M., Smith, H. O., Woese, C. R., Venter, J. C. (1997) *Nature* **390,** 364–370.

32. Fraser C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Boman, C., Garland, S., Fujii, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., Venter, J. C. (1997) *Nature* **390,** 580–586.

33. Cole, S. T., Brosch R., Parkhill J., Garnier, C., Churcher, D., Harris, S., Gordon, V., Eiglmeier, K., Gas, S., Barry III, C. E., 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, J., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., Barrell, B. G. (1998) *Nature* **393,** 537–544.

34. Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., Gwinn, M., Hickey, E. K., Clayton, R., Ketchum, K. A., Sodergren, E., Hardham, J. M., McLeod, M. P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J. K., Chidambaram, M., Utterback, T., McDonald, M. D., Fujii, C., Garland, S., Hatch, B., Horst, K., Roberts, K., Sandusky, M., Weidman, J., Smith, H. O., Venter, C. J. (1998) *Science* **281,** 375–388.

35. Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kikuchi, H. (1998) *DNA Res.* **5,** 147–155.

36. http://www.tigr.org/tdb/mdb/mdb.html.

37. http://www.mcs.anl.gov/home/gaasterl/genomes.html.

38. Koch, R. (1882) *BerlinerKlin Wochenschr* **19,** 221–230.
39. Cole, S. T., Smith, D. R. (1994) *in* Tuberculosis (Bloom, B. R., Ed.), pp. 239–252, ASM Press, Washington, D.C.
40. Smith, D. R. (1996) *Trends Biol. Sci.* **14,** 290–293.
41. Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L. & Humphery-Smith, I. (1995) *Electrophoresis* **16,** 1090–1094.
42. Cordwell, S. J., Basseal, D. J., Bjellqvist, B. J., Shaw, D. C., Humphery-Smith I. (1997) *Electrophoresis* **18,** *1393–1398.*
43. Pasquali, C., Frutiger, S., Kobayashi, R., Latter, G. I., Schwender, B., Volpe, T., Warner, J. R., McLaughlin, C. S. (1994) *Electrophoresis* **15,** 1466–1486.
44. VanBogelen, R. A., Sankar, P., Clark, R. L., Bogan, J. L., Neidhardt, F. C. (1992) *Electrophoresis* **13,** 1014–1054.
45. VanBogelen, R. A., Abshire, K. Z., Moldover, B., Olsen, E. R., Neidhardt, F. C. (1997) *Electrophoresis* **18,** 1243–1251.
46. Link, A. J., Robinson, K., Church, G. M. (1997) *Electrophoresis* **18,** 1259–1313.
47. Link, A. J., Hays, L. G., Carmack, E. B., Yates, III, J. R. (1997) *Electrophoresis* **18,** 1314–1334.
48. Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuenta, R., Payne, W. E. (1997) *Electrophoresis* **18,** 1347–1360.
49. Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93,** 14440–14445.
50. Urquhart, B. L. Atsalos, T. E., Roach, D., Basseal, D. J., Bjellqvist, B. J., Britton, W. L., Humphery-Smith, I. (1997) *Electrophoresis* **18,** 1384–1392.
51. Britton, W. J., Hellqvist, L., Ivanyi, J., Basten, A. (1987) *Scand. J. Immunol.* **26,** 149–159.
52. Daugelat, S., Gulle, H., Schoel, B., Kaufmann, H. E. (1992) *J. Infect. Dis.* **166,** 186–190.
53. Lee, B., Horwitz, M. A. (1995) *J. Clin. Invest.* **96,** 245–249.
54. Wallis, R. S., Paranjape, R., Phillips, M. (1993) *Infect. & Immun.* **6,** 627–632.
55. Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C., Stover, C. K. (1996) *J. Bacteriol.* **178,** 1274–1282.
56. Sonnenberg, M. G., Belisle, J. T. (1997) *Infect. & Immun.* **65,** 4515–4524.
57. Wasinger, V. C., Bjellqvist, B. J., Humphery-Smith, I. (1997) *Electrophoresis* **18,** 1373–1383.
58. Görg, A., Boguth, G., Obermaier, C., Weiss, W. (1998) *Electrophoresis* **19,** 1516–1519.
59. Arnott, D., O'Connell, K. L. King, K. L., Stults, J. T. (1998) *Anal Biochem.* **258,** 1–18.
60. Muller, E. C., Thiede, B., Zimny-Arndt, U., Scheler, C., Prehm, J., Muller-Werdan, U., Wiltmann-Liebold, B., Otto, A., Jungblut, P. (1996) *Electrophoresis* **17,** 1700–1712.
61. Bairoch, A. (1997) *in* Proteome Research: New Frontiers in Functional Genomics (Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F., Eds.), pp. 107–108, Springer-Verlag, Berlin.
62. Wasinger, V. C., Humphery-Smith, I. (1998) *FEBS Lett.*, in press.
63. Patton, W. F., Pluskal, M. G., Shea, W. M., Biecker, J. L., Lopez, M. F., Zimmerman, R., Belanger, L. M., Hatch, P. D. (1990) *Biotechniques* **8,** 518–527.
64. Voris, B. P., Young, D. A. (1980) *Anal. Biochem.* **104,** 478–484.
65. Young, D. A. (1984) *Clin. Chem.* **30,** 2104–2108.
66. Klose, J., Zeindl, E. (1984) *Clin. Chem.* **30,** 2014–2020.
67. Levinson, R. M., Anderson, G. M., Cohn, J. A., Blackshear, P. J. (1990) *Electrophoresis* **11,** 269–279.
68. Klose, J., Kobaltz, U. (1995) *Electrophoresis* **16,** 1034–1059.
69. Atland, K., von Eckardstein, A., Banzhoff, A., Wagner, M., Rossmann, U., Hackler, R., Becher, P. (1987) *Electrophoresis* **8,** 52–62.
70. Williams, M. R., Sanchez, J. -C., Williams, K. L., Hochstrasser, D. F. (1996) *Electrophoresis* **17,** 830–838.
71. Guptasarma, P. (1995) *Bioessays* **17,** 987–997.
72. Guptasarma, P. (1996) *Bioessays* **18,** 325–332.
73. Blattner, F. R., Burland, V., Plunkett, G., Sofia, H. J., Daniels, D. L. (1994) *Nucleic Acids Res.* **21,** 5408–5417.
74. Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Dover, J., Du, Z., Favello, A., Fulton, L., Gattung, S., Geisel, C., Kirsten, J., Kucaba, T., Hilier, L., Jier, M., Johnston, L., Langston, Y., Latrielle, P., Louis, E. J., Macri, C., Mardis, E., Menezes, S., Mouser, L., Nhan, M., Rifkin, L., Riles, L., St Peter, H., Trevaskis, E., Vaughan, K., Vignati, D., Vaudin, M. (1994) *Science* **265,** 2077–2082.
75. Humphery-Smith, I., Cordwell, S. J., Blackstock, W. P. (1997) *Electrophoresis* **18,** 1217–1242.
76. Wodicka, L., Dong, H., Mittman, M., Ho, M., Lockhart, D. J. (1997) *Nature Biotechnol.* **15,** 1359–1367.
77. Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, Davis, R. W. (1998) *Trends Biotechnol.* **16,** 301–302.
78. Anderson, L., Seilhamer, J. (1997) *Electrophoresis* **18,** 533–537.