Complementary Analysis of the *Mycobacterium tuberculosis* Proteome by Two-dimensional Electrophoresis and Isotope-coded Affinity Tag Technology*

Frank Schmidt‡, Samuel Donahoe§, Kristine Hagens¶, Jens Mattow¶, Ulrich E. Schaible¶, Stefan H. E. Kaufmann¶, Ruedi Aebersold§, and Peter R. Jungblut‡

Classical proteomics combined two-dimensional gel electrophoresis (2-DE) for the separation and quantification of proteins in a complex mixture with mass spectrometric identification of selected proteins. More recently, the combination of liquid chromatography (LC), stable isotope tagging, and tandem mass spectrometry (MS/MS) has emerged as an alternative quantitative proteomics technology. We have analyzed the proteome of *Mycobacterium tuberculosis*, a major human pathogen comprising about 4,000 genes, by (i) 2-DE and mass spectrometry (MS) and by (ii) the isotope-coded affinity tag (ICAT) reagent method and MS/MS. The data obtained by either technology were compared with respect to their selectivity for certain protein types and classes and with respect to the accuracy of quantification. Initial datasets of 60,000 peptide MS/MS spectra and 1,800 spots for the ICAT-LC/MS and 2-DE/MS methods, respectively, were reduced to 280 and 108 conclusively identified and quantified proteins, respectively. ICAT-LC/MS showed a clear bias for high *M* proteins and was complemented by the 2-DE/MS method, which showed a preference for low *M* proteins and also identified cysteine-free proteins that were transparent to the ICAT-LC/MS method. Relative quantification between two strains of the *M. tuberculosis* complex also revealed that the two technologies provide complementary quantitative information; whereas the ICAT-LC/MS method quantifies the sum of the protein species of one gene product, the 2-DE/MS method quantifies at the level of resolved protein species, including post-translationally modified and processed polypeptides. Our data indicate that different proteomic technologies applied to the same sample provide complementary types of information that contribute to a more complete understanding of the biological system studied.

Molecular & Cellular Proteomics 3:24–42, 2004.

© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.

* From the ‡Core Facility Protein Analysis and ¶Department of Immunology, Max Planck Institute for Infection Biology, D-10117 Berlin, Germany and the §Institute for Systems Biology, Seattle, Washington 98103-8904

Received, August 3, 2003, and in revised form, October 6, 2003 Published, MCP Papers in Press, October 13, 2003, DOI 10.1074/mcp.M300074-MCP200

Classical proteomics studies of *Mycobacterium tuberculosis* have combined two-dimensional electrophoresis (2-DE) with mass spectrometry (MS) and have revealed about 1,800 distinct spots separated by 2-DE (1). About 350 of these were identified, and the comparison of the protein patterns of virulent and attenuated strains identified several proteins that are being studied further as potential vaccine candidates (1, 2). More than two million deaths/year and eight million new infections are caused by *M. tuberculosis*, the bacterium responsible for tuberculosis (3). Genomics, transcriptomics, and proteomics, which form the rationale basis to develop new therapeutic and preventive strategies, have been applied to combat this disease. The complete genome of *M. tuberculosis* comprises about 4,000 genes, which were classified in six protein classes and 30 subclasses (4).

The smallest unit of the proteome, the protein species, is defined by its chemical structure (5). Therefore, each modification of a protein leads to a new protein species. These are successfully resolved by 2-DE if they differ by at least one charge or by at least several hundred daltons in mass. Quantification by 2-DE requires a high degree of pattern reproducibility, which is difficult to achieve in a multistep and parallel procedure.

To alleviate limitations of the 2-DE/MS method, more recently internally standardized gel-free quantitative proteomics methods have been developed. Of these the prototypical method is isotope-coded affinity tag (ICAT) reagent labeling and tandem mass spectrometry (MS/MS) (6). Proteins contained in two sample mixtures are covalently labeled with the isotopically light or heavy form of the ICAT reagents, respectively, and the samples are combined and proteolyzed. After purification of the labeled peptides via the affinity tag that is part of the reagents, they are analyzed by LC-MS/MS. The peptides eluting from the final chromatography step are subjected to data-dependent MS/MS, providing for peptide

1 The abbreviations used are: 2-DE, two-dimensional electrophoresis; MS, mass spectrometry; MS/MS, tandem mass spectrometry; LC, liquid chromatography; ICAT, isotope-coded affinity tag; MALDI, matrix-assisted laser desorption ionization.
quantification based on the relative signal intensities of the heavy and light forms of a particular peptide detected in the MS scan and for peptide identification by MS/MS and sequence database searching. This approach has been successfully applied to a wide variety of biological samples (7–14). Higher mass accuracy and manual peptide selection without the time pressure from data-dependent procedures were achieved by combining ICAT labeling with MALDI quadrupole time-of-flight mass spectrometry (15). Furthermore, the labeling protocols have been optimized (16), resulting in a robust quantitative technology.

Here we compare the results obtained by analyzing a microbial pathogen comprising about 4,000 genes using the 2-DE/MS and ICAT-LC/MS methods. Two strains of the *M. tuberculosis* complex were compared. Both technologies showed biases for and against certain types and classes of proteins and quantified proteins at different levels. 2-DE/MS complemented ICAT-LC/MS for low *M* and cysteine-free proteins and protein species separation. The ICAT technology complemented the 2-DE/MS with some functional protein classes. ICAT-LC/MS was superior for high *M* proteins and membrane proteins.

**EXPERIMENTAL PROCEDURES**

2-DE—After growing *M. tuberculosis* H37Rv and *Mycobacterium bovis* BCG in Middlebrook 7H9 broth as described previously (1), cellular proteins were dissolved in 9 M urea, 70 mM dithiothreitol, 2% carrier ampholytes Servalyte 2–4 (Serva, Heidelberg, Germany), and protease inhibitors (*N*-*p*-tosyl-L-lysine chlormethyl ketone, leupeptin, E64, pepstatin A; 25 μM) (1). Carrier ampholyte isoelectric focusing was combined with SDS-polyacrylamide gel electrophoresis (1). Gels of 0.75 mm were used. For analytical gels (gel width of 0.75 mm) to be stained with Coomassie Brilliant Blue G-250, 60 g of protein, respectively, were applied at the anodic side of the gel. The 2-DE technique used has a resolution power of 5,000 spots. Subtractive analysis was performed manually, and some of the clear intensity differences were quantified by the program TOPSPOT (available for download free of charge from www.mpiib-berlin.mpq.de/2D-PAGE/).

MALDI-MS—Proteins separated by 2-DE were identified by peptide mass fingerprinting after in-gel tryptic digestion as described previously (1). A Voyager Elite mass spectrometer (Perseptive Biosystems, Framingham, MA) was used with a mass accuracy of 30 ppm after internal calibration. Proteins were identified by MS-FIT or MASCOT (Matrix Science Ltd., London, United Kingdom, www.matrixscience.com) database searches. The identified proteins had a sequence coverage higher than 30%. For comparison with the ICAT results the dataset described earlier (1, 2) was used. Spot positions and identities with additional information and hyperlinks to sequence and pathway databases are available at www.mpiib-berlin.mpq.de/2D-PAGE/.

ICAT Reagent Labeling and Chromatography—Total cell extract was prepared for the two strains, and proteins were dissolved in buffer (6 M urea, 0.05% SDS, 5 mM Tris, pH 8.3, 5 mM EDTA) to obtain a protein concentration of 700 μg/200 μl and reduced with 5 mM tributylphosphine for 30 min at 37 °C. After addition of 350 nmol ICAT reagent to each sample (0.5 nmol of ICAT/μg of protein; final ICAT concentration, 1.75 μM) proteins were incubated for 90 min in the dark at room temperature with gentle stirring. After incubation, dithiothreitol was added to a final concentration of 10 mM to quench residual-free reagent. Labeled samples were mixed, diluted 4-fold with water so that the final urea concentration was 1.5 M, and digested with 1:25 trypsin/protein for 5 h at 37 °C. To remove the remaining ICAT reagent and other contaminants and to separate peptides, the resulting peptide mixture was combined, acidified to pH 3, and loaded on a 4.6 mm × 200 mm Polygel calf pocket phosphopeptide exchange chromatography column (Poly LC Inc., Columbus, MD) with 5-μm particles and 300-Å pores at a flow rate of 800 μl/min. The column was washed by buffer A (5 mM KH₂PO₄, 25% acetonitrile, pH 3.0), and the peptides were eluted by buffer B (20 mM KH₂PO₄, 350 mM KCl, 25% acetonitrile, pH 3.0). The column was developed over a 50-min dual-step salt gradient.

The collected fractions were dried by speed-vac and resuspended in 2× phosphate-buffered saline pH 7.2 for avidin purification using a self-packed UltraLink monomeric avidin column (Ferrett 400 μl of packed beads in a glass Pasteur pipette. The column was washed with water, and biotinylated peptides were eluted at 0.3% formic acid. Avidin column eluent was dried down by speed-vac, and the pellet was resuspended in reverse phase buffer A (5% acetic acid, 0.005% heptfluorobutyric acid). The biotinylated peptides were analyzed using a reverse phase capillary chromatography (75 μm, 10-cm self-packed C18 column, Monitor, Column Engineering, Ontario, Canada) at a flow rate of 250 nl/min.

Ion Trap-MS—Peptide identification by collision-induced dissociation was carried out by data-dependent precursor ion selection using the dynamic exclusion option on a Finnigan LCQ ion trap mass spectrometer. The MS/MS spectra were searched against a protein sequence database (*M. tuberculosis* H37Rv, ftp://ftp.sanger.ac.uk/pub/tb/sequences/TB.pep; 3,924 entries) using the SEQUEST software tool, and the abundance ratios of isotopically labeled cysteinyl peptide pairs were calculated using the XPRESS program (14).

Only strictly tryptic cysteinyl peptides and peptides with an Xcorr score value higher than 3 were used for identification. The first 800 identifications from SEQUEST, sorted by Xcorr, were verified by spot check using Mascot. Also, quantification of the first 800 cysteine-containing peptides was inspected manually. The SEQUEST results and the manual changes were implemented and stored in a relational MySQL database, and a Web interface was developed to ask multicriteria questions for intelligent data searches (www.mpiib-berlin.mpq.de/2D-PAGE/ functional classification tool). The database entries fulfilling these search criteria are reported by a browser such as Netscape or Internet Explorer.

**RESULTS**

Data Amount and Reduction—The *M. tuberculosis* H37Rv genome has been predicted to comprise about 4,000 genes (4). However, it cannot be expected that proteins representing every predicted gene are present in a bacterial culture grown under specific conditions. From *M. tuberculosis* H37Rv cells in late exponential phase about 1,800 distinct protein spots were resolved on silver-stained 2-DE gels (1). Analysis of the same samples by ICAT-LC/MS identified about 60,000 peptides. Both datasets contained redundancies. Because of post-translational modifications and processing, many genes are represented on 2-DE gels by multiple spots. In the ICAT experiment a particular peptide may occur in different chromatographic fractions and will therefore be sequenced multiple times. To obtain comparable datasets data reduction was necessary for both approaches. The data analysis flow chart is shown in Fig. 1.
The criterion to have comparable datasets was the accessibility of data. In 2-DE gels the best accessible proteins are within the high intensity spots, which resulted in high identification scores. In ICAT-LC/MS, where the abundance cannot be directly measured, the accessibility is also determined by scoring factors. Therefore, as a comparable dataset the most intense spots from a 2-DE gel and the most reliably identified proteins from the ICAT analysis were chosen. The 560 spots initially identified by 2-DE/MS and present in both strains were listed in order of decreasing silver-staining intensity. Of these the 160 most intense spots were chosen for further analysis. After removing redundancies caused by post-translational modifications, 108 unique proteins constituted the final 2-DE dataset that was used throughout the study (Table I).

The primary ICAT-LC/MS data contained numerous cysteine-free peptides and peptides missing lysines or arginines at their C terminus. These primary data were reduced to 2,000 cysteinyl peptides with double tryptic termini. The 800 peptides with the highest SEQUEST scores were further manually evaluated, resulting in 280 uniquely identified proteins (Table II). This final ICAT-MS dataset was used throughout the study. Within the 108 and 280 proteins identified by 2-DE/MS and ICAT-LC/MS, respectively, 27 were common to both (Fig. 2).

For peptide quantification, the same 800 spectra were evaluated manually, and if necessary, quantification calculated by the XPRESS program was corrected manually by adjusting the scan intervals, the tolerance, and the predicted shift of light and heavy peptide forms.

All the identified proteins were inserted into a 2D-PAGE database (www.mpiib-berlin.mpg.de/2D-PAGE). Within this database the position of the identified spots on 2-DE gels, including $M_r$ and $pI$ values, MS data, and other characteristics, is available. Links to sequence, classification, and pathway databases provide further information. The relational structure of the database is constructed to allow flexible data mining procedures (18).

**Protein Class Biases of 2-DE/MS and ICAT-LC/MS Proteomics**—The genome of *M. tuberculosis* was classified into six functional classes (Fig. 3) (4) and further subdivided into 30 subclasses comprising 80 protein families. Three of these protein families (respiration, IS elements, and PE family) were further classified into subfamilies. To identify protein classes that were over- or underrepresented in the two datasets for either method the ratio of the number of identified proteins of
each class and the total number of proteins was calculated. These ratios were compared with the ratio of the total number of members of each family and the total number of predicted proteins (normalized ratio).

At the level of protein classes it is obvious that both experimental approaches overrepresented the number of proteins from the class “small molecule metabolism,” whereas the protein classes “other” and “unknowns” were clearly under-
### TABLE II

*Intensity differences detected by ICAT*

No., number of identified and quantified peptides of proteins; if ≥2, the average value of the ratio was used.

| Open reading frame | Protein name                                      | No. | BCG | H37Rv |
|--------------------|---------------------------------------------------|-----|-----|-------|
| Rv0005             | DNA gyrase subunit B                              | 1   | 1   | 5.0   |
| Rv0007             | Conserved hypothetical (= MLB1770.07)             | 4   | 1   | 1.3   |
| Rv0018c            | Putative phosphoprotein phosphatase               | 2   | 1   | 0.5   |
| Rv0087             | Probable formate hydrogenase subunit 5            | 1   | 1   | 0.8   |
| Rv0108c            | Hypothetical protein                              | 3   | 1   | 1.1   |
| Rv0120c            | Elongation factor G                               | 1   | 1   | 3.3   |
| Rv0141c            | Hypothetical protein                              | 5   | 1   | 1.3   |
| Rv0147             | Aldehyde dehydrogenase                           | 8   | 1   | 1.4   |
| Rv0149             | Putative oxidoreductase                           | 1   | 1   | 0.8   |
| Rv0174             | Mce family, part of mce1 operon                   | 8   | 1   | 1.4   |
| Rv0175             | Conserved hypothetical                            | 4   | 1   | 0.8   |
| Rv0183             | Probable oxidoreductase                           | 2   | 1   | 0.6   |
| Rv0189c            | Dihydroxy-acid dehydratase                        | 4   | 1   | 0.9   |
| Rv0192             | Conserved hypothetical protein                    | 1   | 1   | 0.04  |
| Rv0211             | Phosphoenolpyruvate carboxykinase                 | 1   | 1   | 0.7   |
| Rv0216             | Conserved hypothetical                            | 4   | 1   | 0.1   |
| Rv0223c            | Aldehyde dehydrogenase (possible betB)           | 2   | 0   | 1.0   |
| Rv0231             | Acyl-CoA dehydrogenase                            | 1   | 1   | 1.7   |
| Rv0233             | Ribonucleoside-diphosphate reductase B2           | 1   | 1   | 7.1   |
| Rv0234c            | Succinate-semialdehyde dehydrogenase              | 2   | 1   | 0.6   |
| Rv0244c            | Acyl-CoA dehydrogenase                            | 2   | 1   | 0.6   |
| Rv0247c            | Probable iron-sulphur protein                     | 1   | 1   | 0.5   |
| Rv0264c            | Conserved hypothetical protein                    | 1   | 1   | 0.5   |
| Rv0281             | Conserved hypothetical protein                    | 3   | 1   | 0.4   |
| Rv0303             | Similar to NADPH dehydrogenase                   | 1   | 1   | 2.0   |
| Rv0315             | Probable b-1,3-glucanase                          | 3   | 1   | 0.3   |
| Rv0337c            | Aspartate aminotransferase                        | 2   | 1   | 0.6   |
| Rv0414c            | Thiamine synthesis, thiazole moiety               | 2   | 1   | 1.0   |
| Rv0457c            | Probable peptidase                                | 1   | 1   | 0.6   |
| Rv0458             | Aldehyde dehydrogenase                            | 2   | 1   | 3.3   |
| Rv0462             | Probable dihydrodipamide dehydrogenase           | 1   | 1   | 0.5   |
| Rv0469             | Unknown mycolic acid methyltransferase            | 2   | 1   | 1.3   |
| Rv0470c            | Unknown mycolic acid methyltransferase            | 3   | 1   | 0.8   |
| Rv0489             | Phosphoglycerate mutase I                         | 3   | 1   | 0.7   |
| Rv0503c            | Cyclopropane mycolic acid synthase 2              | 6   | 1   | 2.5   |
| Rv0525             | Conserved hypothetical protein                    | 1   | 1   | 0.9   |
| Rv0547c            | Putative oxidoreductase                           | 1   | 1   | 0.3   |
| Rv0551c            | Acyl-CoA synthase                                 | 1   | 1   | 0.8   |
| Rv0570             | Ribonucleotide reductase, class II                | 1   | 0   | 1.0   |
| Rv0571c            | Conserved hypothetical (adjacent open reading frames U650S and U650T) | 2   | 1   | 0.9   |
| Rv0634c            | Putative glyoxylase II                            | 1   | 1   | 1.4   |
| Rv0637             | Conserved hypothetical protein                    | 1   | 1   | 0.6   |
| Rv0640             | 50 S ribosomal protein L11                        | 6   | 1   | 1.1   |
| Rv0642c            | Methoxymycolic acid synthase 4                    | 7   | 1   | 1.0   |
| Rv0655             | ABC transporter                                   | 3   | 1   | 1.3   |
| Rv0663             | Probable arylsulfatase                            | 1   | 1   | 0.06  |
| Rv0672             | Acyl-CoA dehydrogenase (also known as aidB)      | 3   | 1   | 1.1   |
| Rv0684             | Elongation factor G                               | 3   | 1   | 0.5   |
| Rv0688             | Putative oxidoreductase                           | 1   | 1   | 1.4   |
| Rv0721             | 30 S ribosomal protein S5                         | 2   | 1   | 0.7   |
| Rv0734             | Probable methionine aminopeptidase                | 1   | 1   | 1.4   |
| Rv0738             | Possible membrane protein                         | 2   | 1   | 1.1   |
| Rv0751c            | Methylmalonate semialdehyde oxidoreductase        | 1   | 1   | 5.0   |
| Rv0753c            | Methylmalonate semialdehyde dehydrogenase        | 2   | 1   | 0.8   |
| Rv0761c            | Zinc-containing alcohol dehydrogenase            | 7   | 1   | 0.9   |
| Rv0773c            | Putative g-glutamyl transpeptidase                | 1   | 1   | 0.9   |
| Rv0780             | Phosphoribosylaminomimidazole-succinocarboxamide | 1   | 1   | 0.7   |
| Rv0791c            | Possible monooxygenasemonooxygenase               | 1   | 1   | 0.9   |
| Open reading frame | Protein name | No. | BCG | H37Rv |
|-------------------|--------------|-----|-----|-------|
| Rv0803            | Phosphoribosylformylglycinamidine synthase II | 4   | 1   | 0.7   |
| Rv0815c           | Thiosulfate sulfurtransferase               | 3   | 1   | 0.5   |
| Rv0823c           | Transcriptional regulator (NifR3/Smm1 family) | 1   | 1   | 0.5   |
| Rv0824c           | Acyl-[ACP] desaturase/fatty acid desaturase | 15  | 1   | 0.8   |
| Rv0853c           | Pyruvate (or indolepyruvate) decarboxylase  | 1   | 1   | 1.7   |
| Rv0859            | β oxidation complex subunit (acetyl-CoA)    | 5   | 1   | 1.3   |
| Rv0873            | Acyl-CoA dehydrogenase                      | 4   | 1   | 1.7   |
| Rv0888            | Possible membrane protein                   | 1   | 1   | 3.3   |
| Rv0905            | Enoyl-CoA hydratase/isomerase superfamily   | 2   | 1   | 1.1   |
| Rv0925c           | Hypothetical protein                        | 1   | 1   | 0.9   |
| Rv0928            | PstS component of phosphate uptake          | 2   | 1   | 0.2   |
| Rv0951            | Succinyl-CoA synthase b chain               | 6   | 1   | 0.7   |
| Rv0952            | Succinyl-CoA synthase α chain               | 4   | 1   | 0.7   |
| Rv0957            | Phosphoribosylaminomimidazolecarboxamidine  | 1   | 1   | 0.7   |
| Rv0971c           | Enoyl-CoA hydratase/isomerase superfamily   | 1   | 1   | 0.6   |
| Rv0976c           | Hypothetical protein                        | 1   | 1   | 0.8   |
| Rv1006            | Hypothetical protein                        | 1   | 1   | 0.9   |
| Rv1056            | Conserved hypothetical protein              | 1   | 1   | 1.4   |
| Rv1071c           | Enoyl-CoA hydratase/isomerase superfamily   | 2   | 1   | 0.6   |
| Rv1074c           | Acetyl-CoA C-acetyltransferase              | 1   | 1   | 1.1   |
| Rv1093            | Serine hydroxymethyltransferase             | 2   | 1   | 0.3   |
| Rv1094            | Acyl-[ACP] desaturase                       | 1   | 1   | 0.9   |
| Rv1097c           | Probable membrane-spanning protein         | 2   | 1   | 1.0   |
| Rv1103c           | Hypothetical protein                        | 2   | 1   | 2.0   |
| Rv1110            | Very similar to LytB                       | 1   | 1   | 1.1   |
| Rv1122            | 6-Phosphogluconate dehydrogenase (Gram+)   | 4   | 1   | 0.8   |
| Rv1180            | Polyketide synthase                        | 2   | 1   | 5.0   |
| Rv1181            | Polyketide synthase                        | 4   | 1   | 7.6   |
| Rv1206            | Acyl-CoA synthase                          | 1   | 1   | 0.6   |
| Rv1248c           | 2-Oxoglutarate dehydrogenase               | 2   | 1   | 1.0   |
| Rv1269c           | Conserved hypothetical protein              | 2   | 1   | 2.0   |
| Rv1274            | Lipoprotein                                | 4   | 1   | 0.7   |
| Rv1278            | Hypothetical protein                       | 1   | 1   | 0.4   |
| Rv1295            | Homoserine synthase/threonine synthase      | 3   | 1   | 0.7   |
| Rv1297            | Transcription termination factor rho        | 1   | 1   | 0.5   |
| Rv1299            | Peptide chain release factor 1              | 1   | 1   | 0.9   |
| Rv1309            | ATP synthase γ chain                       | 1   | 1   | 0.7   |
| Rv1310            | ATP synthase β chain                       | 1   | 1   | 1.1   |
| Rv1314c           | Conserved hypothetical (= P53523)          | 2   | 1   | 0.9   |
| Rv1323            | Acetyl-CoA C-acetyltransferase (also known as thiL) | 5   | 1 | 0.7 |
| Rv1324            | Conserved hypothetical (some similarity to thioredoxins) | 6 | 1 | 0.8 |
| Rv1326c           | 1,4-α-glucan branching enzyme              | 1   | 1   | 0.6   |
| Rv1339            | Conserved hypothetical protein             | 2   | 1   | 0.6   |
| Rv1364c           | SigB regulation protein                    | 1   | 1   | 0.5   |
| Rv1384            | Carbamoyl-phosphate synthase subunit       | 1   | 1   | 0.7   |
| Rv1392            | S-adenosylmethionine synthase              | 2   | 1   | 1.1   |
| Rv1416            | Riboflavin synthase β chain                | 1   | 1   | 0.5   |
| Rv1436            | Glyceraldehyde-3-phosphate dehydrogenase   | 9   | 1   | 0.9   |
| Rv1438            | Triosephosphate isomerase                  | 1   | 1   | 0.8   |
| Rv1449c           | Transketolase                              | 2   | 1   | 0.5   |
| Rv1454c           | Probable quinone oxido-reductase           | 1   | 1   | 1.4   |
| Rv1458c           | ABC-type transporter                       | 1   | 1   | 0.6   |
| Rv1462            | Conserved hypothetical protein             | 2   | 1   | 0.2   |
| Rv1483            | 3-Oxoacyl-[ACP] reductase (also known as MabA) | 2 | 1 | 0.3 |
| Rv1513            | Conserved hypothetical protein             | 1   | 0   | 1.0   |
| Rv1523            | Possible methyl-sterol transferase         | 1   | 1   | 0.7   |
| Rv1527c           | Polyketide synthase                        | 3   | 1   | 0.3   |
| Rv1543            | Probable fatty-acyl CoA reductase          | 2   | 1   | 0.8   |
| Rv1544            | Probable ketoacyl reductase                | 3   | 1   | 1.0   |
| Rv1551            | Conserved hypothetical                     | 1   | 1   | 0.6   |
### TABLE II—continued

| Open reading frame | Protein name                                      | No. | BCG | H37Rv |
|--------------------|--------------------------------------------------|-----|-----|-------|
| Rv1609             | Anthranilate synthase component I                | 2   | 1   | 0.6   |
| Rv1640c            | C-term lysyl-tRNA synthase                       | 1   | 1   | 3.3   |
| Rv1641             | Initiation factor IF-3                            | 4   | 1   | 1.7   |
| Rv1654             | Acetylglutamate kinase                            | 3   | 1   | 0.4   |
| Rv1658             | Arginosuccinate synthase                          | 2   | 1   | 0.3   |
| Rv1703c            | Arginosuccinate synthase                          | 1   | 1   | 0.6   |
| Rv1731             | Succinate-semialdehyde dehydrogenase             | 2   | 1   | 1.3   |
| Rv1732c            | Conserved hypothetical protein                    | 5   | 1   | 0.8   |
| Rv1770             | Conserved hypothetical protein                    | 1   | 1   | 0.7   |
| Rv1771             | Oxidoreductase, possible                          | 1   | 1   | 0.7   |
| Rv1784             | Conserved hypothetical protein                    | 1   | 1   | 0.6   |
| Rv1793             | Conserved hypothetical protein                    | 2   | 1   | 0.5   |
| Rv1794             | Conserved hypothetical (similar to MTCY15F10.23; 25.7% identity in 307aa) | 2   | 1   | 1.0   |
| Rv1812c            | Probable dehydrogenase                           | 1   | 1   | 0.3   |
| Rv1819c            | Probable multidrug resistance pump               | 1   | 1   | 0.5   |
| Rv1837c            | glmB malate synthase                              | 2   | 1   | 0.7   |
| Rv1843c            | Inosine-5'-monophosphate dehydrogenase           | 3   | 1   | 1.7   |
| Rv1872c            | L-lactate dehydrogenase                          | 4   | 1   | 2.5   |
| Rv1876             | Bacterioferritin                                  | 1   | 1   | 0.7   |
| Rv1886c            | Antigen 85B, mycolyltransferase                  | 2   | 1   | 1.7   |
| Rv1908c            | Catalase-peroxidase                               | 2   | 1   | 1.1   |
| Rv1925             | Acyl-CoA synthase                                 | 1   | 1   | 3.3   |
| Rv1932             | Thiol peroxidase                                  | 13  | 1   | 0.6   |
| Rv1942c            | Hypothetical protein                              | 1   | 1   | 0.8   |
| Rv1997             | Probable cation transport ATPase                 | 1   | 1   | 1.4   |
| Rv1998c            | Conserved hypothetical                            | 3   | 1   | 0.2   |
| Rv2000             | Hypothetical protein                              | 1   | 1   | 0.6   |
| Rv2004c            | Unknown                                          | 3   | 1   | 0.9   |
| Rv2032             | Conserved hypothetical protein                    | 1   | 1   | 0.8   |
| Rv2051c            | Probable membrane protein                         | 1   | 1   | 0.8   |
| Rv2062c            | Cobalt insertion                                 | 1   | 1   | 1.7   |
| Rv2068c            | Class A b-lactamase                               | 1   | 1   | 2.5   |
| Rv2091c            | Potential transmembrane region                    | 2   | 1   | 1.0   |
| Rv2102             | Conserved hypothetical protein                    | 1   | 1   | 1.0   |
| Rv2115c            | ATPase of AAA family                             | 1   | 1   | 1.2   |
| Rv2121c            | ATP phosphoribosyltransferase                    | 4   | 1   | 1.1   |
| Rv2135c            | Conserved hypothetical (= Q49773 B2126 C1 148)   | 1   | 1   | 3.3   |
| Rv2141c            | ArgE/DapE/Acr1/Cpg2/yscS family                  | 2   | 1   | 1.3   |
| Rv2161c            | Similar to alkanal monooxygenase β chain          | 5   | 1   | 6.3   |
| Rv2187             | Acyl-CoA synthase                                 | 1   | 1   | 7.6   |
| Rv2202c            | Carbohydrate kinase                              | 2   | 1   | 0.7   |
| Rv2213             | Aminopeptidase A/l                                | 1   | 1   | 0.6   |
| Rv2214c            | Probable epoxide hydrolase                        | 3   | 1   | 0.8   |
| Rv2220             | Glutamine synthase class I                        | 66  | 1   | 1.1   |
| Rv2245             | β-ketoacyl-ACP synthase (meromycolate)            | 2   | 1   | 0.8   |
| Rv2246             | β-ketoacyl-ACP synthase (meromycolate)            | 8   | 1   | 0.6   |
| Rv2247             | Acetyl(proplyl) CoA carboxylase β subunit         | 4   | 1   | 0.5   |
| Rv2249c            | Glycerol-3-phosphate dehydrogenase               | 2   | 1   | 0.9   |
| Rv2289c            | CDP-diacylglycerol phosphatidylhydrolase         | 1   | 1   | 3.3   |
| Rv2294             | Aminotransferase                                  | 3   | 1   | 0.9   |
| Rv2301             | Probable cutinase                                 | 1   | 1   | 1.4   |
| Rv2373c            | DnaJ homologue                                   | 1   | 1   | 0.6   |
| Rv2416c            | Conserved hypothetical protein                    | 1   | 1   | 0.4   |
| Rv2421c            | Conserved hypothetical protein                    | 1   | 1   | 0.5   |
| Rv2427c            | γ-glutamyl phosphate reductase                    | 1   | 1   | 0.8   |
| Rv2428             | Alkyl hydroperoxide reductase                    | 2   | 1   | 0.03  |
| Rv2454c            | Oxidoreductase, β subunit                         | 4   | 1   | 0.6   |
| Rv2460c            | ATP-dependent Clp protease proteolytic subunit    | 5   | 1   | 1.3   |
| Rv2465c            | Phosphopentose isomerase                          | 1   | 1   | 0.7   |
| Rv2467             | Probable aminopeptidase                           | 2   | 1   | 0.7   |
| Open reading frame | Protein name | No. | BCG | H37rv |
|--------------------|--------------|-----|-----|-------|
| Rv2468c            | Unknown      | 1   | 1   | 0.7   |
| Rv2476c            | Hypothetical protein | 1 | 1 | 0.9 |
| Rv2477c            | ABC-transporter ATP-binding protein | 1 | 1 | 0.6 |
| Rv2495c            | Dihydrolipoamide acetyltransferase | 2 | 1 | 0.5 |
| Rv2521             | Bacterioferritin comigratory protein | 3 | 1 | 0.6 |
| Rv2524c            | Fatty acid synthase | 5 | 1 | 0.7 |
| Rv2557             | Conserved hypothetical protein | 23 | 1 | 4.0 |
| Rv2558             | Conserved hypothetical protein | 2 | 1 | 2.5 |
| Rv2575             | Conserved hypothetical protein | 1 | 1 | 3.3 |
| Rv2610c            | Glycosyltransferase | 1 | 1 | 1.7 |
| Rv2623             | Conserved hypothetical | 5 | 1 | 0.6 |
| Rv2626c            | Conserved hypothetical protein | 45 | 1 | 0.4 |
| Rv2629             | Hypothetical protein | 3 | 1 | 0.6 |
| Rv2632c            | Conserved hypothetical protein | 2 | 1 | 10.0 |
| Rv2700             | Conserved hypothetical (= U1764H) | 1 | 1 | 1.1 |
| Rv2711             | Iron-dependent repressor, IdeR | 1 | 1 | 1.1 |
| Rv2751             | Conserved hypothetical protein | 1 | 1 | 1.1 |
| Rv2752c            | Conserved hypothetical protein | 1 | 1 | 0.8 |
| Rv2753c            | Dihydricopicolinate synthase | 2 | 1 | 0.8 |
| Rv2765             | Unknown | 2 | 1 | 0.8 |
| Rv2784c            | Lipoprotein | 2 | 1 | 0.8 |
| Rv2790c            | Nonspecific lipid transport protein | 3 | 1 | 0.6 |
| Rv2851c            | Conserved hypothetical protein | 1 | 1 | 1.4 |
| Rv2868c            | Essential gene of unknown function | 10 | 1 | 0.6 |
| Rv2889c            | Elongation factor EF-Ts | 1 | 1 | 0.5 |
| Rv2890c            | 30 S ribosomal protein S2 | 3 | 1 | 0.7 |
| Rv2923c            | Conserved hypothetical protein | 1 | 1 | 0.4 |
| Rv2935             | Phenolphthiocerol synthesis (pksF) | 3 | 1 | 0.2 |
| Rv2940c            | Myocerosic acid synthase | 6 | 1 | 0.3 |
| Rv2953             | Conserved hypothetical protein | 4 | 1 | 0.8 |
| Rv2970c            | Probable lipase/esterase | 3 | 1 | 1.4 |
| Rv2971             | Oxidoreductase of Aldo/keto reductase family | 6 | 1 | 0.24 |
| Rv2982c            | Glycerol-3-phosphate dehydrogenase | 1 | 1 | 0.5 |
| Rv2984             | Polyphosphat kinase | 1 | 1 | 0.8 |
| Rv2987c            | 3-Isopropylmalate dehydratase small subunit | 1 | 1 | 0.2 |
| Rv2993c            | Conserved hypothetical protein | 2 | 1 | 0.5 |
| Rv3001c            | Ketol-acid reductoisomerase | 8 | 1 | 2.0 |
| Rv3010c            | Phosphofructokinase I | 6 | 1 | 0.6 |
| Rv3033             | Hypothetical protein | 1 | 1 | 1.0 |
| Rv3040c            | Hypothetical protein | 1 | 1 | 1.1 |
| Rv3044             | Putative Felli-dicatrate transporter | 1 | 1 | 1.0 |
| Rv3045             | Alcohol dehydrogenase | 2 | 1 | 0.3 |
| Rv3051c            | Ribonucleoside diphosphate reductase a chain | 3 | 1 | 0.8 |
| Rv3075c            | Conserved hypothetical | 3 | 1 | 0.8 |
| Rv3086             | Zinc-containing alcohol dehydrogenase | 1 | 1 | 5.0 |
| Rv3089             | Acyl-CoA synthase | 1 | 1 | 8.3 |
| Rv3090             | Hypothetical protein | 1 | 1 | 0.5 |
| Rv3094c            | Conserved hypothetical protein | 1 | 1 | 0.9 |
| Rv3130c            | Conserved hypothetical protein | 1 | 1 | 0.2 |
| Rv3131             | Conserved hypothetical protein | 5 | 1 | 1.0 |
| Rv3133c            | Two-component response regulator | 1 | 1 | 0.6 |
| Rv3134c            | Conserved hypothetical | 2 | 1 | 0.7 |
| Rv3139             | Acyl-CoA dehydrogenase | 3 | 1 | 1.3 |
| Rv3145             | NADH dehydrogenase chain A | 2 | 1 | 0.4 |
| Rv3150             | NADH dehydrogenase chain F | 2 | 1 | 0.4 |
| Rv3151             | NADH dehydrogenase chain G | 5 | 1 | 0.7 |
| Rv3197             | Probable ABC transporter | 1 | 1 | 0.8 |
| Rv3208             | Transcriptional regulator (TetR/AcrR family) | 2 | 1 | 0.7 |
| Rv3224             | Putative oxidoreductases | 7 | 1 | 0.6 |
| Rv3240c            | SecA, preprotein translocase subunit | 3 | 1 | 1.1 |
represented (Fig. 3). Both 2-DE/MS and ICAT-LC/MS approaches revealed similar numbers/class. However, at the resolution of the protein subclass level (Fig. 4), the ICAT-LC/MS method compared with the 2-DE/MS method preferentially identified proteins of the subclasses “central intermediary metabolism,” “polyketide and non-ribosomal peptide synthesis,” “degradation of macromolecules,” “cell envelope,” and “transport/binding proteins.” In contrast, 2-DE preferred relative to the ICAT method the protein subclasses “lipid biosynthesis,” “synthesis and modification of macromolecules,” “chaperones/heat-shock proteins,” “protein and peptide secretion,” and “adaptations and atypical conditions.” Both methods overestimated relative to the normalized ratio the subclasses “degradation,” “energy metabolism,” “amino acid biosynthesis,” and “lipid biosynthesis and detoxification” and underestimated “cell envelope,” “transport/binding proteins,” and “virulence.” None of the proteins in the subclasses “cell division,” “IS elements, repeated sequences,” “PE and PPE families,” “cytochrome P450 enzymes,” “cyclases,” and “chelatases” were experimentally verified by either of the two methods within the reduced datasets. With the restricted number of experimentally identified proteins a more detailed analysis at the protein family or subfamily level was not considered. These data suggest that either method as executed in this study was significantly biased in favor of highly abundant proteins.

| Open reading frame | Protein name | No. | BCG | H37rv |
|--------------------|--------------|-----|-----|-------|
| Rv3241c            | Member of S30AE ribosomal protein family | 2   | 1   | 1.1   |
| Rv3246c            | Two-component response regulator | 3   | 1   | 0.8   |
| Rv3248c            | Adenosylhomocysteine | 1   | 1   | 0.6   |
| Rv3264c            | Glucose-1-phosphate thiomydyltransferase | 1   | 1   | 0.9   |
| Rv3273c            | C-term similar to carbolic anhydrase | 7   | 1   | 0.1   |
| Rv3274c            | Acyl-CoA dehydrogenase | 1   | 1   | 1.1   |
| Rv3280             | Acetyl/propionyl CoA carboxylase b subunit | 2   | 1   | 0.7   |
| Rv3283             | Thiosulfate sulfurtransferase | 4   | 1   | 0.7   |
| Rv3285             | Acetyl/propionyl CoA carboxylase α subunit | 6   | 1   | 0.6   |
| Rv3295             | Transcriptional regulator (TetR/AcrR family) | 1   | 1   | 1.3   |
| Rv3313c            | Probable adenosine deaminase | 2   | 1   | 0.7   |
| Rv3332             | N-acetylglucosamine-6-P-deacetylase | 2   | 1   | 2.5   |
| Rv3389c            | Putative dehydrogenase | 3   | 1   | 0.8   |
| Rv3396c            | GMP synthase | 2   | 1   | 0.3   |
| Rv3411c            | Inosine-5′-monophosphate dehydrogenase | 2   | 1   | 0.7   |
| Rv3464             | dTDP-glucose 4,6-dehydratase | 2   | 1   | 1.0   |
| Rv3520c            | Probable coenzyme F420-dependent enzyme | 2   | 1   | 1.1   |
| Rv3534c            | 4-Hydroxy-2-oxovalerate aldolase | 4   | 1   | 0.7   |
| Rv3535c            | Acetaldehyde dehydrogenase | 3   | 1   | 0.9   |
| Rv3556c            | Acetyl-CoA C-acetyltransferase | 1   | 1   | 1.4   |
| Rv3564             | Acyl-CoA dehydrogenase | 1   | 1   | 0.5   |
| Rv3678c            | Transcriptional regulator (LysR family) | 1   | 1   | 1.0   |
| Rv3684             | Probable lyase, cysteine metabolism | 6   | 1   | 0.6   |
| Rv3708c            | Aspartate semialdehyde dehydrogenase | 4   | 1   | 0.8   |
| Rv3709c            | Aspartokinase | 1   | 1   | 1.3   |
| Rv3720             | C-term similar to cyclopropane fatty acid synthases | 2   | 1   | 1.4   |
| Rv3721c            | DNA polymerase II, α (dnaA) and β (dnaB) | 4   | 1   | 1.1   |
| Rv3722c            | Hypothetical protein | 1   | 1   | 0.4   |
| Rv3726             | Putative alcohol dehydrogenase, zinc-type | 4   | 1   | 0.9   |
| Rv3734c            | Conserved hypothetical | 2   | 1   | 1.0   |
| Rv3761c            | Acyl-CoA dehydrogenase | 1   | 1   | 0.8   |
| Rv3772             | Histidinol-phosphate aminotransferase | 2   | 1   | 0.9   |
| Rv3794             | Involved in arabinogalactan synthesis | 1   | 1   | 1.0   |
| Rv3799c            | Acetyl/propionyl CoA carboxylase β subunit | 2   | 1   | 0.5   |
| Rv3801c            | Acyl-CoA synthase | 3   | 1   | 1.4   |
| Rv3804c            | Antigen 85A, mycolyltransferase | 5   | 1   | 2.0   |
| Rv3825c            | Polyketide synthase | 2   | 1   | 2.0   |
| Rv3837c            | Putative phosphoglycerate mutase | 1   | 1   | 1.1   |
| Rv3856c            | Conserved hypothetical protein | 1   | 1   | 0.8   |
| Rv3859c            | Ferredoxin-dependent glutamate synthase | 1   | 1   | 0.3   |
| Rv3881c            | Conserved hypothetical protein | 2   | 1   | 1.7   |
| Rv3913c            | Thioredoxin reductase | 2   | 1   | 1.3   |
Molecular Mass Bias—To estimate the bias of either method for the identification of proteins of a particular mass range the percentages of proteins predicted from the genome of *M. tuberculosis* H37Rv and of proteins identified by 2-DE/MS and ICAT-LC/MS were sorted by *M*<sub>r</sub> in bins of 10-kDa width and into an additional bin for proteins >100 kDa (Fig. 5a). With 61.7% most of the proteins were predicted in the *M*<sub>r</sub> range between 10 and 40 kDa. In the range between 10 and 60 kDa 3,300 of 3,924 (84%) proteins were predicted. With respect to the *M*<sub>r</sub> of identified proteins the 2-DE and ICAT methods were complementary. For the low mass range (*M*<sub>r</sub> 10–30 kDa) proteins identified by 2-DE/MS were clearly overrepresented, whereas for all the mass ranges >30 kDa proteins identified by ICAT-LC/MS were overrepresented. The overrepresentation was most pronounced for proteins >100 kDa relative to both 2-DE/MS and normalized percentages. Both methods underestimated proteins with a *M*<sub>r</sub> <10 kDa, and furthermore, 2-DE gave a poor representation of proteins with *M*<sub>r</sub> >60 kDa.

Isoelectric Point Biases—To study biases of the two methods for certain pI ranges, the pI range from 3 to 13 was divided into bins of 1 pI unit, and proteins were assigned to these bins based on the calculated pI value of the unmodified polypeptide chain (Fig. 5b). The pI values were calculated by summing the pI values of the single amino acids and dividing them by the number of the amino acids (19). The relative distribution of the normalized mycobacterial proteins showed a low number of proteins in the pI range of 7–8, an observation that has also been described for several other organisms (20–22). Both methods overrepresented proteins between pI 4 and 6, with 2-DE/MS showing stronger representation for pI range 4–5 and ICAT-LC/MS for pI range 5–6. More than 50% of all proteins identified by either the ICAT-LC/MS or 2-DE/MS method had a calculated pI between 5 and 6, and more than 70% had a calculated pI between 5 and 7. Proteins of pI ranges >6 were underrepresented by both methods, and proteins of predicted pI values >11 were not detected at all.

Hydropathy Bias—Hydrophobic proteins often cause problems for protein analysis due to insolubility in commonly used solvents and loss by binding to surfaces. A measure for hydrophilicity and hydrophobicity is the grand average hydropathy (GRAVY) score, which is calculated as the arithmetic mean of the sum of the hydropathic indices of each amino acid in a protein (23). This index was used to compare the number of normalized proteins and those identified by ICAT-LC/MS and 2-DE/MS in specific hydrophatic ranges (Fig. 5c). Fifty-nine percent of the normalized mycobacterial proteins have a hydrophobic character, and 41% have a hydrophilic character. The hydropathy distribution has the shape of a Gaussian curve with a maximum of more than 50% of the...
normalized proteins concentrated in the neutral range between the values −0.2 and 0.2. Of the experimentally identified proteins 75% (ICAT-LC/MS) and about 60% (2-DE/MS) fell into this neutral range. 2-DE/MS slightly overestimated proteins in the hydrophobic range, whereas ICAT-LC/MS gave slight underestimation. Proteins in the strong hydrophilic range were underestimated by both 2-DE/MS and ICAT-LC/MS. These results reveal that 2-DE/MS has advantage for the identification of hydrophobic proteins and that both approaches had problems with the analysis of hydrophilic proteins.

Quantification of Differences in Protein Abundance between M. tuberculosis H37Rv and M. bovis BCG—Absolute quantification of proteins is a difficult task and at best achieved by amino acid composition analysis (24). Therefore, most proteomic experiments attempt to obtain relative quantitative information by determining the abundance ratios of proteins present in two samples. In the 2-DE/MS method this is achieved by comparing optical densities of matched spots in 2-DE gels, and in the ICAT-LC/MS method it is achieved by calculating the ratios of signal intensities for pairs of differentially isotopically labeled peptides. Comparing virulent with attenuated strains of the M. tuberculosis complex with the classical proteomics approach, we detected 32 spots only present in the virulent strains and obvious intensity differences by visual inspection (1, 2). Reproducibility of sample handling is improved by the ICAT-LC/MS approach due to the fact that aliquots from two biological samples are mixed directly after labeling and treated identically throughout the whole analysis procedure, the two samples thus serving as mutual internal standards. A higher number of proteins of different abundance in the two strains were detected by ICAT-LC/MS. Proteins showing intensity differences by more than 3-fold were extracted from Table II and are summarized in Table III. After quantification by the XPRESS tool, the calculated ratios were manually verified for the 280 proteins extracted, and if necessary, they were corrected. This analysis revealed three proteins as unique for the virulent H37Rv strain. These three proteins are potential vaccine candidates. Two of them, Rv0223c and Rv1513, were already predicted by genomic analyses (25) as belonging to regions deleted in BCG strains. Rv1513 is a conserved hypothetical protein of unknown function and is specific to M. tuberculosis but absent from the wild type M. bovis genome, as are genes of the flanking clusters of Rv1506c and Rv1516c. Even though a considerable number of flanking genes are annotated as “conserved hypotheticals,” this gene cluster may be involved in carbohydrate metabolism due to the fact

![Fig. 4. Comparison of proteins identified by ICAT-LC/MS and 2-DE/MS versus normalized proteins.](image)
FIG. 5. Bias of ICAT-LC/MS and 2-DE/MS against protein characteristics. The percentages of normalized, ICAT-LC/MS-identified, and 2-DE/MS-identified proteins are displayed according to their characteristics. a, molecular mass in 10-kDa increments. Proteins with a molecular mass smaller than 10 kDa and greater than 100 kDa were grouped in one column each. Theoretical molecular masses are extracted from the Sanger Institute database. b, pI values in one-step increments. Theoretical pI values are extracted from the Sanger database. c, hydropathy indices in 0.2-step increments. Hydropathy is displayed as the grand average hydropathy (GRAVY) score, which is calculated as the arithmetic mean of the sum of the hydrophatic indices of each amino acid. The left end of the scale represents hydrophobic proteins, and the right end represents hydrophilic proteins.
that Rv1516c is a sugar transferase homologue and that gmdA, epiA, and Rv1520, which code for a GDP-mannose 4,6-dehydratase, a nucleotide-sugar epimerase, and another putative sugar transferase, respectively, are part of the same cluster. Rv0223c is a probable aldehyde hydrogenase that is present in the genome of *M. bovis* wild type but not of the leprosy agent *Mycobacterium leprae*, a mycobacterial species with large genetic deletions when compared with *M. tuberculosis* (26). The other variant, Rv0570, was newly detected by ICAT-LC/MS and is a ribonucleotide reductase present in the genome of wild type *M. bovis* but absent from *M. leprae*. Furthermore, 20 proteins were increased and 23 decreased by more than 3-fold in *M. tuberculosis* H37Rv compared with *M. bovis* BCG (Table III). The genes of these proteins, of which only a small subset was also detected by the 2-DE/MS approach, are present in BCG, although probably down-regulated or not induced by so far unknown regulatory mechanisms active in *M. tuberculosis* under the same culture conditions. Furthermore, because this analysis was performed with mycobacterial samples from one type of growth condition, i.e. logarithmic growth phase in 7H9 fully complemented medium, the proteins down- or up-regulated in BCG in comparison to *M. tuberculosis* may not be controlled in the same manner in either of the two mycobacterial strains. The proteins up-regulated in BCG in comparison with *M. tuberculosis*, although present in the latter strain, are of various functional categories. Of particular interest were the proteins coded by genes Rv2935 and Rv2940c, which belong to a cluster of genes involved in glycolipid synthesis including synthesis of polyketides and mycocerosates, and Rv1527c, which is located in a cluster of genes involved in polyketide and lipid synthesis. These data suggest that under the specific growth conditions used in this study the lipid synthesis pathway in BCG is regulated differently than that of *M. tuberculosis*. As a certain mycocerosate and its synthesis machinery have been implicated in virulence in *M. tuberculosis*, these data are of potential clinical importance (27).

Comparison of Quantification by ICAT-LC/MS and 2-DE/MS—Because the number of proteins with different abundance detected by both the 2-DE/MS and ICAT-LC/MS methods was low (Table III), we searched to explain this observation. One reason may be that the number of identified proteins is low in comparison to the expressed proteins in *M. tuberculosis* (Fig. 2) so that the chance of finding numerous proteins common to both datasets is low. To search for further reasons we analyzed the 27 proteins both approaches identified as quantitatively regulated. In 2-DE gels the spots of these 27 proteins were quantified with the help of the evaluation software TOPSPOT. The intensity values of spots from gels derived from four independent sample preparations for each mycobacterial strain were averaged. Accepting a divergence of 30%, only eight of the 27 proteins showed the same intensity difference detected by 2-DE and ICAT-LC/MS (Table IV). One example is shown in Fig. 6. Enoyl-CoA hydratase (Rv0905) showed an intensity relationship of 1:1.12 and 1:1.1 for 2-DE/MS and ICAT-LC/MS, respectively. Thus far, this protein was only found in one spot of each 2-DE pattern. The protein was identified with a sequence coverage of 46%, and
the peptides of the protein covered 62% of the peptide mass fingerprint. All of the five most intense peaks belonged to Rv0905. The remaining peaks comprised contamination from matrix, trypsin, or stain. No additional protein reached an acceptable scoring factor, suggesting that Rv0905 represents the only protein in this spot (28). These data indicate that for the simple situation encountered for Rv0905, i.e. probable single spot pattern and no co-migrating proteins, the two methods provided comparable results.

More complex situations resulted in a divergence of the ratios obtained by the two methods. Four spots contained more than one protein. Here, only ICAT-LC/MS allowed us to

| Open reading frame | Protein name | No. | BCG | H37Rv |
|--------------------|--------------|-----|-----|-------|
| Rv0570             | Ribonucleotide reductase, class II | 1   | +   |       |
| Rv1513             | Conserved hypothetical protein*   | 1   | +   |       |
| Rv0223c            | Aldehyde dehydrogenase (possible betB)* | 2   |     |       |
| Rv2632c            | Conserved hypothetical protein    | 2   | 1   | 10.0  |
| Rv3089             | Acyl-CoA synthase                  | 1   | 1   | 8.3   |
| Rv2187             | Acyl-CoA synthase                  | 1   | 1   | 7.6   |
| Rv1181             | Polypeptide synthase               | 4   | 1   | 7.6   |
| Rv0233             | Ribonucleoside-diphosphate reductase B2 | 1   | 1   | 7.1   |
| Rv2161c            | Similar to alkanal monooxygenase β chain | 5   | 1   | 6.3   |
| Rv1180             | Polypeptide synthase               | 2   | 1   | 5.0   |
| Rv0005             | DNA gyrase subunit B               | 1   | 1   | 5.0   |
| Rv0751c            | Methylmalonate semialdehyde oxidoreductase | 1   | 1   | 5.0   |
| Rv3086             | Zinc-containing alcohol dehydrogenase | 1   | 1   | 5.0   |
| Rv3726             | Putative alcohol dehydrogenase, zinc-type | 4   | 1   | 5.0   |
| Rv2557             | Conserved hypothetical protein     | 23  | 1   | 4.0   |
| Rv0120c            | Elongation factor G               | 1   | 1   | 3.3   |
| Rv0458             | Aldehyde dehydrogenase            | 2   | 1   | 3.3   |
| Rv0888             | Possible membrane protein          | 1   | 1   | 3.3   |
| Rv1640c            | C-term lysyl-tRNA synthase         | 1   | 1   | 3.3   |
| Rv1925             | Acyl-CoA synthase                  | 1   | 1   | 3.3   |
| Rv2135c            | Conserved hypothetical (= Q49773 B2126 C1 148) | 1   | 1   | 3.3   |
| Rv2289             | CDP-diacylglycerol phosphatidylhydroxase | 1   | 1   | 3.3   |
| Rv2575             | Conserved hypothetical protein     | 1   | 1   | 3.3   |

| Open reading frame | Protein name | No. | BCG | H37Rv |
|--------------------|--------------|-----|-----|-------|
| Rv2429             | Alkyl hydroperoxide reductase       | 2   | 35.0| 1     |
| Rv0192             | Conserved hypothetical protein      | 1   | 25.0| 1     |
| Rv0663             | Proable arylsulfatase               | 1   | 16.9| 1     |
| Rv0216             | Conserved hypothetical protein      | 4   | 10.3| 1     |
| Rv3273             | C-term similar to carbonic anhydrase| 7   | 9.2 | 1     |
| Rv0928             | PstS component of phosphate uptake  | 2   | 6.2 | 1     |
| Rv2987c            | 3-Isopropylmalate dehydratase small subunit | 1   | 5.1 | 1     |
| Rv3130c            | Conserved hypothetical protein      | 1   | 5.1 | 1     |
| Rv2935             | Phenolphthiocerol synthesis (pksF)  | 3   | 4.9 | 1     |
| Rv1998c            | Conserved hypothetical protein      | 3   | 4.9 | 1     |
| Rv1462             | Conserved hypothetical protein      | 2   | 4.2 | 1     |
| Rv2971#            | Oxoreductase of Aldo/keto reductase family | 6   | 4.0 | 1     |
| Rv1527c            | Polypeptide synthase                | 3   | 3.5 | 1     |
| Rv0315             | Probable b-1.3-glucanase            | 3   | 3.6 | 1     |
| Rv3859c            | Ferredoxin-dependent glutamate synthase | 1   | 3.3 | 1     |
| Rv1483             | 3-Oxoacyl-[ACP] reductase (also known as MabA) | 2   | 3.2 | 1     |
| Rv1658             | Arginosuccinate synthase            | 2   | 3.2 | 1     |
| Rv1093#            | Serine hydroxymethyltransferase     | 2   | 3.0 | 1     |
| Rv1812c            | Probable dehydrogenase              | 1   | 3.0 | 1     |
| Rv2940c            | Mycocerosic acid synthase           | 6   | 3.0 | 1     |
| Rv3045#            | Alcohol dehydrogenase               | 2   | 3.0 | 1     |
| Rv0547c            | Putative oxidoreductase             | 1   | 3.0 | 1     |
| Rv3396c            | GMP synthase                        | 2   | 3.0 | 1     |
Complementary Analysis of *M. tuberculosis* by 2-DE and ICAT

Table IV: Common proteins between ICAT-LC/MS and 2-DE/MS

| BCG | H37Rv | BCG | H37Rv | Open reading frame | More than one protein in one 2-DE spot (ICAT ratio) | More than one protein in 2-DE gel | Same intensity relation between 2-DE and ICAT |
|-----|-------|-----|-------|-------------------|-----------------------------------------------|-------------------------------|-----------------------------------|
| 1   | 1.00  | 1   | 0.7   | Rv0211            | + Rv2115c (1:1.2)                              | No                            | Yes                               |
| 1   | 0.92# | 1   |       | Rv0462           | No                                             | Yes                           | NA                                |
| 1   | 0.75# | 1   |       | Rv0462           | No                                             | Yes                           | NA                                |
| 1   | (0.86) | 1   | 0.5   | Rv0462           | No                                             | Yes                           | No                                |
| 1   | 1.28  | 1   | 1     | Rv0642c          | No                                             | No                            | Yes                               |
| 1   | 0.60  | 1   | 0.5   | Rv0815c          | No                                             | No                            | Yes                               |
| 1   | 1.12  | 1   | 1.1   | Rv0905           | No                                             | No                            | Yes                               |
| 1   | 0.41# | 1   |       | Rv0952           | No                                             | Yes                           | NA                                |
| 1   | 1.06# | 1   |       | Rv0952           | No                                             | Yes                           | NA                                |
| 1   | (1.00) | 1   | 0.7   | Rv0952           | No                                             | Yes                           | No                                |
| 1   | 1.25  | 1   | 1.1   | Rv1074c          | No                                             | Yes                           | NA                                |
| 1   | 1.01  | 1   | 1.1   | Rv1392           | No                                             | No                            | Yes                               |
| 1   | 1.36  | 1   | 0.5   | Rv1416           | + Rv3678c (1:1)                                | No                            | No                                |
| 1   | 1.73  | 1   | 0.8   | Rv1438           | No                                             | No                            | No                                |
| 1   | 1.20  | 1   |       | Rv1794           | No                                             | No                            | Yes                               |
| 1   | 1.05  | 1   | 0.6   | Rv1932           | No                                             | No                            | No                                |
| 1   | 1.22  | 1   | 0.6   | Rv2246           | No                                             | No                            | No                                |
| 1   | 1.29  | 1   | 0.7   | Rv2468c          | + Rv2031c                                     | No                            | No                                |
| 1   | 1.53  | 1   | 0.6   | Rv2521           | No                                             | No                            | No                                |
| 1   | 1.10  | 1   | 0.6   | Rv2632           | No                                             | No                            | No                                |
| 1   | 1.41# | 1   |       | Rv2626c         | No                                             | Yes                           | NA                                |
| 1   | 0.88# | 1   |       | Rv2626c         | No                                             | Yes                           | NA                                |
| 1   | 0.25# | 1   |       | Rv2626c         | No                                             | Yes                           | NA                                |
| 1   | 1.88# | 1   |       | Rv2626c         | No                                             | Yes                           | NA                                |
| 1   | 0.33# | 1   |       | Rv2626c         | No                                             | Yes                           | NA                                |
| 1   | 1.92# | 1   |       | Rv2626c         | No                                             | Yes                           | NA                                |
| 1   | 1.05# | 1   |       | Rv2626c         | No                                             | Yes                           | NA                                |
| 1   | (0.91) | 1   | 0.4   | Rv2626c         | NA                                             | Yes                           | No                                |
| 1   | 0.72  | 1   | 0.2   | Rv2971           | No                                             | No                            | No                                |
| 1   | 0.52# | 1   |       | Rv3042           | No                                             | Yes                           | NA                                |
| 1   | 0.83# | 1   |       | Rv3045           | No                                             | Yes                           | NA                                |
| 1   | (0.72) | 1   | 0.3   | Rv3045           | No                                             | Yes                           | No                                |
| 1   | 0.32  | 1   | 0.8   | Rv3075c         | + Rv0389                                      | No                            | Yes                               |
| 1   | 1.18  | 1   | 0.6   | Rv3133c          | No                                             | No                            | No                                |
| 1   | 0.18  | 1   | 0.6   | Rv3224           | No                                             | No                            | No                                |
| 1   | 1.05  | 1   | 1.4   | Rv3274c          | No                                             | No                            | No                                |
| 1   | 1.36  | 1   | 0.8   | Rv3389c          | No                                             | No                            | No                                |
| 1   | 1.02# | 1   |       | Rv3411c         | No                                             | Yes                           | NA                                |
| 1   | 1.46# | 1   |       | Rv3411c         | No                                             | Yes                           | NA                                |
| 1   | (1.21) | 1   | 0.7   | Rv3411c         | NA                                             | Yes                           | No                                |
| 1   | 3.25  | 1   | 2     | Rv3804c         | No                                             | No                            | No                                |

Distinguish which one(s) of the proteins present in the spot changed their abundance. Therefore, quantification by 2-DE of spots comprising a mixture of two or more proteins is not useful unless also stable isotope-tagging methods are being used (17). One spot with an intensity ratio of 1:1 between BCG and H37Rv contained two proteins, phosphoenolpyruvate carboxykinase (Rv0211) and an ATPase of ATP:ADP antiporter family (Rv2115c) (Fig. 7). ICAT-LC/MS analysis revealed a ratio of 1:0.68 for Rv0211 and 1:1.22 for Rv2115c, clearly distinguishing between these two proteins. In these four cases only ICAT-LC/MS allowed us to quantify the protein amount.

For five genes, the corresponding proteins were detected in different spots. In these cases, each protein species could be independently quantified by the 2-DE/MS method, whereas the ICAT-LC/MS method quantified the sum of the protein species as a single protein. Succinyl-CoA synthase α chain (Rv0952) was identified in three 2-DE spots (Fig. 8). Each of them had a different intensity relationship between BCG and H37Rv of 1:0.41, 1:1, and −/+/. ICAT-LC/MS did not distinguish between these three protein species and calculated a 1:0.7 relationship. Adding up the intensities of all 2-DE spots containing Rv0952 a ratio of 1:1 was calculated between BCG and H37Rv assimilating the ICAT-LC/MS result. These cases exemplify the potential of 2-DE to separate proteins at the protein species level, a capability that is usually not provided by the ICAT-LC/MS method. This is of particular importance if differential post-translational modification leads to different
electrophoretic mobility. For a further 10 proteins for which the ratios determined by 2-DE/MS and ICAT-LC/MS differed, the reasons for the discrepancy are presently unknown. It is likely that additional data, e.g. the identification of additional spots representing a particular gene product or the detection of additional proteins in specific 2-DE gel spots, will help to clarify the apparent discrepancies.

These data clearly show that the 2-DE/MS and ICAT-LC/MS cannot be expected to and in fact do not provide identical quantitative results. Only in cases where a protein exists as a single protein species or where all protein species are represented by one spot containing only one protein will quantification in both methods be comparable. This was only true for 22% of the cases investigated.

**DISCUSSION**

As was shown for the ribosomal proteome the resolution of a complete proteome is feasible provided the complexity is not too high (5). Problems with dynamic range of protein concentration, sensitivity, and resolution increase with the complexity of the samples to be analyzed. In addition, intrinsic properties of certain classes of proteins can make their separation and/or detection difficult. Limitations for the analysis of basic and hydrophobic proteins by 2-DE have been described previously (5, 20). The presence of cysteine in the protein is a prerequisite for ICAT analysis as proteins lacking cysteines cannot be quantified. From the mycobacterial genome 81% of the proteins contain at least one cysteine. Including the limitation by the MS measurement to peptides between $M_r$ 500 and 4,000 the percentage of accessible proteins is reduced to 75%. Cysteine content in the protein subclasses chaperones/heat-shock proteins and PE and PPE families is significantly reduced (4). Only five of 16 predicted chaperones/heat-shock proteins contain at least one cysteine. Heat-shock proteins like DnaK, GroEL2, GroES, and HspX are represented as intense spot series by 2-DE but are not present within the ICAT-labeled proteins. Instead, LC/MS peptides of these proteins were found with high scores within the unlabeled fraction of the 60,000 peptides. Proteins of the PE and PPE families were not identified by either 2-DE/MS or ICAT-LC/MS. Only 42 of 167 proteins of the PE and PPE families subclass contain at least one cysteine. Indeed, 26 proteins of the PE and PPE families were detected within the 200 best scores of the unlabeled proteins. These 26 proteins have a mean $M_r$ of 88,961 and are therefore not in the optimal $M_r$ range for 2-DE. However, 12 of these proteins have an $M_r < 60$ kDa showing that the $M_r$ cannot be the only reason for
excluding them in the 2-DE/MS analysis. The mean pI value of 4.39 and the hydropathy values also should not cause problems for 2-DE. For this protein class LC/MS seems to be of advantage compared with 2-DE/MS. Because of the low amount of cysteines in this class of proteins, ICAT labeling is not of advantage here, and quantification is not possible.

Fig. 4 illustrates lack of the subclasses polyketide and non-ribosomal peptide synthesis, central intermediary metabolism, and degradation of macromolecules by 2-DE/MS analysis, whereas by ICAT-LC/MS all of these three subclasses were overrepresented. The subclass polyketide and non-ribosomal peptide synthesis contained 41 proteins, and 21 of them had an M_r >100,000. This may be the reason for absence within the 2-DE patterns and overrepresentation in ICAT-LC/MS. The other two protein subclasses contain only one protein with an M_r >100,000, leaving the reason unresolved as to why these subclasses were not represented by 2-DE/MS.

Membrane proteins were also underrepresented by the ICAT-LC/MS approach, although an earlier report described the analysis of 491 microsomal proteins and therefore showed compatibility of the method with the analysis of poorly soluble proteins (14). The underrepresentation of membrane proteins in this study is therefore most likely explained by the protein solubilization conditions used. Underrepresented protein classes other and unknowns could contain falsely assigned genes. We have shown by 2-DE/MS (29) that gene prediction programs have neglected six genes in *M. tuberculosis*. The only way to avoid falsely predicted genes is to accept a gene as such only after obtaining evidence for its existence at the protein level. Genes in the genome lists should be annotated by a comment “confirmed at the protein level” with reference.

Proteins can only be accurately quantified by 2-DE/MS if the spot intensity is in the linear range of the staining method used and if the spot consists only of one protein species. Furthermore, we show that quantitative results obtained by the two methods cannot be compared directly. The ICAT-LC/MS method quantifies protein composition, whereas the 2-DE/MS method quantifies each protein species individually. Therefore, ICAT-LC/MS and 2-DE/MS are complementary.
Quantitative proteomic methods, and the optimal use of either method depends on the biological objectives. It should be noted that both methods are undergoing significant improvements. To avoid the quantification problems caused by non-parallel separation and identification in the classical 2-DE/MS proteomics an ICAT-2-DE/MS strategy was developed (17). The advantages of protein species separation are combined with parallel quantification and quantification of several proteins within one spot. ICAT-2-DE/MS will not allow the quantification of large proteins by the parallel process and needs again to be complemented by ICAT-LC/MS. Fluorescence labeling by difference gel electrophoresis (30) exchanges optical density measurements by fluorescence measurements, hereby increasing the dynamic range of optical measurements. Experience will show whether the labeling process with difference gel electrophoresis, ICAT, or others can be performed with 100% yield for all proteins. Maybe even these methods have to be accepted to be complementary because of the different amino acids that are labeled. The ICAT-LC/MS method has been advanced by the development of second generation isotope-tagging reagents that are characterized by the use of 12C/13C as the stable isotope, resulting in the elimination of the chromatographic shift of deuterated tags and the introduction of an acid-cleavable linker that leads to improved MS/MS spectral quality due to the reduced size of the isotope tag. Furthermore, advanced software tools for the analysis of LC-MS/MS data have been developed that significantly accelerate the pace and increase the consistency of data analysis (31, 32). Collectively, these recent technical advances suggest that the performance of quantitative proteomic technologies is rapidly approaching the stage at which the routine and complete analysis of cellular proteomes will become reality.

Acknowledgments—K. P. Pleissner and T. Eifert are acknowledged for excellent help in bioinformatics.

2 X.-J. Li, H. H. Zhang, J. A. Ranish, and R. Aebersold, submitted for publication.
**Complementary Analysis of *M. tuberculosis* by 2-DE and ICAT**

1. Jungblut, P. R., Schaible, U. E., Mollenkopf, H.-J., Zimny-Arndt, U., Rau-nach, B., Mattow, J., Halada, P., Lamer, S., and Kaufmann, S. H. E. (1999) Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of mycobacterial pathogens. *Mol. Microbiol.* 33, 1103–1117.

2. Mattow, J., Jungblut, P. R., Schaible, U. E., Mollenkopf, H.-J., Lamer, S., Hagens, K., Müller, E.-C., and Kaufmann, S. T. H. (2001) In search for a novel tuberculosis vaccine: identification of proteins from *Mycobacterium tuberculosis* missing in *Mycobacterium bovis* BCG strains. *Electrophoresis* 22, 2936–2946.

3. Kaufmann, S. H. E. (2000) Is the development of a new tuberculosis vaccine possible? *Nat. Med.* 6, 955–960.

4. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eigelemeier, K., Gas, S., Barry, C. E., III, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentle, S., Hamlin, N., Holroyd, S., Hornsbys, T., Jagels, K., Barrett, D. M., et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.

5. Jungblut, P., Thiede, B., Zimny-Arndt, U., Muller, E. C., Scheler, C., Wittmann-Liebold, B., and Otto, A. (1996) Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis* 17, 839–847.

6. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using iso-topo-coded affinity tags. *Nat. Biotechnol.* 17, 994–999.

7. Ideker, T., Thorsson, V., Rish, J. A., Christiansen, R., Buhrer, J., Eng, K. J., Burman, G., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934.

8. Shiio, Y., Donohoe, S., Yi, E. C., Goodlett, D. R., Aebersold, R., and Eisenman, R. N. (2002) Quantitative proteomic analysis of Myc oncoprotein function. *EMBO J.* 21, 5088–5096.

9. Guina, T., Purvine, S. O., Yi, E. C., Eng, J., Goodlett, D. R., Aebersold, R., and Eng, J. (2003) The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* 33, 349–355.

10. Baliga, N. S., Pan, M., Goo, Y. A., Yi, E. C., Goodlett, D. R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W. V., and Hood, L. (2002) Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14913–14918.

11. von Haller, P. D., Yi, E., Donohoe, S., Vaughn, K., Keller, A., Nesvizhskii, A. I., Eng, J., Li, X. J., Goodlett, D. R., Aebersold, R., and Watts, J. D. (2003) The application of new software tools to quantitative protein profiling via ICAT and tandem mass spectrometry. I. Statistically annotated datasets for peptide sequences and proteins identified via the application of ICAT and tandem mass spectrometry to proteins co-purifying with T cell lipid rafts. *Mol. Cell Proteomics* 2, 426–427.

12. Arur, S., Uche, U. E., Rezau, K., Fang, M., Scranlon, V., Cowan, A. E., Mohler, W., and Han, D. K. (2003) Annexin I is an endogenous ligand that mediates apoptotic cell engulfment. *Dev. Cell* 4, 587–598.

13. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced micosomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* 19, 946–951.

14. Griffin, T. J., Gygi, S. P., Rist, B., and Aebersold, R. (2001) Quantitative proteomic analysis by using MALDI quadrupole time-of-flight mass spectrometer. *Anal. Chem.* 73, 978–986.

15. Smolka, M. B., Zhou, H., Purkayastha, S., and Aebersold, R. (2001) Optimization of the isotope-coded affinity tagging-labeling procedure for quantitative proteome analysis. *Anal. Biochem.* 297, 25–31.

16. Smolka, M., Zhou, H., and Aebersold, R. (2002) Quantitative protein profiling using two-dimensional gel electrophoresis, isotope-coded affinity tag labeling, and mass spectrometry. *Mol. Cell Proteomics* 1, 19–29.

17. Pleissner, K.-P., Elfer, T., and Jungblut, P. R. (2002) A European pathogenic microorganism proteome database: construction and maintenance. *Comp. Funct. Genomics* 3, 97–100.

18. Rickard, E. C., Strohl, M. M., and Nielsen, R. G. (1991) Correlation of electrophoretic mobilities from capillary electrophoresis with physicochemical properties of proteins and peptides. *Anal. Biochem.* 197, 197–207.

19. Büttner, K., Bernhardt, J., Scharf, C., Schmid, R., Mader, U., Eymann, C., Antelmann, H., Völker, U., and Hecker, M. (2001) A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*. *Electrophoresis* 22, 2908–2935.

20. VanBogelen, R. A., Abshire, K. Z., Moldover, B., Olson, E. R., and Neidhardt, F. C. (1997) Escherichia coli proteome analysis using the gene-protein database. *Electrophoresis* 18, 1243–1251.

21. Link, A. J., Hays, L. G., Carmack, E. B., and Yates, J. R., III (1997) Identifying the major proteome components of Haemophilus influenzae type-strain NCTC 8143. *Electrophoresis* 18, 1314–1334.

22. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.

23. Jungblut, P., Dzonara, M., Klose, J., and Wittmann-Liebold, B. (1992) Identification of tissue proteins by amino acid analysis after purification by two-dimensional electrophoresis. *J. Protein Chem.* 11, 603–612.

24. Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schnolick, G. K., Rane, S., and Small, P. M. (1999) Comparative genomics of *BCG* vaccines by whole genome DNA microarrays. *Science* 284, 1520–1523.

25. Cole, S. T., Eigelemeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., Ungallal, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsbys, T., Jagels, K., Barrett, D. M., et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.

26. Cox, J. S., Chen, B., McNeil, M., and Jacobs, W. R. (1999) Complex lipid determination tissue specific replication of *Mycobacterium tuberculosis* in mice. *Nature* 402, 79–83.

27. Schmidt, F., Schmid, M., Mattow, J., Pleissner, K. P., and Jungblut, P. R. (2003) Iterative procedure to improve peptide mass fingerprint identifi-cation of two-dimensional electrophoresis separated proteins. *J. Am. Soc. Mass Spectrom.* 14, 943–956.

28. Jungblut, P. R., Müller, E.-C., Mattow, J., and Kaufmann, S. H. E. (2001) Proteomics reveals open reading frames in *Mycobacterium tuberculosis* not predicted by genomics. *Infect. Immun.* 69, 5905–5907.

29. Uenue, M., Morgan, M. E., and Minden, J. S. (1997) Difference gel elec-trophoresis: a single method for detecting changes in protein extracts. *Electrophoresis* 18, 2071–2077.

30. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392.

31. Keller, A., Purvine, S., Nesvizhskii, A. I., Stolyar, S., Goodlett, D. R., and Kolker, E. (2002) Experimental protein mixture for validating tandem mass spectral analysis. *OMICS* 6, 207–212.