Anal. Chem. **2004,** *76,* 5769−5776

Proteomic Profiling of Intact Mycobacteria by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry

Justin M. Hettick,*,† Michael L. Kashon,† Janet P. Simpson,† Paul D. Siegel,† Gerald H. Mazurek,‡ and David N. Weissman†

National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, 1095 Willowdale Road, Morgantown, West Virginia 26505, and National Center for HIV, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

Current methods for the identification of mycobacteria in culture are time-consuming, requiring as long as 12 weeks for positive identification. One potential approach to rapid mycobacterial identification is to utilize proteomic profiling of cultures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). In this report, we have applied MALDI-TOF MS to proteomic profiling of cultured microorganisms representing six species of the genus *Mycobacterium*. We find that analysis of acetonitrile/trifluoroacetic acid cellular extracts produces data similar to that of the analysis of deposited whole cells, while minimizing human contact with the microorganisms and rendering them nonviable. A matrix composition of α-cyano-4-hydroxycinnamic acid with fructose yields highly reproducible MALDI-TOF spectra. Statistical analysis of MALDI-TOF MS data allows differentiation of each individual mycobacterial species on the basis of unique mass fingerprints. The methodology allows identification of a number of unique (potentially diagnostic) biomarkers as targets for protein identification by MS/MS experiments. In addition, we observe a number of signals common to all mycobacterial species studied by MALDI-TOF MS, which may be genus-specific biomarkers. The potentially genus-specific biomarkers occur at low mass (<2 kDa) and are likely to be lipids and cell wall components such as mycolic acids. This study demonstrates the potential for mass spectrometry-based identification/classification of mycobacteria.

It is estimated that more than one-third of the worldwide human population is infected with *Mycobacterium tuberculosis*, and ∼1.8 million die of the disease each year.[1,2] In low-prevalence countries such as the United States, it is often important to differentiate between infection with *M. tuberculosis* and other (nontuberculous) mycobacteria. Thus, improved diagnostic methodologies resulting in rapid and accurate mycobacterial identification would be highly desirable. Current approaches to early identification of *M. tuberculosis* in culture include polymerase chain reaction[3] and the analysis of derivatized cell wall components, such as mycolic acids, by high-performance liquid chromatography (HPLC).[4,5]

Mass spectrometry is poised to make important contributions to the field of chemotaxonomy. With the development of matrix-assisted laser desorption/ionization (MALDI)[6,7] and electrospray ionization,[8] as well as improvements in time-of-flight mass spectrometry (TOF MS),[9] it has become feasible to produce and analyze gas-phase ions of biological molecules with high resolution and mass accuracy. Experiments such as peptide mass mapping[10,11] and sequencing by tandem mass spectrometry (MS/MS)[12] are increasingly important to protein identification experiments.[13]

Recently, there has been significant interest in the mass spectrometric analysis of bacteria (both intact and lysed), and several reviews have been presented.[14−16] The possibility of using pattern recognition software for sensitive, rapid, and automated identification of intact microorganisms on the basis of mass spectrometry data is an attractive goal, and a number of studies have been published demonstrating the identification of various intact bacteria by MALDI.[17−30] However, other researchers have demonstrated that the profiles generated by this methodology can vary significantly with changes in sample preparation; therefore,

* To whom correspondence should be addressed. E-mail: ayf2@cdc.gov. Phone: (304) 285-6065. Fax: (304) 285-6126.
† National Institute for Occupational Safety and Health.
‡ National Center for HIV, STD and TB Prevention.

(1) Dye, C.; Scheele, S.; Dolin, P.; Pathania, V.; Raviglione, M. C. *JAMA, J. Am. Med. Assoc.* **1999**, *282*, 677−686.
(2) Iademarco, M. F.; Castro, K. G. *Semin. Respir. Infect.* **2003**, *18*, 225−240.
(3) Eisenach, K. D.; Cave, M. D.; Bates, J. H.; Crawford, J. T. *J. Infect. Dis.* **1990**, *161*, 977−981.
(4) Butler, W. R.; Jost, K. C., Jr.; Kilburn, J. O. *J. Clin. Microbiol.* **1991**, *29*, 2468−2474.
(5) Hagen, S. R.; Thompson, J. D. *J. Chromatogr., A* **1995**, *692*, 167−172.
(6) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299−2301.
(7) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom.* **1987**, *78*, 53−68.
(8) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Anal. Chem.* **1985**, *57*, 675−679.
(9) Cotter, R. J. *Anal. Chem.* **1999**, *71*, 445A−451A.
(10) Yates, J. R., III.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397−408.
(11) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(12) Biemann, K. In *Methods in Enzymology*; McCloskey, J. A., Ed.; Academic Press: San Diego, CA, 1990; Vol. 193, pp 886−887.
(13) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837−845.
(14) Lay, J. O., Jr. *Mass Spectrom. Rev.* **2001**, *20*, 172−194.
(15) van Baar, B. L. M. *FEMS Microbiol. Rev.* **2000**, *24*, 193-219.
(16) Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001**, *20*, 157−171.

10.1021/ac049410m Not subject to U.S. Copyright. Publ. 2004 Am. Chem. Soc.
Published on Web 09/03/2004

optimization of whole-cell MALDI experiments has been an active area of research.[26−32] In addition, MALDI-TOF mass spectra of bacteria can vary with the growth medium selected. Although the experiments have demonstrated that unique biomarkers exist regardless of growth medium,[22] it is important to keep experimental conditions constant when species identification by mass spectrometry is attempted. Alternatively, several groups have developed HPLC electrospray mass spectrometry methods for the characterization of bacterial cell lysates, extracts, or both.[33−37]

The purpose of the current study is to utilize MALDI-TOF MS for rapid identification of several mycobacterial species. In particular, we are interested in establishing the groundwork for differentiating *M. tuberculosis* from other mycobacterial strains, as rapid, sensitive, specific methodologies for the positive identification of this pathogen are critical to stemming the spread of disease worldwide. To do so, it is necessary to optimize the MALDI experiment such that spectra (1) contain maximum information and (2) are highly reproducible. Many papers in the literature are devoted to proteomic profiling of bacteria; however, there seems to be no consensus on whether direct analysis of the whole cell (involving codeposition of bacteria with the MALDI matrix)[19−23,26−28] or analysis of a cell extract[25,29,30] is preferable. In the current report, we find that analysis of cell extracts produces data similar to the analysis of whole cells. Furthermore, since the extraction procedure inactivates mycobacteria, analysis of extracts minimizes pathogen exposures in the mass spectrometry laboratory.

(17) Welham, K. J.; Domin, M. A.; Scannell, D. E.; Cohen, E.; Ashton, D. S. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 176−180.
(18) Holland, R. D.; Wilkes, J. G.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O., Jr. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1227−1232.
(19) Claydon, M. A.; Davey, S. N.; Edwards-Jones, V.; Gordon, D. B. *Nat. Biotechnol.* **1996**, *14*, 1584−1586.
(20) Bright, J. J.; Claydon, M. A.; Soufian, M.; Gordon, D. B. *J. Microbiol. Methods* **2002**, *48*, 127−138.
(21) Lynn, E. C.; Chung, M.; Tsai, W.; Han, C. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2022−2027.
(22) Bernardo, K.; Pakulat, N.; Macht, M.; Krut, O.; Seifert, H.; Fleer, S.; Hunger, F.; Kronke, M. *Proteomics* **2002**, *2*, 747−753.
(23) Jarman, K. H.; Daly, D. S.; Petersen, C. E.; Saenz, A. J.; Valentine, N. B.; Wahl, J. H. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1586−1594.
(24) Krishnamurthy, T.; Ross, P. L. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1992−1996.
(25) Haag, A. M.; Taylor, S. N.; Johnston, K. H.; Cole, R. B. *J. Mass Spectrom.* **1998**, *33*, 750−756.
(26) Williams, T. L.; Andrzejewski, D.; Lay, J. O., Jr.; Musser, S. M. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 342−351.
(27) Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Gantt, S. L.; Jarman, K. H.; Kingsley, M. T.; Wahl, K. L. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1580−1585.
(28) Gantt, S. L.; Valentine, N. B.; Saenz, A. J.; Kingsley, M. T.; Wahl, K. L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1131−1137.
(29) Domin, M. A.; Welham, K. J.; Ashton, D. S. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 222−226.
(30) Wang, Z.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 456−464.
(31) Ramirez, J.; Fenselau, C. *J. Mass Spectrom.* **2001**, *36*, 929−936.
(32) Smole, S. C.; King, L. A.; Leopold, P. E.; Arbeit, R. D. *J. Microbiol. Methods* **2002**, *48*, 107−115.
(33) Krishnamurthy, T.; Davis, M. T.; Stahl, D. C.; Lee, T. D. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 39−49.
(34) Vaidyanathan, S.; Rowland, J. J.; Kell, D. B.; Goodacre, R. *Anal. Chem.* **2001**, *73*, 4134−4144.
(35) Ho, Y. P.; Hsu, P. H. *J. Chromatogr., A* **2002**, *976*, 103−111.
(36) Williams, T. L.; Leopold, P. E.; Musser, S. M. *Anal. Chem.* **2002**, *74*, 5807−5813.
(37) Williams, T. L.; Callahan, J. H.; Monday, S. R.; Feng, P. C. H.; Musser, S. M. *Anal. Chem.* **2004**, *76*, 1002−1007.

EXPERIMENTAL SECTION

Reagents. Middlebrook 7H9 broth and OADC media additive were acquired from Becton Dickinson (Sparks, MD) and prepared according to manufacturer's instructions. The peptides/proteins bradykinin, melittin (bee venom), insulin (bovine), cytochrome *c* (equine), and β-casein (bovine) were acquired from Sigma Chemical Co. (St. Louis, MO). Fructose, α-cyano-4-hydroxycinnamic acid (CHCA), and sinapic acid (SPA) were acquired from Aldrich Chemical Co. (Milwaukee, WI). Distilled deionized water (DDI; 18 MΩ) was produced by a Milli-Q system (model A10, Millipore, Billerica, MA) and filtered through a 0.22-μm filter (Millipak 40, Millipore).

All reagents were used as received, with the exception of CHCA, which was recrystallized as follows: ∼300 mg of CHCA was added to 250 mL of DDI and the resultant mixture vigorously stirred. Concentrated NH_4OH was added until the CHCA was dissolved. CHCA was precipitated by titrating the solution to pH ∼2 with 1.0 N HCl and recovered by vacuum filtration. The precipitate was washed 3 times with 1% HCl (cold).

Cell culture broth was prepared as follows: 67.5 mL Middlebrook 7H9 broth was added to 100-mL culture bottles (Fisher Scientific, Pittsburgh, PA) and autoclaved for 20 min, followed by addition of 7.5 mL of Middlebrook OADC.

Stock solutions of MALDI matrixes and additive were prepared as follows: CHCA (30 mg/mL in methanol), SPA (saturated in methanol), and fructose (22 mg/mL in methanol).

Mycobacteria. *Mycobacterium fortuitum* (ATCC 6841), *Mycobacterium kansasii* (ATCC 12478), *Mycobacterium intracellulare* (ATCC 13950), *M. tuberculosis* H37Ra (ATCC 25177), and *Mycobacterium avium* (ATCC 700736) were acquired from the American Type Culture Collection (Manassas, VA). *Mycobacterium bovis* BCG (Bacillus Calmette-Guerin, Pasteur, TMCC 1011) was provided by Prof. Ian Orm of Colorado State University (Fort Collins, CO). Long-term storage of cells was performed at −80 °C.

Inocula were prepared by suspending a small loop of cells in PBS/0.5% Tween 80. For the intraculture studies, 75-mL cultures were performed in 100-mL culture bottles incubated at 37 °C on an incubator shaker (model C24KC, New Brunswick Scientific, Edison, NJ). For the interculture studies, 15-mL cultures were performed in 18 × 150 mm culture tubes and incubated at 37 °C on a culture roller (model TC-7, New Brunswick Scientific). Cultures of *M. fortuitum*, *M. kansasii*, *M. intracellulare*, *M. tuberculosis* H37Ra, *M. bovis* BCG, and *M. avium* were cultured for 2, 11, 19, 25, 25, and 25 days, respectively (late log phase to early stationary-phase growth). Prior to mass spectrometry experiments, all mycobacterial samples were adjusted to OD 1.0 ± 0.1 at 600 nm, either by dilution or concentration (∼10^8 cfu/mL).

Sample Preparation. MALDI analysis was performed on both (1) whole cells and (2) an acid/organic cell extract. Ten-milliliter aliquots of 1.0 OD cell suspension (cell suspension A) were pelleted by centrifugation (Sorvall Legend RT, Kendro Laboratory Products, Newtown, CT) at 4000*g* for 30 min. The cell pellets were washed with 10 mL of DDI and resuspended in 100 μL of DDI (MALDI of whole cells; cell suspension B) or 100 μL of 50/50 CH_3CN/4% trifluoroacetic acid (TFA) (MALDI of organic/acid

extract). For MALDI analysis of the organic/acid extract, the cells were repelleted and the supernatant fluid was taken for analysis.

For sample preparation optimization experiments, the following variables were examined: cell or supernatant fluid concentration, duration of cell extraction, stability of cell extract supernatant fluid, matrix concentration, and matrix composition. Individual 10-μL samples of *M. fortuitum* and *M. kansasii* were prepared by adding stock matrix solution to either cell suspension B or extract supernatant fluid in ratios ranging from 250:1 to 0.2:1. Diluent was chosen such that the final solution composition was 50/50 organic/2% TFA. Two separate 1-μL aliquots of each sample were deposited on a gold sample chip (Ciphergen Biosystems, Inc., Fremont, CA) and allowed to air-dry. Between analyses, the gold chips were washed as per manufacturer's instructions.

To characterize the variability associated with the methodology, an intraculture design was implemented in which eight samples were prepared from a single culture from each of the species listed above. Using this design, all variability across spectra within a species can be attributed to sample preparation and the process of data acquisition via MALDI-TOF MS. An interculture design in which two samples were prepared from each of eight independent cultures grown from each species was performed in order to further characterize the variability across detected peaks, between independent cultures and again between replicate samples from the same culture. This experimental setup was also utilized to begin proteomic profiling across the different species of mycobacteria.

Mass Spectrometry. Samples were analyzed by MALDI using a linear TOF mass spectrometer (PBS IIc, Ciphergen Biosystems, Inc.).[38,39] All spectra were acquired using a laser power at or slightly above the threshold for ion production (laser step 140−160). A total of 130 individual spectra (single laser shots) were averaged into a composite spectrum for each sample spot. The individual spectra were acquired in sets of 10 from 13 different positions within the sample spot. The 13 positions were kept constant for all samples and spots. The delayed extraction parameters were set to optimally focus a broad range (1−20 kDa) of mass-to-charge (m/z) ratios, and a mass deflector was used to reject ion signals below m/z 500. Data were acquired between 0 and 25 kDa for all bacteria, and all spectra were externally calibrated using a peptide/protein mixture consisting of five components that cover the range of 1−20 kDa (see Reagents, above).

Data Analysis. Initial data analysis was performed using Biomarker Wizard software (Ciphergen Biosystems, Inc.). Individual spectra were calibrated, baseline corrected, and normalized to total ion current (TIC), followed by identification of peak "clusters" representing peaks common to groups of spectra. This is an iterative process in which a first pass identifies all peaks that had a signal-to-noise ratio (S/N) > 5 and occurred in a minimum of 10% of the spectra. A second pass estimated peak intensity for all spectra that contained peaks that had S/N > 2 and were within 0.3% of the first pass peak mass. Finally, an estimated peak intensity was generated at each mass for all spectra that did not fulfill the criteria of the first two iterations. This process was applied to (1) each set of 8 spectra obtained from intraculture design, (2) each set of spectra from a given species from the interculture design, and (3) all 96 spectra representing all species together for proteomic profiling from the interculture design.

All data were analyzed using SAS/STAT software, Version 8.2 of the SAS system for Windows (SAS Institute, Cary, NC). The TIC intensity values were distributed log-normal, and thus, a natural log transformation was utilized on the intensity values. To assess intracultural variability, the coefficient of variation (CV) was calculated for each detected peak within a given species, and a mean CV was calculated. In addition, the percentage of peaks that were detected in first pass, second pass, or estimated in the third pass was calculated. To examine the sources of variability within species, a variance components analysis was performed for each species using the "PROC MIXED" procedure, assigning peaks, cultures, and samples as random factors. This analysis partitions all of the variation in the data into these three factors.

A series of multivariate discriminant analyses were performed in order to assess the potential utility of MALDI-TOF MS for discriminating among the different species of mycobacteria in the interculture experiment. For each peak detected, the mean value from the two samples from each culture were calculated and utilized as input variables. The "PROC DISCRIM" procedure was first applied by utilizing peaks detected with the Biomarker Wizard. The probability of misclassifying a given culture into the incorrect species was assessed using both the cross-validation and cross-list options in SAS. A stepwise variable selection method using the "PROC STEPDISC" procedure was then performed to select a subset of variables that could serve as predictor variables for class membership. This procedure was followed by the "PROC CANDISC" procedure to perform a canonical discriminant analysis on the subset of variables. This procedure creates new variables by taking linear combinations of the original variables and aids in determining the true underlying dimension of the data space. The canonical functions generated allow the calculation of canonical scores, which can be used to discriminate among the various species. Finally, a cluster analysis was performed using the "PROC CLUSTER" procedure using the complete linkage option, and a dendrogram was generated using the "PROC TREE" procedure. This procedure was performed with only the variables selected previously using the stepwise variable selection procedure.

Protein Database Queries. The UniProt database (http://www.uniprot.org)[40] was queried using the Text Search (http://www.pir.uniprot.org/search/textSearch.shtml) tool. The database search was constrained to search for proteins within 0.5% of the measured molecular weight that occurred within any mycobacterial species.

RESULTS AND DISCUSSION

Sample Preparation. Other researchers have demonstrated that the MALDI spectra obtained from bacteria can be highly variable.[26−32] Therefore, it is important to investigate sample preparation such that spectra are acquired under optimal conditions. Although a number of studies utilizing MALDI analysis of bacteria have been presented in the literature, none are optimized

(38) Hutchens, T. W.; Yip, T.-T. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 576−580.

(39) Merchant, M.; Weinberger, S. R. *Electrophoresis* **2000**, *21*, 1164−1177.

(40) Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. L. *Nucleic Acids Res.* **2004**, *32*, D115−D119.

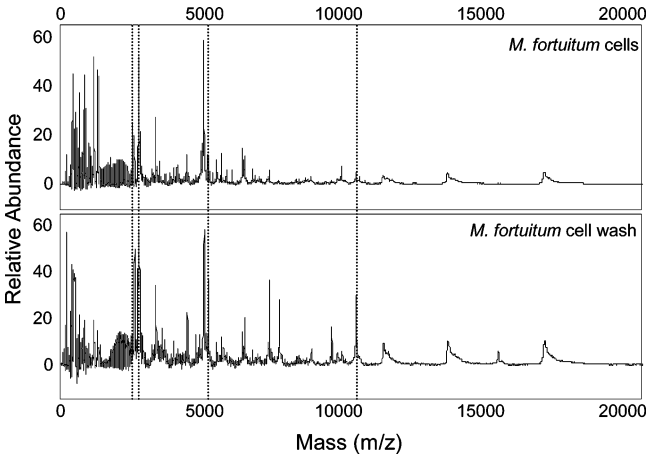

Figure 1. MALDI-TOF mass spectra of *M. fortuitum* using whole cells (top) and cell extract supernatant (bottom). Vertical dashed lines indicate unique biomarkers for *M. fortuitum* at *m*/*z* 2549, 2706, 5088, and 10160.

for mycobacteria. Therefore, we have examined the cell suspension B/supernatant fluid concentration, matrix concentration, and matrix composition for MALDI analysis of mycobacteria. In addition, we investigated the use of an acid/organic extract of the mycobacteria versus codeposition of intact cells with the MALDI matrix. In each case, spectra were examined for number and relative abundance of peaks detected.

Although the majority of the protocols in the literature describe the deposition of intact bacteria with the matrix,[19−23,26−28] other protocols describe the analysis of organic solvent cellular extracts.[25,29,30] No consensus exists on which sample preparation methodology is preferable. Because of the potential pathogenicity of many mycobacteria, biological safety is an important consideration for the analysis protocol. The use of cellular extracts instead of intact cells is an attractive option, as it minimizes operator contact with the infectious agent.

Although most bacteria can be inactivated by incubation in a disinfectant such as ethanol/water, or alternatively, in the matrix solution itself, mycobacteria, because of their complex, waxy cell wall are notoriously resistant to such conventional methods of inactivation. To determine whether the matrix solution inactivated mycobacteria, *M. fortuitum* samples were vortexed in a matrix solution identical to that used for the MALDI experiment. The mycobacteria were pelleted, removed from the matrix solution, and then incubated on agar plates for two weeks. Colony formation was observed, indicating the matrix solution does not inactivate the mycobacteria. To test whether incubation in the acetonitrile (ACN)/TFA solvent inactivated mycobacteria, an identical experiment was performed with *M. fortuitum* cells vortexed in the ACN/TFA solvent. No mycobacterial growth was observed on these plates; therefore, we conclude that incubation in the ACN/TFA solution inactivates the mycobacteria.

We evaluated the data produced from the analysis of both whole cells and cellular extracts in order to determine whether the use of cellular extracts is feasible for mycobacterial analysis by MALDI-TOF MS. Figure 1 compares the MALDI-TOF spectra of *M. fortuitum* obtained by direct deposition of mycobacteria with the matrix (top) with the ACN/TFA extract (bottom). The two spectra are very similar in terms of observed peaks and relative abundances; however some differences are noted. The cell extract

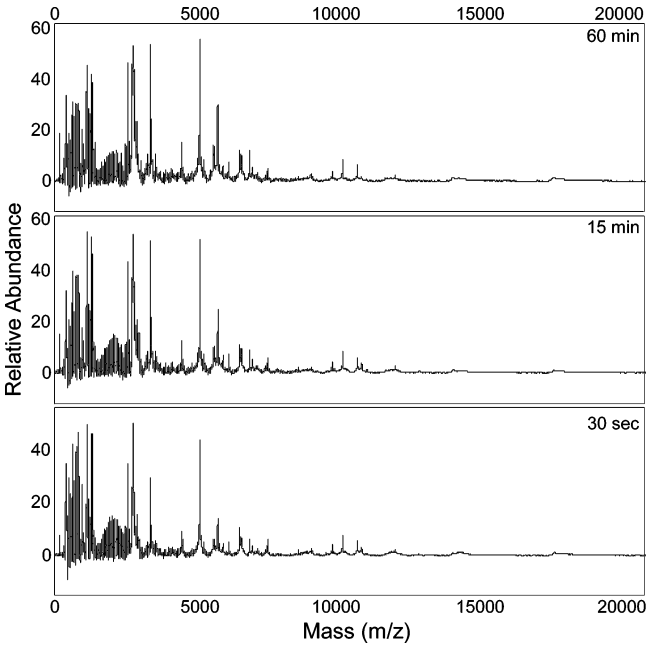

Figure 2. MALDI-TOF mass spectra of *M. fortuitum* cell extracts at 30 s (bottom), 15 min (middle), and 60 min (top)

spectrum has a higher relative abundance for protein signals of $m/z > 5000$, whereas the whole-cell spectrum has a higher relative abundance for some smaller molecules of $m/z < 2000$. Of particular importance, four signals that we have identified as being specific to *M. fortuitum* (indicated by vertical dashed lines in Figure 1) were all observed in much higher relative abundance in the cell extract spectra than in the whole-cell deposition spectra, e.g. *m*/*z* 2549 (50 vs 20% relative abundance), 2706 (60 vs 50%), 5088 (40 vs 15%), 10160 (30 vs 5%).

The high degree of similarity observed between the two spectra is consistent with the hypothesis that the spectra obtained from MALDI of whole cells are not the result of cell wall ablation; rather, they are the result of cellular surface molecules being solubilized into the matrix.[25] However, the higher relative abundance of some smaller molecules in the whole-cell spectrum is also consistent with the possibility of cell wall ablation by the MALDI desorption laser. Because the same information is available from either whole-cell or cell extract fluid analysis and because the cell extracts produce better S/N at higher *m*/*z* ratios and contain no viable mycobacteria, all subsequent spectra used for proteomic profiling in this study were taken from ACN/TFA cellular extracts.

A time-course experiment was performed to determine the optimum duration of extraction of the mycobacteria in the ACN/TFA solvent. Figure 2 presents spectra obtained after 30-s, 15-min, and 60-min cell extractions. No significant difference is noted between these three spectra; therefore, all spectra acquired in the proteomic profiling experiments utilize a 30-s extraction period. The extracts have good long-term stability (weeks) when stored at or below −20 °C, facilitating their use in diagnostic applications where transport of the samples may be required.

To determine the optimal matrix/analyte ratio, a number of spectra (*M. fortuitum*) were acquired by mixing stock matrix solution with cell suspension B or organic/acid supernatant fluid in ratios ranging from 250:1 to 0.2:1. Previous studies have shown that matrix/analyte ratios too high or low can produce poor

Table 1. Summary of Intraculture Variability

species	no. of samples	no. of peaks	mean CV	% peaks detected by first pass	% peaks detected by second pass	% peaks detected by est
M. fortuitum	8	41	0.134	56.7	40.6	2.7
M. kansasii	8	44	0.239	68.4	27.3	4.3
M. intracellulare	8	26	0.266	51.1	36.1	12.5
M. tuberculosis	8	40	0.229	44.1	40.6	15.3
M. bovis BCG	8	25	0.172	77.0	21.9	1.1
M. avium	8	20	0.318	50.6	31.3	18.1
average	8	32.7	0.226	58.0	33.0	9.0

Table 2. Summary of Total Variability, Intraculture, and Interculture

species	no. of cultures	% of total variance attributable to peaks	% of total variance attributable to interculture	% of total variance attributable to intraculture
M. fortuitum	8	84.75	5.78	9.47
M. kansasii	8	92.32	4.28	3.40
M. intracellulare	8	82.66	13.66	3.68
M. tuberculosis	8	85.31	12.21	2.48
M. bovis BCG	8	81.96	12.94	5.10
M. avium	8	65.33	18.18	16.48
average	8	82.06	11.18	6.77

spectra,[28] and our results are consistent with this observation. Spectral quality was degraded (loss of peaks, poor S/N) at the highest and lowest matrix/analyte ratios (data not shown). The best results were consistently obtained for samples containing 5 μL of matrix and 1 μL of cell suspension B or 2 μL of ACN/TFA supernatant fluid. We examined the spectra produced from four different matrix compositions: (1) CHCA, (2) CHCA/fructose, (3) SPA, and (4) SPA/fructose. The best results were obtained for CHCA (consistent with the findings of Haag and co-workers)[25] and CHCA/fructose. The use of fructose as a matrix additive has been well documented,[41−46] and benefits include improved shot-to-shot reproducibility, sample durability, and resolution (due to reduced metastable fragmentation).

Based on the results of these experiments, a sample preparation methodology was finalized for use in mycobacterial profiling experiments. Mycobacteria from 10 mL of OD 1.0 suspension are pelleted, washed with DDI, vortexed in 50/50 acetonitrile/4% TFA for 30 s, and then repelleted, and the supernatant fluid is drawn off. A 2-μL aliquot of this supernatant fluid is added to 5 μL of 1:1 stock CHCA/fructose and diluted to 10 μL such that the final solvent composition is 50/50 organic/2% TFA.

Both an intraculture and an interculture comparison were performed in order to characterize the variability associated with the methodology. The results of the intraculture experiment are presented in Table 1. Between 20 and 44 peak clusters were identified for each species, and the average coefficient of variability was 0.226. Table 1 also includes the percentage of peaks identified on the first pass, second pass, and estimated following implementation of the peak detection algorithm (Biomarker Wizard). In general, the more variable the spectrum-to-spectrum data, the higher the percentage of estimated peaks. The variability between spectra is very reasonable for MALDI-TOF MS of dried droplet samples, which tend to be very inhomogeneous.[43]

A summary of the variance components analysis for the interculture experiment is presented in Table 2. On average, ~7% of the variability in the data from a given species is due to sample-to-sample variation. This indicates that the mycobacterial growth conditions and MALDI-TOF MS analysis are highly reproducible for a given sample. The culture-to-culture variability is also low relative to the total variability. This component represents variation across independently grown cultures and is indicative of relatively homogeneous protein expression patterns across cultures. The data in Table 2 indicate that, on average, 83% of the variability observed in the intraculture experiment is due to variability in the relative abundance of observed m/z ratios; e.g., of the total variability, very little is attributable to differences between cultures. Figure 3 presents MALDI-TOF mass spectra from each of the eight independently cultured *M. fortuitum* samples and illustrates the high degree of reproducibility obtained between cultures. The vertical dashed lines represent the four signals identified as being unique to *M. fortuitum*.

Proteomic Profiling of Mycobacteria. Eight composite spectra were obtained by averaging two individual spectra for each of the six mycobacterium species in the interculture experiment, and representative spectra for each species are presented in Figure 4. A total of 77 peaks were identified with the Biomarker Wizard from the 48 independently grown mycobacterial cultures (8/species). Although there are similarities between spectra, particularly at lower m/z values, signals at m/z values greater than 5000 tend to be unique. Of the species investigated here, only two species, *M. avium* and *M. intracellulare* had few signals at higher m/z ratios. The similarity in the spectra of *M. avium* and *M. intracellulare* is not unexpected, as they are closely related biologically and are often referred to as the *M. avium intracellulare* complex. Even though these species are closely related, they can be unambiguously differentiated on the basis of m/z 4833 (*M. intracellulare*) and 7630 (*M. avium*). Similarly, positive identification of *M. tuberculosis* can be made versus any of the other five species on the basis of six unique m/z values observed in each spectrum (m/z 4617, 4658, 4747, 4788, 5042, and 5003). In addition, m/z 9086 occurs only in *M. tuberculosis* and *M. bovis* BCG (although at lower relative abundance in *M. bovis* BCG than in *M. tuberculosis*). No unique biomarkers were identified for *M. bovis* BCG, consistent with the fact that no genes unique from *M. tuberculosis* occur in *M. bovis*. The genome of *M. bovis* BCG differs from that of M. *tuberculosis* largely through deletion, and it is suggested that the greatest degree of sequence variation between *M. tuberculosis* and *M. bovis* BCG occurs in the genes coding for cell wall and secreted proteins.[47] This difference is likely to play a predominant role in the difference in pathogenicity of *M. tuberculosis* and *M. bovis* BCG, as many cell surface components are well known to be virulence and pathogenicity factors.[48] Despite the >99.95% similarity between the two genomes, the observation

(41) Beavis, R. C.; Lindner, J.; Grotemeyer, J.; Schlag, E. W. *Chem. Phys. Lett.* **1988**, *146*, 310−314.
(42) Castoro, J. A.; Wilkins, C. L. *Anal. Chem.* **1993**, *65*, 2621−2627.
(43) Hettick, J. M.; McCurdy, D. L.; Barbacci, D. C.; Russell, D. H. *Anal. Chem.* **2001**, *73*, 5378−5386.
(44) Hettick, J. M. Ph.D. Dissertation, Texas A&M University, College Station, TX, 2003.
(45) Koomen, J. M.; Russell, W. K.; Hettick, J. M.; Russell, D. H. *Anal. Chem.* **2000**, *72*, 3860−3866.
(46) Koster, C.; Castoro, J. A.; Wilkins, C. L. *J. Am. Chem. Soc.* **1992**, *114*, 7572−7574.

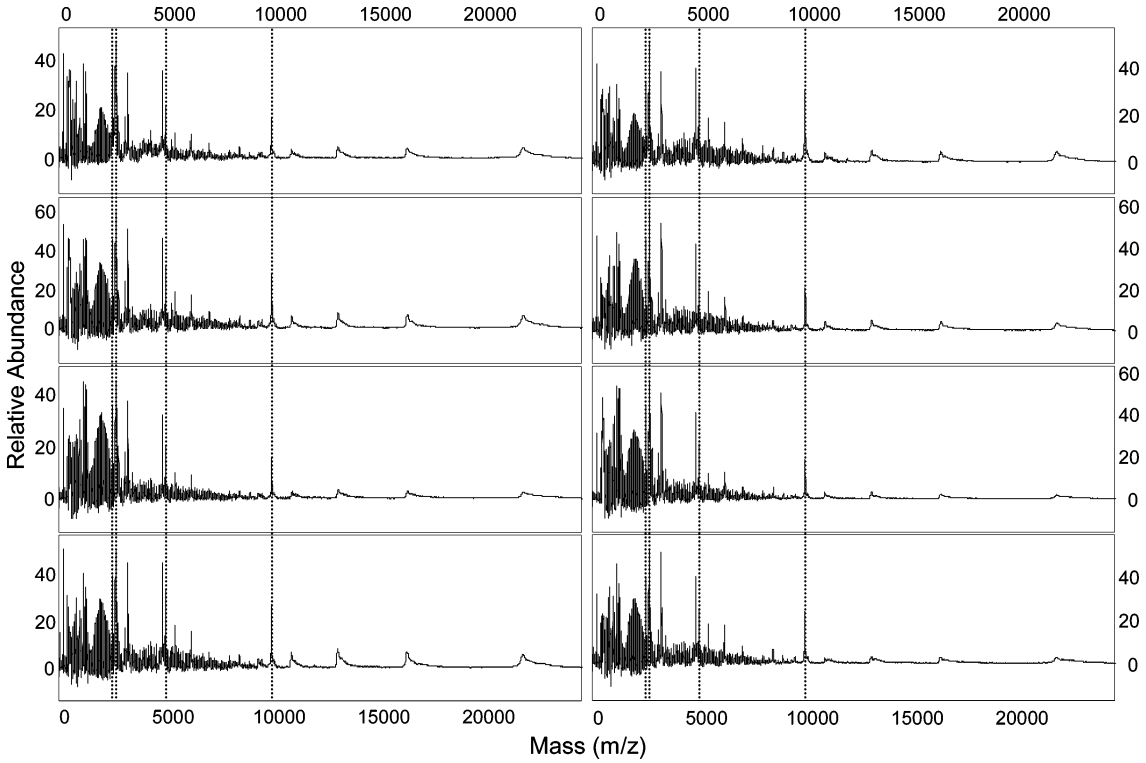

Figure 3. MALDI-TOF mass spectra of eight cultures of *M. fortuitum* grown independently. Vertical dashed lines indicate unique biomarkers for *M. fortuitum* at *m*/*z* 2549, 2706, 5088, and 10160.

Table 3. Summary of Potential Biomarkers Observed by MALDI-TOF MS from Mycobacteria Extracts

species	*m*/*z*
Potential Species-Specific Biomarkers	
M. fortuitum	2549, 2706, 5088, 10160
M. kansasii	2561, 9348, 12333
M. intracellulare	1104, 1233, 1589
M. tuberculosis	4617, 4658, 4747, 4788, 5002, 5043
M. bovis BCG	no *m*/*z* values unique to *M. bovis*.
M. avium	1217, 6359
Potential Genus-Specific Biomarkers	
all	1331, 1370, 1393, 1421, 1436, 1615, 1792

of *m*/*z* 9086 in the absence of *m*/*z* 4617, 4658, 4747, 4788, 5002, and 5043 is unique, allowing unambiguous identification of *M. bovis* BCG by MALDI-TOF MS profiling. The observed biomarkers, potentially species- and genus-specific, are summarized in Table 3. It should be stressed that these biomarkers are only potentially specific, as we have not examined all species of the genus *Mycobacterium*, nor have we examined species outside the genus *Mycobacterium*.

The utility of using protein expression profiles to discriminate among several species of mycobacteria is illustrated through the use of common statistical techniques such as discriminant and cluster analysis. A linear discriminant analysis using all 77 identified peaks in the data set resulted in 100% correct classification of individual cultures into the appropriate species group. Using a subset of 20 variables selected using the stepwise variable selection procedure also resulted in 100% correct classification. Using these same 20 variables in the canonical discriminant analysis indicated that the species mean vectors (the vectors representing the 20 variables included in the model) could be placed in five-dimensional space. However, the first three canonical functions were able to account for over 85% of the total variability between species. A three-dimensional plot of the canonical scores is shown in Figure 5, and it is clear that the different species lie in a distinct spatial arrangement. Each individual spectrum (48 total) is represented by a vertical bar in 3D space. In every case, the eight spectra from within a particular species cluster together, whereas a species-specific cluster is spatially resolved from every other species cluster.

A second method to view the potentially unique profile of protein expression within a given species is to utilize cluster analysis. The dendrogram representing the "distance" that each culture lies from one another in the data space is presented in Figure 6. Similar to Figure 5, all 48 cultures (8 independent cultures from each of 6 species) are displayed in the data space. From this figure, it can be seen that each species clusters together such that all individual cultures within a species are closer to one another than they are to cultures belonging to another species. In addition, biologically similar species cluster together. For example, the *M. avium intracellulare* complex clusters together, and *M. bovis* BCG and *M. tuberculosis* cluster together. This indicates that the spectra for each of the two pairs are more similar to one another than spectra from other species. Likewise, the *M. bovis M. tuberculosis* cluster and the *M. fortuitum M. kansasii* cluster are closer to one another than either to the *M. avium M. intracellulare* complex.

(47) Garnier, T.; Eiglmeier, K.; Camus, J.-C.; Medina, N.; Mansoor, H.; Pryor, M.; Duthoy, S.; Grondin, S.; Lacroix, C.; Monsempe, C.; Simon, S.; Harris, B.; Atkin, R.; Doggett, J.; Mayes, R.; Keating, L.; Wheeler, P. R.; Parkhill, J.; Barrell, B. G.; Cole, S. T.; Gordon, S. V.; Hewinson, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7877−7882.

(48) Evason, D. J.; Claydon, M. A.; Gordon, D. B. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 49−54.

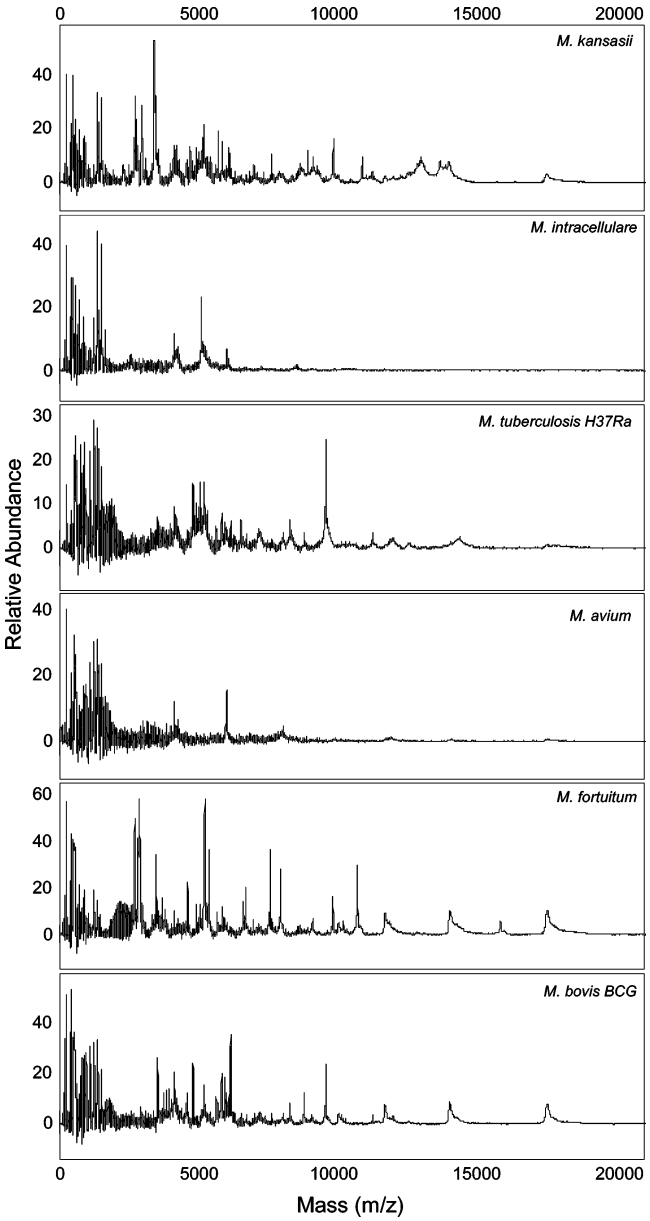

Figure 4. MALDI-TOF mass spectra of cell extract supernatants of six different mycobacterial species.

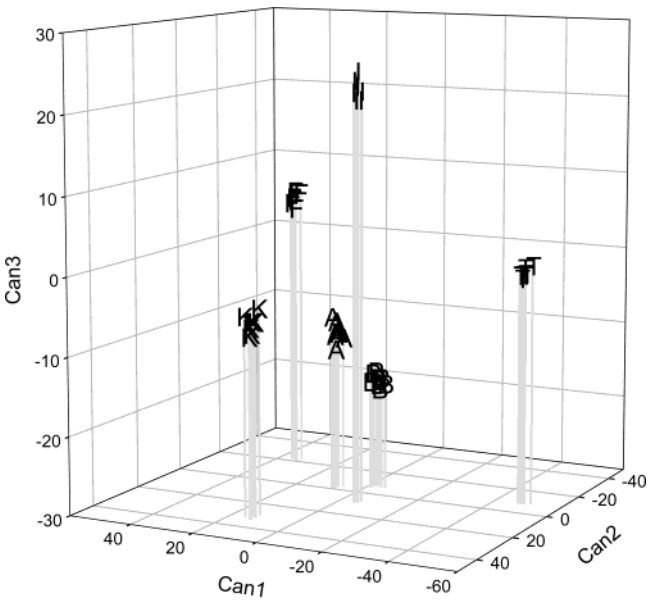

Figure 5. 3D plot of the first three canonical scores (85% of the total variance) for eight independent cultures each of (I) *M. intracellulare*, (F) *M. fortuitum*, (T) *M. tuberculosis* H37Ra, (B) *M. bovis* BCG, (A) *M. avium*, and (K) *M. kansasii*. Each individual culture is represented by a vertical bar in 3D space.

A logical extension of the current experiments is the identification of the biomarkers suggested by statistical analysis. To suggest potential protein candidates for each *M. tuberculosis* biomarker, we searched the UniProt database for proteins that fell within 0.5% of the observed biomarker *m*/*z*. UniProt suggested several possible protein candidates near 9086 Da, a biomarker we found only in *M. bovis* BCG and *M. tuberculosis* H37Ra. The best match for the biomarker at *m*/*z* 9086 is very hypothetical protein Rv2076c/MT2136/Mb2101c (Primary Accession Number Q10684), which has a molecular weight of 9077.[47,49,50] Protein Rv2076c/MT2136/Mb2101c is coded by the genomes of both *M. tuberculosis* and *M. bovis* BCG and is a transmembrane protein. Similarly, protein candidates were identified for *m*/*z* 4617 (hypothetical protein MT0009/ hypothetical protein MT2514), 4658 (hypothetical protein MT0492/ hypothetical protein MT1025.1), 4747 (hypothetical protein MT3744), 5002 (hypothetical protein MT3032), and 5043 (hypothetical protein MT2960/ hypothetical protein MT3135).

Although a molecular mass search serves as a useful method to select potential protein candidates, the genome does not predict posttranslational modification. Thus, the positive identification of proteins on the basis of molecular weight alone is not possible. The confidence level for protein identification is dramatically higher when sequence information is used to query the database. For this reason, we are currently pursuing the identification of these biomarkers via peptide mass mapping and peptide sequencing by tandem mass spectrometry.

Because our analysis utilizes cellular extracts, proteins localized on the surface of the cell are attractive candidates. Although other researchers have identified highly abundant ribosomal proteins in MALDI-TOF MS spectra of whole bacterial cells,[51] this is a less likely source for the biomolecules observed in these experiments. The bacterial cells utilized in the study of Pineda[51] and co-workers underwent a −80 °C freeze/thaw cycle, and it is likely that a significant portion of the bacterial cells were lysed, releasing cytosolic proteins. Mycobacteria are difficult to lyse, and traditional methods such as freeze/thaw cycles or sonication are ineffective; rather high-pressure methods such as the French press are often utilized. Because our sample preparation methodology is unlikely to result in significant cell lysis, it is unlikely that cytosolic proteins are sampled. This hypothesis is consistent with the observations of other researchers who have determined that exposure of intact cells to the matrix solution does not extract components from

(49) Fleischmann, R. D.; Alland, D.; Eisen, J. A.; Carpenter, L.; White, O.; Peterson, J.; DeBoy, R.; Dodson, R.; Gwinn, M.; Haft, D.; Hickey, E.; Kolonay, J. F.; Nelson, W. C.; Umayam, L. A.; Ermolaeva, M.; Salzberg, S. L.; Delcher, A.; Utterback, T.; Weidman, J.; Khouri, H.; Gill, J.; Mikula, A.; Bishai, W.; Jacobs, W. R., Jr.; Venter, J. C.; Fraser, C. M. *J. Bacteriol.* **2002**, *184*, 5479−5490.

(50) Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; et al. *Nature* **1998**, *393*, 537−544.

(51) Pineda, F. J.; Antoine, M. D.; Demirev, P. A.; Feldman, A. B.; Jackman, J.; Longenecker, M.; Lin, J. S. *Anal. Chem.* **2003**, *75*, 3817−3822.

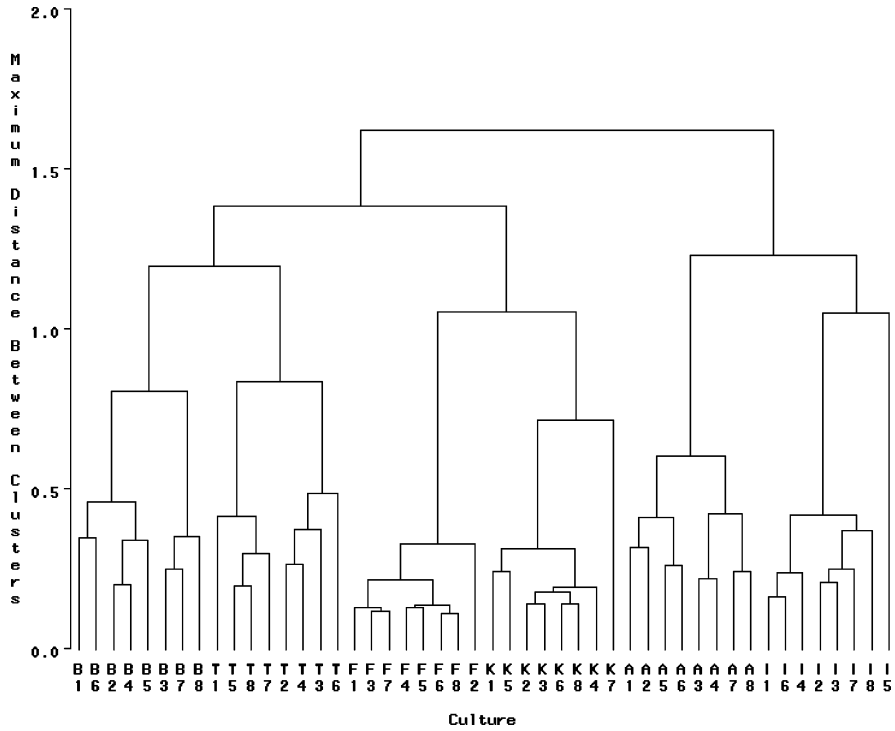

Figure 6. Cluster analysis dendrogram for eight independent cultures (1−8) of each (I) *M. intracellulare*, (F) *M. fortuitum*, (T) *M. tuberculosis* H37Ra, (B) *M. bovis* BCG, (A) *M. avium*, and (K) *M. kansasii*.

within the cell.[48] Therefore, it is likely that the biomarker signals observed in these experiments arise from noncovalently bound surface molecules that are extracted into the solvent, rather than from internal cell components.

In addition to the biomarkers that are observed to be unique to a particular species, we have identified a number of potentially genus-specific biomarkers that occur in all mycobacterial species from this study (m/z 1331, 1370, 1393, 1421, 1436, 1615, and 1792). These signals are too small to be proteins and are more likely to be cell wall components such as mycolic acids or lipids.

Interestingly, a series of peaks at m/z 66, 33, and 22 kDa were observed in MALDI TOF mass spectra, corresponding to the singly, doubly, and triply charged ions of bovine serum albumin (BSA), respectively. BSA is a major component of the growth medium. These peaks are present in the MALDI-TOF spectrum of the growth media and are present at very low levels in the DDI wash. However, the peaks occur with greater relative abundance in the bacterial proteomic profiles than in the wash, leading us to conclude that the BSA is noncovalently bound to the surface of the bacterium prior to undergoing catabolism and is released into the organic/acid extract supernatant fluid.

CONCLUSION

We have successfully applied MALDI-TOF mass spectrometry-based proteomic profiling to the analysis of various mycobacteria, including *M. tuberculosis* H37Ra. Each of the six species examined here can be unambiguously identified on the basis of unique m/z values in the protein profile, and the spectra can be transformed and visualized in 3D space by canonical discriminate and cluster analyses. Statistical analysis of MALDI-TOF data suggests unique biomarkers not only for individual species but potentially for the genus *Mycobacterium*. Furthermore, these biomarkers may have potential for use in species identification by other means, such as immunoassay.

Our data demonstrate that codeposition of the intact mycobacteria is not necessary for successful proteomic profiling; similar data are obtained from the analysis of an acetonitrile/trifluoroacetic acid cell extract. This allows mycobacterial identification to be performed with minimum contact with viable pathogenic mycobacteria. In addition, we find that these organic/acid extracts are stable when stored at or below −20 °C, thus facilitating the transport of nonpathogenic samples from a culture site to the mass spectrometry laboratory.

Sample preparations using CHCA/fructose as a matrix are optimal for profiling of mycobacteria because they yield strong signals across the mass range of interest (0−20 kDa), they are durable, and they produce satisfactory shot-to-shot reproducibility.

In aggregate, these studies demonstrate the feasibility of using MALDI-TOF mass spectrometry to identify cultured mycobacterial species in the laboratory. Further studies, including more exhaustive analysis of species from the genus *Mycobacterium* as well as clinically based studies, will be necessary to build a database of biomarkers before the methodology can be used prospectively in a clinical setting. Finally, protein identification experiments such as peptide sequencing by tandem mass spectrometry are underway to provide confident identification of the biomarkers suggested by these experiments.

ACKNOWLEDGMENT

The authors thank Professor Brian L. Wickes of the Department of Microbiology and Immunology, University of Texas Health Sciences Center at San Antonio, for helpful discussions.

Received for review April 19, 2004. Accepted August 2, 2004.

AC049410M