

MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates

Hercules Moura¹, Adrian R. Woolfitt¹, Maria G. Carvalho², Antonis Pavlopoulos¹, Lucia M. Teixeira³, Glen A. Satten¹ & John R. Barr¹

¹Division of Laboratory Sciences, National Center for Environmental Health and ²Division of Bacterial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, M.S. F-50, Atlanta, GA, USA; and ³Instituto de Microbiologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

OnlineOpen: This article is available free online at www.blackwell-synergy.com

Correspondence: John R. Barr, Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, M.S. F-50, 4770 Buford Highway NE, Atlanta, GA 30341, USA. Tel.: +1 770 488 7848; fax: +1 770 488 0509; e-mail: jbarr@cdc.gov

Received 8 August 2007; revised 12 March 2008; accepted 8 April 2008.
First published online 5 June 2008.

DOI:10.1111/j.1574-695X.2008.00428.x

Editor: Kai Man Kam

Keywords

matrix-assisted laser desorption/ionization time of flight mass spectrometry; biomarkers; protein fingerprints; *Streptococcus pyogenes*; ribosomal proteins; necrotizing fasciitis.

Abstract

A novel mass spectral fingerprinting and proteomics approach using MALDI-TOF MS was applied to detect and identify protein biomarkers of group A *Streptococcus* (GAS) strains. *Streptococcus pyogenes* ATCC 700294 genome strain was compared with eight GAS clinical isolates to explore the ability of MALDI-TOF MS to differentiate isolates. Reference strains of other bacterial species were also analyzed and compared with the GAS isolates. MALDI preparations were optimized by varying solvents, matrices, plating techniques, and mass ranges for *S. pyogenes* ATCC 700294. Spectral variability was tested. A subset of common, characteristic, and reproducible biomarkers in the range of 2000–14 000 Da were detected, and they appeared to be independent of the culture media. Statistical analysis confirmed method reproducibility. Random Forest analysis of all selected GAS isolates revealed differences among most of them, and summed spectra were used for hierarchical cluster analysis. Specific biomarkers were found for each strain, and invasive GAS isolates could be differentiated. GAS isolates from cases of necrotizing fasciitis were clustered together and were distinct from isolates associated with noninvasive infections, despite their sharing the same *emm* type. Almost 30% of the biomarkers detected were tentatively identified as ribosomal proteins.

Introduction

Streptococcus pyogenes, frequently referred to as group A *Streptococcus* (GAS), is one of the most common and versatile human bacterial pathogens. It causes a variety of diseases, ranging from mild and quite frequent noninvasive infections of the upper respiratory tract and skin to severe invasive infections that include necrotizing fasciitis and streptococcal toxic shock syndrome (Facklam, 2002). This bacterial species is also associated with such life-threatening poststreptococcal sequelae as acute rheumatic fever and glomerulonephritis. For many years, the evaluation of epidemiologic relationships between GAS isolates was based on serological typing for detection of a bacterial cell surface protein, the M protein, which is considered as a major

virulence factor of these microorganisms. More recently, several DNA-based typing methods have been applied to evaluate the diversity of GAS isolates and to elucidate their association with different diseases (Facklam *et al.*, 1999; McGregor *et al.*, 2004; Doktor *et al.*, 2005). A major improvement to serological M typing was *emm* gene sequencing analysis (Beall *et al.*, 1997; Teixeira *et al.*, 2001). However, new typing methods are still needed to improve strain differentiation and to contribute with new insights into the epidemiology and pathogenesis of GAS infections. A better understanding of GAS infections would lead to improvements in the rapid and precise detection of the microorganism and the development of more effective strategies for treatment and prevention of GAS diseases (Currie, 2006).

A practical and rapidly evolving application of mass spectrometry (MS)-based methods in microbiology is the rapid identification of microorganisms and strain differentiation (van Baar, 2000; Fenselau & Demirev, 2001; Lay, 2001). Two soft ionization methods, matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) and liquid chromatography-electrospray ionization mass spectrometry (LC-ESI MS), are particularly useful in the identification and characterization of proteins in complex mixtures. They play a central role in proteomics research (Aebersold & Mann, 2003; Patterson & Aebersold, 2003; Gingras *et al.*, 2005; Lambert *et al.*, 2005). MS methods offer detection limits superior to traditional methods for the analysis of many relevant microorganisms, in addition to offering other significant advantages. Because of their simplicity, speed, and accuracy, MS methods have been successfully applied for biomarker discovery and for characterizing various bacterial agents (Ruelle *et al.*, 2004; Shaw *et al.*, 2004; Stackebrandt *et al.*, 2005; Valentine *et al.*, 2005; Wunschel *et al.*, 2005a, b; Pignone *et al.*, 2006; Liu *et al.*, 2007). It is now well accepted that the mass spectra obtained by MALDI-TOF MS provide characteristic patterns of proteins (fingerprints composed of unique biomarkers) from whole organisms that can be used to identify bacteria, viruses, protozoa, and fungi (Magnuson *et al.*, 2000; Amiri-Eliasi & Fenselau, 2001; Moura *et al.*, 2003; Krader & Emerson, 2004; Wunschel *et al.*, 2005a). However, interpretation of MS data from complex mixtures is difficult, and early reports using visual analysis of mass spectra confirmed the difficulties in the differentiation of microorganisms (Fenselau & Demirev, 2001; Lay, 2001). Currently, there are two major approaches to identify microorganisms by MS: a fingerprint-statistical-based approach (Holland *et al.*, 1996; Jarman *et al.*, 1999, 2000; Dickinson *et al.*, 2004; Satten *et al.*, 2004; Valentine *et al.*, 2005; Wunschel *et al.*, 2005a, b) and a proteomics approach (Demirev *et al.*, 1999, 2001; Pineda *et al.*, 2000, 2003; Pribil & Fenselau, 2005; Pribil *et al.*, 2005).

Here we report the development and application of MALDI-TOF MS analysis with statistical analysis as a potential complementary method for GAS characterization and strain differentiation. We have applied the MALDI-TOF MS with RANDOM FOREST (RF) analysis, hierarchical cluster analysis, and proteomic database searching to a limited number of invasive GAS and noninvasive GAS isolates. Using a combination of novel data processing methods for visual peak comparison, statistical analysis, and proteomics database searching, we were able to demonstrate the power of this combined approach on a small number of well-characterized and diverse GAS isolates. We believe that the combined approach will strengthen the ability of MALDI-TOF MS to differentiate microorganisms.

Materials and methods

Chemicals

All chemicals used during this study were purchased from Sigma-Aldrich (St. Louis, MO) except where indicated. Buffers and culture media were obtained from the Scientific Resources Program at the Centers for Disease Control and Prevention (CDC).

Bacterial strains

A total of nine GAS strains, comprising the *S. pyogenes* genome strain (ATCC 700294) and eight clinical isolates, were included in this study (Table 1). The panel of selected GAS clinical isolates was obtained from patients with different syndromes who lived in the United States and Brazil. All clinical isolates were collected for previous studies and were obtained using the ethical guidelines of either the CDC or the Instituto de Microbiologia, Universidade Federal do Rio de Janeiro. These isolates were identified previously using conventional phenotypic tests (Facklam, 2002) and were characterized by sequencing of *emm* gene-specific PCR products (*emm*-typing) (Beall *et al.*, 1997; Teixeira *et al.*, 2001). In addition, a reference strain of *Streptococcus agalactiae* (ATCC 13813), another β hemolytic streptococcal species that plays a role as an important agent of infections in humans, was included for comparative purposes. Reference strains of *Enterococcus faecalis* (ATCC 29212), *Escherichia coli* (ATCC 25922), and *Staphylococcus aureus* (ATCC 29213) were included as outgroups for statistical purposes.

Growth conditions to obtain bacterial cells for MS analysis

Bacterial strains were initially grown on blood agar (trypticase soy agar containing 5% defibrinated sheep blood) plates at 37 °C for 18–24 h. All microbiological procedures were carried out in a certified Biosafety Level 2 cabinet equipped with high-efficiency particulate air filters. Bacterial cells

Table 1. Characteristics of *Streptococcus pyogenes* isolates included in this study

Strain	Source	Origin	<i>emm</i> type
Th 892-97	Throat infection	Brazil	<i>emm1</i>
Th 834-97	Throat infection	Brazil	<i>emm1</i>
Th 1476-97	Throat infection	Brazil	<i>emm1</i>
Sk 3961-98	Skin infection	Brazil	<i>emm12</i>
Th 3159-00	Throat infection	Brazil	<i>emm12</i>
NF 307-03	Necrotizing Fasciitis	USA	<i>emm1</i>
NF 312-03	Necrotizing fasciitis	USA	<i>emm1</i>
NF 325-03	Necrotizing fasciitis	USA	<i>emm12</i>
ATCC 700294	Genome strain	USA	<i>emm1</i>

were inoculated in tubes containing 30 mL of Todd Hewitt broth (THB). After incubation at 37 °C for 18–24 h, cells were washed three times with 10 mL of Tris-sucrose buffer (0.01 M Tris, 0.025 M sucrose, pH 7.0) by centrifugation under refrigeration (4–10 °C). Cells were then suspended in 500 µL of Milli-Q grade water (Millipore) and maintained at –80 °C until they were prepared for MS analysis. To test batch variability, cells from strain ATCC 700294 grown on different occasions over a 3-week (batch 2) and a 6-week (batch 3) interval were prepared and analyzed. The clusters of distinctive peaks or biomarkers desorbed from the first batch of *S. pyogenes* were used as reference spectra to evaluate method reproducibility, batch-to-batch reproducibility, and interference of growth media on biomarker expression and detection. A subset of cells from batch 2 was subjected to 2×10^8 rads of γ irradiation before processing for MALDI-TOF MS analysis. This has been shown to destroy bacterial viability without disrupting the protein structure (Shaw *et al.*, 2004). To test the interference of growth media on biomarker detection, *S. pyogenes* ATCC 700294 was also grown on blood agar plates, and isolated colonies (1, 2, 3, 4, and 5 colonies) were collected and analyzed directly without further treatment. In addition, 30 colonies were pooled, washed three times with Tris-sucrose buffer as described above, and analyzed. The other bacterial samples, including the eight GAS clinical isolates, and the reference strains used throughout the study were cultured three times over a 2-week period and washed as described above.

Sample preparation

Washed bacterial cells were prepared for MS analysis as described previously (Shaw *et al.*, 2004) with a few modifications. Briefly, whole bacteria were adjusted to concentrations of 10^6 and 10^7 organisms μL^{-1} using Milli-Q grade water, and then premixed with equal volumes of the respective matrix solutions just before spotting on the MALDI target. MALDI matrices consisted of 10 mg mL^{-1} saturated solutions of diaminobenzoic acid (DHB), α -cyano-4-hydroxycinnamic acid (CHCA), or 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid; SA). DHB was dissolved in Milli-Q grade water containing 0.1% trifluoroacetic acid (TFA), whereas CHCA and SA were mixed, respectively, with 50% and 70% acetonitrile (ACN) and Milli-Q grade water containing 0.1% TFA. A 192-well stainless-steel sample target plate with numbered circular wells [Applied Biosystems (AB), Framingham, MA] was utilized for MALDI-TOF analysis. The plates were washed with Milli-Q grade water, treated with methanol, and allowed to dry at room temperature. When dry, 0.5 µL of the premixed suspensions containing matrices and whole bacterial forms or mass standards for calibration (Sequazyme Peptide Mass Standards Kit, AB) were spotted in four separate proximal wells to create quadruplicates

of samples and controls. Bovine cytochrome *c* (1 mM) was added to one well of each sample as an internal standard. After drying, the plate was inserted into the instrument for MALDI-TOF MS analysis.

MALDI-TOF MS analysis

Mass spectra were acquired using a MALDI-TOF/TOF mass spectrometer (AB 4700 Proteomics Analyzer) equipped with a nitrogen laser (Nd:YAG) at 337 nm, and with a 200 Hz repetition rate. Analyses were performed on at least three different days in linear-delayed extraction positive ion mode, using an accelerating voltage of 20 kV. For optimum data quality, mass spectra in the 1000–6000 *m/z* range were acquired using either CHCA or DHB as the matrices; for the 2000–14 000 *m/z* range, SA was used as matrix. A minimum of 11 individual spectra representing 10 accumulated sub-spectra were obtained from each well. The acceptance criteria were based on 1000 laser shots per spot, and they required signal intensities between 2000 and 55 000 counts, with a signal-to-noise of 10 or greater.

Data processing

Mass spectra from at least four harvestings were processed as follows: profile spectral data were exported as text-format *m/z*-intensity lists with a unified *m/z* scale, using custom Microsoft Visual Basic for Applications (VBA) macros in DATA EXPLORER™, the AB viewing application. The text data were further processed and viewed using a suite of custom Microsoft VISUAL BASIC.NET (VB.NET) programs, the VBA macros, and VB.NET programs 'MULTISPEC PROCESSOR' and 'MULTISPEC VIEWER'. The latter was designed to display hundreds of spectra at once in a number of formats, including a simulated sodium dodecyl sulfate–polyacrylamide gel electrophoresis view, making it extremely valuable for visual analysis of the data sets, which may be comprised of several thousand individual spectra. These spectra were normalized to the base peak, smoothed using a 21-point, 2-pass Gaussian algorithm, and finally standardized and denoised using a custom FORTRAN program (Satten *et al.*, 2004). We used PAST software v1.34 (<http://folk.uio.no/ohammer/past/doc1.html>) for hierarchical cluster analysis, with the single summed spectra (one spectrum representing each organism) for input. We used RF v5.1 (http://www.stat.berkeley.edu/users/breiman/RandomForests/cc_home.htm) for classification and identification, in this case with nine summed spectra from three separated harvestings (one month apart) of each organism as a training set and three summed spectra as unknowns. The recompilation of the Fortran RF code for each experimental condition was automatically driven by VB.NET programs, and custom viewing applications were developed to aid in the interpretation of the RF results.

Tentative peak matching and database searching

Tentative identification of the prominent peaks was carried out using the Tag–Ident proteomics tool or ExPASy Sequence Retrieval System (<http://us.expasy.org>). In addition, 'MS DB FILTER', a custom VB.NET algorithm, was used to construct a CDC-modified database filtered from UNIPROT (<http://www.ebi.ac.uk/uniprot/index.html>). MS DB FILTER excludes any Swiss-Prot and TrEMBL or UNIPROT entry described as a fragment. It strips out signal and prepeptide sequences and applies a rule to add or remove initial methionine, as described by Pineda (Pineda et al., 2003). The CDC-modified filtered database was used for data mining the deduced proteome from several bacterial species used in this study that have the whole genome sequenced. Custom algorithms within MULTISPEC VIEWER were used to generate peak lists from the acquired mass spectra and to match these peaks with the CDC-modified database to provide tentative protein identifications. The extracted peak lists were also submitted to the 'Rapid Microorganism Identification Database' (<http://pinedalab.jhsph.edu/microOrgID>).

Results

***Streptococcus pyogenes* ATCC 700294 baseline spectra and effect of bacterial growth media on biomarker detection**

Three batches of *S. pyogenes* genome strain grown in liquid media were analyzed by MALDI-TOF MS. Spectra obtained on 10^6 – 10^8 whole bacteria μL^{-1} revealed complex spectral patterns with 20–40 peaks in the m/z range corresponding to 1000–6000 Da or 2000–14 000 Da. Observed mass spectra are believed to consist primarily of protonated peptide and protein signals, although some signals below m/z 4000 may represent other materials. The mass range corresponding to 2000–14 000 Da was preferred for this study because the peaks in this region were more abundant and more consistent. A representative example of *S. pyogenes* mass spectra is shown in Fig. 1a, while Fig. 1b depicts the reproducible patterns from denoised summed spectra revealed using MULTISPEC VIEWER. A γ -irradiated subset from batch 2 is included in Fig. 1b and shows few changes in the summed spectra as compared with the MALDI-TOF spectra obtained on whole live organisms. Visual analysis of processed spectra using MULTISPEC VIEWER revealed only a minor variation from day to day and from well to well in the quadruplicate analysis of the suspensions containing 10^6 or 10^7 bacterial cells. Figure 2 shows the spectra and a gel view representation of the mass spectra obtained from a single colony and from pools of 2, 3, 4, 5, and 30 colonies of *S. pyogenes* grown on blood agar and loaded directly onto the MALDI target. These spectra showed altered peaks, along with a set of

similar but noisier peaks when compared with those obtained from cells grown in liquid media. Although a higher level of background noise was present in cells grown on blood agar, a key set of peaks corresponding to m/z 4451, 5951, 6310, 7962, 8181, and 9030 and representing potential biomarkers were consistently seen in all wells, as is easily observed in the denoised spectra depicted in Fig. 2.

MULTISPEC VIEWER revealed differences among genus and species

Three batches of washed cells of *S. agalactiae* (ATCC 13813), *E. faecalis* (ATCC 29212), *S. aureus* (ATCC 29213), and *E. coli* (ATCC 25922) revealed typical spectral patterns in the m/z range of 2000–14 000 Da. Figure 3 depicts spectra of the three bacterial standards as compared with *S. pyogenes* ATCC 700294, and *S. agalactiae* ATCC 13813. Bacterial isolates belonging to the same genera and species and grown under the same conditions displayed similar, reproducible sets of peaks such as peaks at m/z 4451, 5951, and 8181 for *S. pyogenes* and *S. agalactiae*. In addition, a different pattern was revealed for each species, and these differences include peaks at m/z 6221, 7326, and 9527 for *E. faecalis*, peaks at m/z 4473, 5524, and 6887 for *S. aureus*, and peaks at m/z 6409, 7272, and 8056 for *E. coli*.

Statistical analyses improved interspecies and strain discrimination in *Streptococcus*

Differences between two species within the genus *Streptococcus* can be easily seen in Fig. 3. The peaks corresponding to m/z 3158, 6310, and 7962 were unique for *S. pyogenes*, while peaks corresponding to m/z 3609, 6180, and 7310 were unique for *S. agalactiae*. However, we preferred to use a statistical analysis of the MALDI-TOF MS data to quickly reveal significant differences among the strains analyzed. Streptococcal strains could be discriminated from each other by RF (Supplementary Fig. S1). Additionally, cluster analysis with PAST showed similarities in spectra and was used to cluster species and strains on the basis of these spectral similarities (Fig. 4). The RF analysis (Supplementary Fig. S1) yielded an estimated overall classification error of 1.7%, and the RF algorithm successfully classified all the bacterial species and the majority of the GAS strains in the training set. However, the RF algorithm was unable to differentiate a subset of *emm1* GAS strains isolated from patients with throat infection, and the estimated classification error for this subset was 22%. Results with three clustering algorithms within PAST, such as paired group, single linkage, and Ward's method, using similarity measure methods such as Euclidean, Dice, and Jaccard, among others, were analogous and quite consistent regardless of the analysis method applied, so that the different bacterial species and strains studied could be reliably and

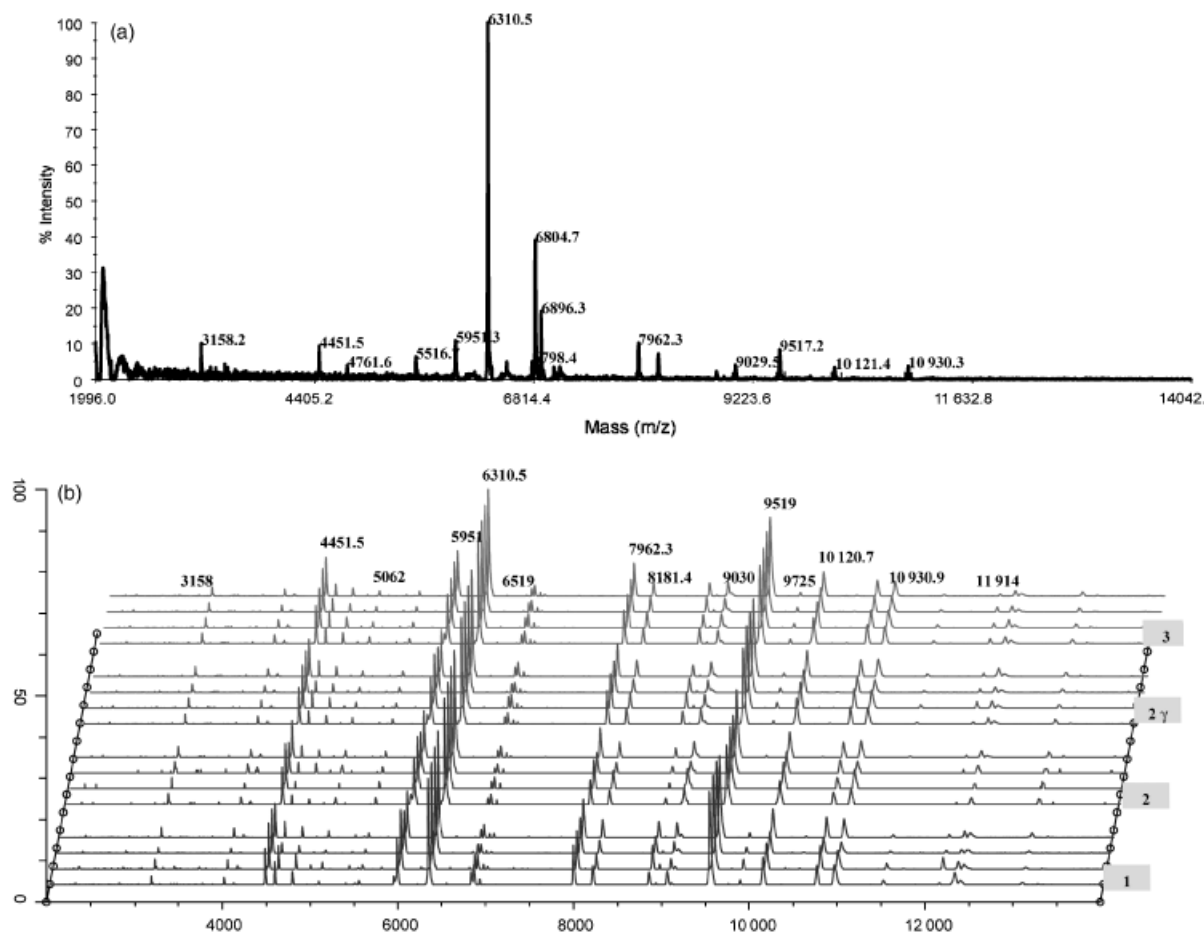


Fig. 1. MALDI-TOF mass spectra (m/z 2000–14 000 Da) for *Streptococcus pyogenes* ATCC 700294 whole cells grown in liquid media (THB). (a) Represents typical raw mass spectra of 10^7 cells. (b) Depicts the reproducible patterns in denoised spectra revealed using MULTISPEC VIEWER after analyzing several runs of *S. pyogenes* (batches 1, 2, and 3), including a subset from batch 2 that was γ irradiated. Analyses were in quadruplicate in each case.

reproducibly separated from each other (Fig. 4). As expected, GAS isolates were clustered together, and even subclusters could be detected, including the three strains isolated from necrotizing fasciitis cases that were grouped together, and were separate from noninvasive GAS isolates.

Tentative identification of peaks by database search

The peak lists generated using MULTISPEC VIEWER were compared with the CDC-modified filtered protein database; 97 tentative matches were uncovered, including several peaks listed as ribosomal proteins. A total of 21 ribosomal proteins detected in the MALDI-TOF mass spectra of each GAS strain analyzed are listed in Table 2. Tentative matches from the CDC-modified database filtered to select ribosomal proteins were overlaid with a typical summed spectrum of *S. pyogenes* (Supplementary Fig. S2). In addition, subsets of

summed m/z values of biomarkers detected from the first batch of *S. pyogenes* listed in Table 2 served as reference standards to evaluate both their batch-to-batch reproducibility and interference of growth media on biomarker expression. Similar and consistent results were obtained with processed peak lists of *E. coli* strain ATCC 25922, *S. aureus* strain ATCC 29213, and *E. faecalis* strain ATCC 29212, where many of the tentatively identified peaks also corresponded to ribosomal proteins (data not shown). The peak lists submitted to the 'Rapid Microorganism Identification Database' returned correct matching results for each genus/species tested.

Discussion

MALDI-TOF MS detection of biomarkers desorbed from whole cells is an emerging technique potentially applicable to the identification of microorganisms at different levels

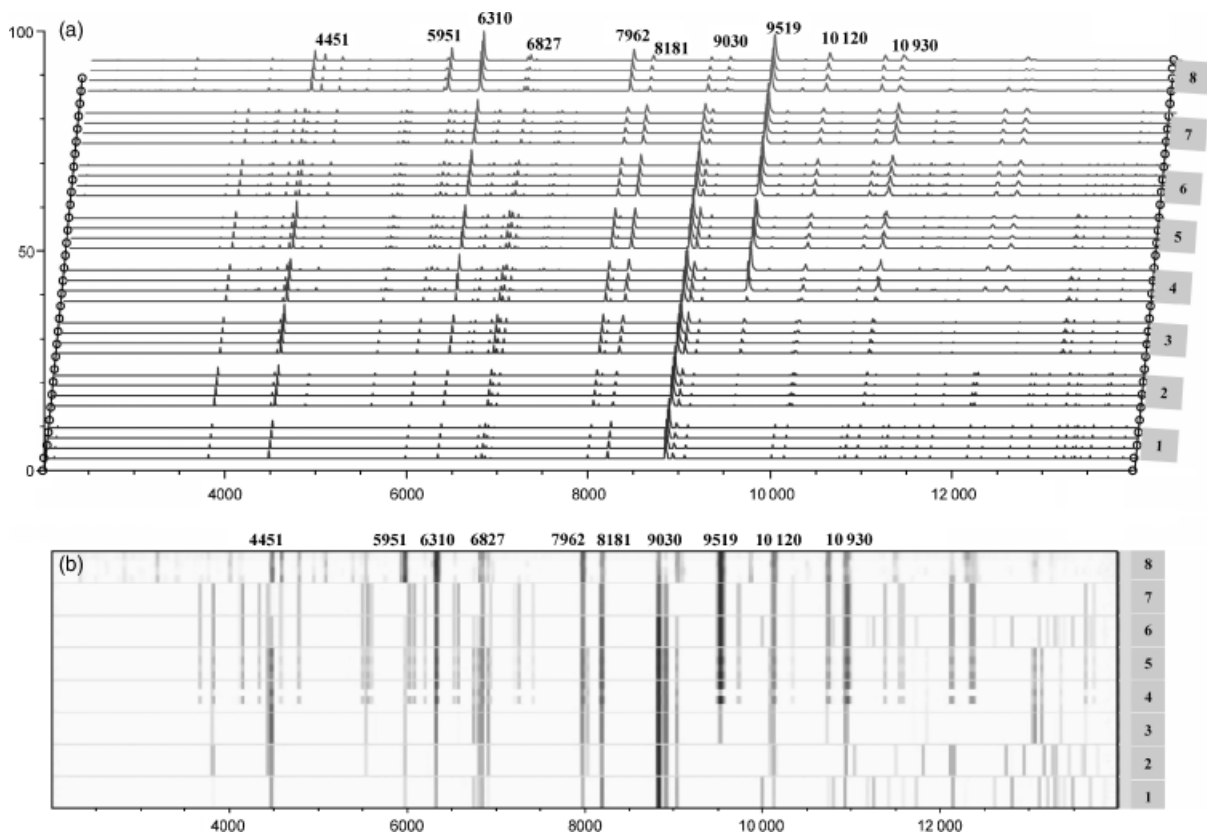


Fig. 2. MULTISPEC VIEWER depicting the influence of different colony numbers and culture media (blood agar and THB) in spectra variability of *Streptococcus pyogenes* ATCC 700294. (a) Mass spectra. Annotated denoised peaks are consistently seen in all wells. (b) Gel view feature to the mass spectra. From bottom to top, numbers 1, 2, 3, 4, 5, and 6 denote groups of four spectra obtained, respectively, from 1, 2, 3, 4, 5, and 30 colonies of *S. pyogenes* grown on blood agar and directly plated onto the MALDI target; numbers 7 and 8 are spectra obtained from washed cells from 30 colonies grown on blood agar (7) and 10^8 cells grown in THB (8).

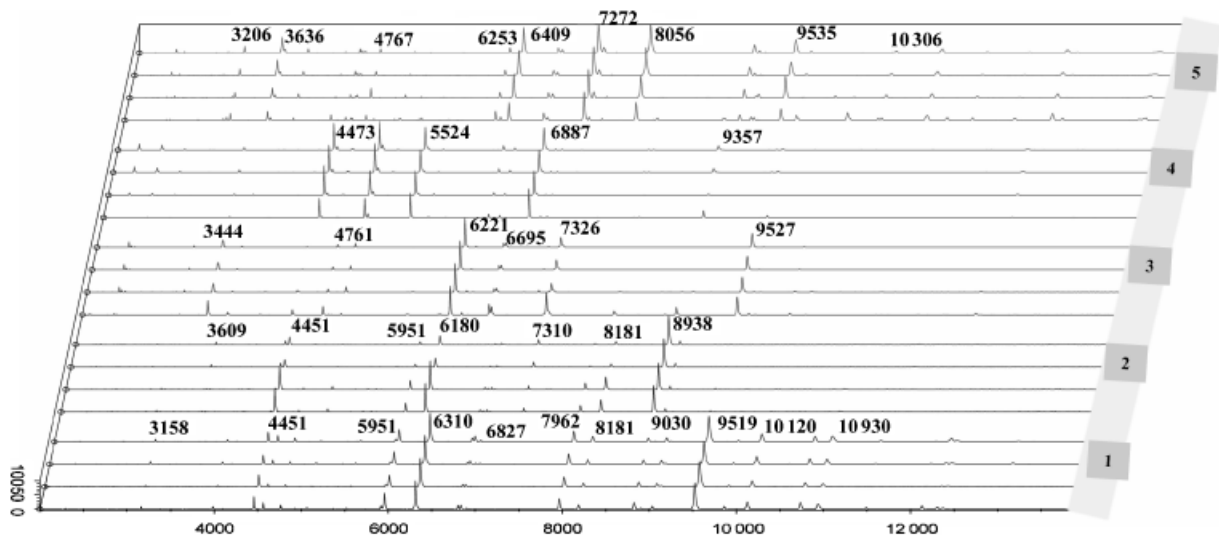


Fig. 3. MULTISPEC VIEWER of MALDI-TOF denoised mass spectra (m/z 2000–14 000 Da) for whole cells of (1) *Streptococcus pyogenes* ATCC 700294, (2) *Streptococcus agalactiae*, (3) *Enterococcus faecalis*, (4) *Staphylococcus aureus*, and (5) *Escherichia coli*, revealing differences among genus and species. Quadruplicate analyses are shown for each organism.

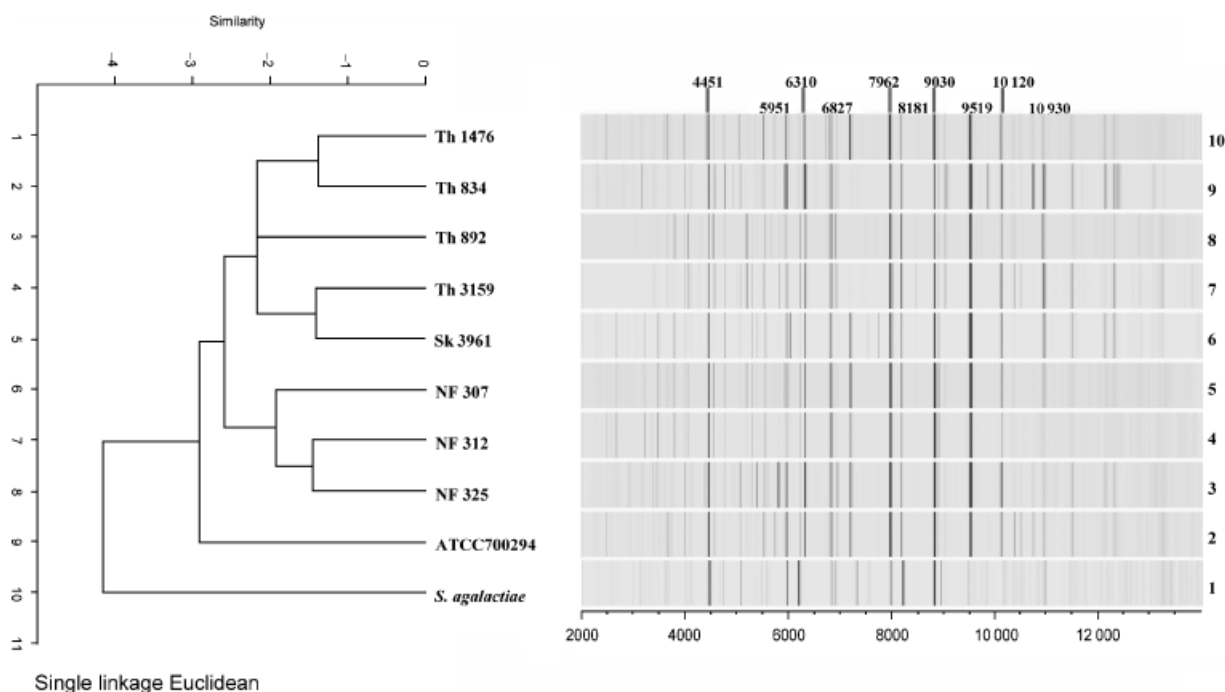


Fig. 4. Dendrogram obtained with PAST (<http://folk.uio.no/ohammer/past/index.html>) and corresponding gel view of mass spectra using MULTISPEC VIEWER, where differences between *Streptococcus agalactiae* and *Streptococcus pyogenes* could be detected visually. GAS isolates were clustered together and subclusters of invasive and noninvasive isolates of *S. pyogenes* could be detected. For this analysis, all spectra for each strain were summed to give one representative spectrum per organism.

(van Baar, 2000; Fenselau & Demirev, 2001; Lay, 2001). An increasing number of reports have shown the successful use of fingerprinting by comparing statistically processed mass spectra with a reference database (Jarman *et al.*, 1999, 2000; Dickinson *et al.*, 2004; Satten *et al.*, 2004; Valentine *et al.*, 2005; Wunschel *et al.*, 2005a, b). Other reports indicate that robust identification can be achieved on the basis of proteome database queries (Demirev *et al.*, 1999, 2001; Pineda *et al.*, 2000, 2003; Pribil & Fenselau, 2005; Pribil *et al.*, 2005). In a recent report, consistently correct identifications were obtained when only the most abundantly expressed proteins (e.g. ribosomal proteins) were used in the database (Pineda *et al.*, 2003). Once the MALDI-TOF MS spectra are collected, multiple forms of data analysis can be carried out on these data to discriminate the genus, species, and even strains of the microorganisms.

In the present study, using a high-resolution MALDI-TOF MS instrument along with a combination of mass spectral visual analysis, fingerprinting, and the proteomics approach, we were able to detect and tentatively identify protein biomarkers of nine selected GAS isolates, including *S. pyogenes* strain ATCC 700294, the genome strain. Processed spectra were used for visual and statistical analyses, and they were used to query proteome databases, including a CDC-modified filtered database.

Reproducibility of MALDI-TOF MS spectra is an important concern that has been addressed in a number of studies (Gantt *et al.*, 1999; Saenz *et al.*, 1999; Evason *et al.*, 2000; Jarman *et al.*, 2000; Fenselau & Demirev, 2001; Valentine *et al.*, 2005; Wunschel *et al.*, 2005a, b). In the present study, different MALDI preparations were initially tested using ATCC 700294 and varying solvents, matrices, plating techniques, and mass ranges. Spectral variability was analyzed using the bacterial cells grown under different conditions and culture media. Overall, the spectra of replicate samples were similar, even spectra acquired from wells with different concentrations of bacteria or harvested on different days and from various culture media. γ Irradiation did not seem to alter peaks detected by MALDI-TOF MS and is an alternative to process virulent isolates before analysis by mass spectrometry. A common subset of biomarkers was found consistently with all the conditions studied, and these biomarkers were consistent with previously reported studies (Valentine *et al.*, 2005). MALDI-TOF MS patterns of *S. pyogenes* were reproducible within sets of replicates and over time, thus indicating the potential of this method as a sensitive, specific, and rugged assay for *S. pyogenes* analysis.

The usefulness of the combined algorithms for data processing applied in this study, along with previously reported methods for spectral standardizing and denoising

Table 2. Ribosomal proteins detected in the MALDI-TOF MS of each *Streptococcus pyogenes* strain analyzed

m/z	Isolate									ID	
	1	2	3	4	5	6	7	8	9		
4451.5											RL36_STRPY ¹
5361.4											RL34_STRPY
5912.9											RL33_STRPY
6310.5											Q9A1V6 (RL30)
6733.7											RL32_STRPY
6797.0											RL28_STRPY
6841.0											RS21_STRPY ²
6942.3											R14A_STRPY
7333.7											RL35_STRPY
7962.3											RL29_STRPY
8181.4											RS20_STRPY
9072.6											RS18_STRPY
9854.0											R31B_STRPY
10120.7											RS16_STRPY
10491.1											RS19_STRPY
10930.9											RL24_STRPY
10950.4											RS6_STRPY ³
11482.4											RS10_STRPY
12124.8											RL7_STRPY
12299.4											RL22_STRPY
13294.5											RS13_STRPY

Isolates: 1, ATCC700294; 2, NF307-03; 3, NF312-03; 4, NF325-03; 5, Th892-97; 6, Th834-97; 7, Th1476-97; 8, Th3159-00; 9, Sk3961-00. Custom algorithms within MULTISPEC VIEWER were used to generate peak lists from the acquired mass spectra and to match these peaks. A total of 21 biomarkers corresponding to ribosomal proteins were tentatively identified by use of the CDC-modified database. Dark boxes correspond to each strain of *Streptococcus pyogenes* where the biomarkers were detected. RL36_STRPY¹, 50S ribosomal protein RL36; RS21_STRPY², 30S ribosomal protein S21; RS6_STRPY³, RP detected only in NF and M12 strains.

(Satten *et al.*, 2004), was demonstrated by the robust differentiation by RF of all species and most of the strains tested. In addition, plotted peaks pointed to the existence of strain-specific biomarkers. This plot can be used to determine which peaks should be investigated further for biomarker identification by sequencing. Dendrograms obtained with the PAST cluster software confirmed the results from RF, indicating the potential of MALDI-TOF MS for discrimination among genus, species, and strains.

The genome of at least 12 strains of *S. pyogenes* has been completely sequenced; thus, tentative peak identification was easily accomplished using the CDC-modified filtered protein entries from Swiss-Prot/TrEMBL and UNIPROT databases. Our finding that the several peaks tentatively identified corresponded to ribosomal proteins is intriguing because it points to another advantage of the method as it

uncovers several important gene products in only one experiment. Ribosomal proteins are extremely ancient molecules, and one-third of ribosomal protein families are conserved among *Bacteria*, *Archae*, and *Eucharia* (Lecompte *et al.*, 2002). In addition, our findings pointed to the usefulness of small protein analysis in the range of 2000–14 000 Da. These small proteins have not been explored in proteomic studies using two-dimensional gel electrophoresis to detect differences/similarities among organisms (Patterson & Aebersold, 2003; Lambert *et al.*, 2005), but they are promising biomarkers, as the sequenced genomes of the *S. pyogenes* isolates (<http://www.ncbi.nlm.nih.gov/entrez/>) predict an average of 650 proteins in the 2000–14 000 Da molecular mass range, a third of them with an estimated pI 7.0 or higher.

MALDI-TOF MS-based methods are very sensitive and have been used before for genus and species differentiation, but they have rarely been used for strain characterization (Stackebrandt *et al.*, 2005). Here we report one of the first uses of MALDI-TOF MS for analyzing GAS strains that were discriminated by this method although they were closely related by a genetic typing (*emm* typing), traditionally used for tracing GAS isolates. Additionally, once MALDI-TOF MS spectra are collected, they can be reanalyzed at a later date as more species and strains have their genomes sequenced or after more advanced statistical analyses become available.

GAS strains have been analyzed by different methods, but each has limitations, and thus far, all have failed to differentiate strains according to their sources (invasive and noninvasive) (Beres *et al.*, 2006). In the present work, invasive GAS isolates from cases of necrotizing fasciitis were clustered together, yet separate from isolates associated with mild, noninvasive throat and skin infections, despite sharing the same *emm* type. There was no correlation between the *emm* type and the data obtained by mass spectrometry. We plan to further investigate the consistency of these findings using a larger number of strains isolated from patients with necrotizing fasciitis and will investigate the possibility of disclosing the existence of a subset of biomarkers common to these invasive isolates, whose identity can be further confirmed by sequencing with tandem mass spectrometry analysis (MS/MS). These initial results suggest that specific biomarkers can be used for early detection of the disease. GAS is the most common causative agent of necrotizing fasciitis, a rare but disfiguring and frequently lethal condition. Analysis of the nine selected GAS strains suggests the existence of a group of unique spectral markers that can be useful to characterize GAS isolated from NF cases. To our knowledge, this is the first report with consistent evidence of biomarker differences among invasive and noninvasive GAS isolates.

In conclusion, we have developed a methodology for proteomic analysis of bacteria using MALDI-TOF MS with

statistical analysis. Visual analysis of multiple spectra was sufficient to differentiate among the different genera and species analyzed. However, the combined MS visual analysis, fingerprinting, and proteomics approach yielded a more robust biomarker identification using a database search, which allowed strain differentiation among GAS isolates, including the separation of a group of isolates that caused necrotizing fasciitis from another group of isolates associated with noninvasive infections. Small ribosomal proteins were consistently found among biomarkers detected in this study. Even though the current study used a limited number of carefully selected, well-characterized, and diverse group of GAS isolates, we believe that the data suggest that MALDI-TOF MS may constitute a powerful tool to distinguish GAS isolates and to investigate virulence-related characteristics, reinforcing the concept that this method is a rapid, reproducible, high-throughput alternative for the characterization of microorganisms.

Acknowledgement

We wish to thank Dr Richard Facklam for many helpful discussions and his valuable advice during the course of this investigation.

Statement

Reuse of this article is permitted in accordance with the Creative Commons Deed, Attribution 2.5, which does not permit commercial exploitation.

References

- Aebersold R & Mann M (2003) Mass spectrometry-based proteomics. *Nature* **422**: 198–207.
- Amiri-Eliasi B & Fenselau C (2001) Characterization of protein biomarkers desorbed by MALDI from whole fungal cells. *Anal Chem* **73**: 5228–5231.
- Beall B, Facklam R, Hoenes T & Schwartz B (1997) Survey of *emm* gene sequences and T-antigen types from systemic *Streptococcus pyogenes* infection isolates collected in San Francisco, California; Atlanta, Georgia; and Connecticut in 1994 and 1995. *J Clin Microbiol* **35**: 1231–1235.
- Beres SB, Richter EW, Nagiec MJ, Sumbly P, Porcella SF, DeLeo FR & Musser JM (2006) Molecular genetic anatomy of inter- and intraserotype variation in the human bacterial pathogen group A *Streptococcus*. *Proc Natl Acad Sci USA* **103**: 7059–7064.
- Currie BJ (2006) Group A streptococcal infections of the skin: molecular advances but limited therapeutic progress. *Curr Opin Infect Dis* **19**: 132–138.
- Demirev PA, Ho YP, Ryzhov V & Fenselau C (1999) Microorganism identification by mass spectrometry and protein database searches. *Anal Chem* **71**: 2732–2738.
- Demirev PA, Lin JS, Pineda FJ & Fenselau C (2001) Bioinformatics and mass spectrometry for microorganism identification: proteome-wide post-translational modifications and database search algorithms for characterization of intact *H. pylori*. *Anal Chem* **73**: 4566–4573.
- Dickinson DN, La Duc MT, Satomi M, Winefordner JD, Powell DH & Venkateswaran K (2004) MALDI-TOFMS compared with other polyphasic taxonomy approaches for the identification and classification of *Bacillus pumilus* spores. *J Microbiol Methods* **58**: 1–12.
- Doktor SZ, Beyer JM, Flamm RK & Shortridge VD (2005) Comparison of *emm* typing and ribotyping with three restriction enzymes to characterize clinical isolates of *Streptococcus pyogenes*. *J Clin Microbiol* **43**: 150–155.
- Evason DJ, Claydon MA & Gordon DB (2000) Effects of ion mode and matrix additives in the identification of bacteria by intact cell mass spectrometry. *Rapid Commun Mass Spectrom* **14**: 669–672.
- Facklam R (2002) What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin Microbiol Rev* **15**: 613–630.
- Facklam R, Beall B, Efstratiou A *et al.* (1999) *emm* typing and validation of provisional M types for group A streptococci. *Emerg Infect Dis* **5**: 247–253.
- Fenselau C & Demirev PA (2001) Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom Rev* **20**: 157–171.
- Gantt SL, Valentine NB, Saenz AJ, Kingsley MT & Wahl KL (1999) Use of an internal control for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis of bacteria. *J Am Soc Mass Spectrom* **10**: 1131–1137.
- Gingras AC, Aebersold R & Raught B (2005) Advances in protein complex analysis using mass spectrometry. *J Physiol* **563**: 11–21.
- Holland RD, Wilkes JG, Rafii F, Sutherland JB, Persons CC, Voorhees KJ & Lay JO Jr (1996) Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **10**: 1227–1232.
- Jarman KH, Daly DS, Petersen CE, Saenz AJ, Valentine NB & Wahl KL (1999) Extracting and visualizing matrix-assisted laser desorption/ionization time-of-flight mass spectral fingerprints. *Rapid Commun Mass Spectrom* **13**: 1586–1594.
- Jarman KH, Cebula ST, Saenz AJ, Petersen CE, Valentine NB, Kingsley MT & Wahl KL (2000) An algorithm for automated bacterial identification using matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* **72**: 1217–1223.
- Krader P & Emerson D (2004) Identification of archaea and some extremophilic bacteria using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. *Extremophiles* **8**: 259–268.
- Lambert JP, Ethier M, Smith JC & Figeys D (2005) Proteomics: from gel based to gel free. *Anal Chem* **77**: 3771–3787.
- Lay JO Jr (2001) MALDI-TOF mass spectrometry of bacteria. *Mass Spectrom Rev* **20**: 172–194.

- Lecompte O, Ripp R, Thierry JC, Moras D & Poch O (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* **30**: 5382–5390.
- Liu H, Du Z, Wang J & Yang R (2007) Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl Environ Microbiol* **73**: 1899–1907.
- Magnuson ML, Owens JH & Kelty CA (2000) Characterization of *Cryptosporidium parvum* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl Environ Microbiol* **66**: 4720–4724.
- McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B & Bessen DE (2004) Multilocus sequence typing of *Streptococcus pyogenes* representing most known emm types and distinctions among subpopulation genetic structures. *J Bacteriol* **186**: 4285–4294.
- Moura H, Ospina M, Woolfitt AR, Barr JR & Visvesvara GS (2003) Analysis of four human microsporidian isolates by MALDI-TOF mass spectrometry. *J Eukaryot Microbiol* **50**: 156–163.
- Patterson SD & Aebersold RH (2003) Proteomics: the first decade and beyond. *Nat Genet* **33** (Suppl), 311–323.
- Pignone M, Greth KM, Cooper J, Emerson D & Tang J (2006) Identification of mycobacteria by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. *J Clin Microbiol* **44**: 1963–1970.
- Pineda FJ, Lin JS, Fenselau C & Demirev PA (2000) Testing the significance of microorganism identification by mass spectrometry and proteome database search. *Anal Chem* **72**: 3739–3744.
- Pineda FJ, Antoine MD, Demirev PA, Feldman AB, Jackman J, Longenecker M & Lin JS (2003) Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. *Anal Chem* **75**: 3817–3822.
- Pribil P & Fenselau C (2005) Characterization of enterobacteria using MALDI-TOF mass spectrometry. *Anal Chem* **77**: 6092–6095.
- Pribil PA, Patton E, Black G, Doroshenko V & Fenselau C (2005) Rapid characterization of *Bacillus* spores targeting species-unique peptides produced with an atmospheric pressure matrix-assisted laser desorption/ionization source. *J Mass Spectrom* **40**: 464–474.
- Ruelle V, Moulaj BE, Zorzi W, Ledent P & Pauw ED (2004) Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **18**: 2013–2019.
- Saenz AJ, Petersen CE, Valentine NB, Gantt SL, Jarman KH, Kingsley MT & Wahl KL (1999) Reproducibility of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for replicate bacterial culture analysis. *Rapid Commun Mass Spectrom* **13**: 1580–1585.
- Satten GA, Datta S, Moura H, Woolfitt AR, Carvalho Mda G, Carlone GM, De BK, Pavlopoulos A & Barr JR (2004) Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens. *Bioinformatics* **20**: 3128–3136.
- Shaw EI, Moura H, Woolfitt AR, Ospina M, Thompson HA & Barr JR (2004) Identification of biomarkers of whole *Coxiella burnetii* phase I by MALDI-TOF mass spectrometry. *Anal Chem* **76**: 4017–4022.
- Stackebrandt E, Pauker O & Erhard M (2005) Grouping myxococci (*Coralloccoccus*) strains by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry: comparison with gene sequence phylogenies. *Curr Microbiol* **50**: 71–77.
- Teixeira LM, Barros RR, Castro AC, Peralta JM, Carvalho MGS, Talkington DF, Vivoni AM, Facklam RR & Beall B (2001) Genetic and phenotypic features of *Streptococcus pyogenes* strains isolated in Brazil that harbor new emm sequences. *J Clin Microbiol* **39**: 3290–3295.
- Valentine N, Wunschel S, Wunschel D, Petersen C & Wahl K (2005) Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. *Appl Environ Microbiol* **71**: 58–64.
- van Baar BL (2000) Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiol Rev* **24**: 193–219.
- Wunschel DS, Hill EA, McLean JS, Jarman K, Gorby YA, Valentine N & Wahl K (2005a) Effects of varied pH, growth rate and temperature using controlled fermentation and batch culture on matrix assisted laser desorption/ionization whole cell protein fingerprints. *J Microbiol Methods* **62**: 259–271.
- Wunschel SC, Jarman KH, Petersen CE, Valentine NB, Wahl KL, Schauki D, Jackman J, Nelson CP & White ET (2005b) Bacterial analysis by MALDI-TOF mass spectrometry: an inter-laboratory comparison. *J Am Soc Mass Spectrom* **16**: 456–462.

Supplementary material

The following Supplementary material is available for this article online:

Fig. S1. Results of classification of datasets using RF.

Fig. S2. Tentative protein identification using the CDC-modified database.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1574-695X.2008.00428.x> (This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.