Available online at www.sciencedirect.com
ScienceDirect
International Journal of Mass Spectrometry 259 (2007) 140–146
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Specific identification of *Bacillus anthracis* strains

Thaiya Krishnamurthy [a,*], Samir Deshpande [b], Johannes Hewel [c], Hongbin Liu [c], Charles H. Wick [a], John R. Yates III [c]

[a] *Edgewood Chemical and Biological Center, Aberdeen Proving Ground, MD 21010-5424, USA*
[b] *Science & Technology Corporation, Edgewood, MD 21040, USA*
[c] *Scripps Research Institute, La Jolla, CA 92037, USA*

Received 25 May 2006; received in revised form 5 September 2006; accepted 8 September 2006
Available online 9 October 2006

Abstract

Accurate identification of human pathogens is the initial vital step in treating the civilian terrorism victims and military personnel afflicted in biological threat situations. We have applied a powerful multi-dimensional protein identification technology (MudPIT) along with newly generated software termed Profiler to identify the sequences of specific proteins observed for few strains of *Bacillus anthracis*, a human pathogen. Software termed Profiler was created to initially screen the MudPIT data of *B. anthracis* strains and establish the observed proteins specific for its strains. A database was also generated using Profiler containing marker proteins of *B. anthracis* and its strains, which in turn could be used for detecting the organism and its corresponding strains in samples. Analysis of the unknowns by our methodology, combining MudPIT and Profiler, led to the accurate identification of the *anthracis* strains present in samples. Thus, a new approach for the identification of *B. anthracis* strains in unknown samples, based on the molecular mass and sequences of marker proteins, has been ascertained.
Published by Elsevier B.V.

Keywords: *Bacillus anthracis*; Strains; MudPIT; Profiler; Identification

1. Introduction

Bacillus anthracis is one of the human pathogens established to be potential biological warfare agents [1]. More recently, it has also known to be used against civilian targets [2]. Accurate determination of the identity of the agent is vital for the successful treatment of the victims. Recently, mass spectrometric analysis of pathogenic and non-pathogenic bacteria has been investigated in detail [3–20]. We had established the identity of the protein biomarkers for some human pathogens using complex procedure for the isolation of bacterial proteins and matrix assisted laser desorption ionization-mass spectrometric (MALDI-MS) method [4]. Genus and species specific marker proteins were established. However, extraction of protein markers was extensive and time consuming. Alternatively, we mixed the intact bacterial cells with UV absorbing matrix (sinapinic acid) and subjected to direct MALDI-MS analysis [5]. The signals were better and more proteins in larger amounts were observed than in the previous method [5]. Proteins specific for genus, species and strains (1–3 proteins) were established by analyzing the recorded spectra ([5], T. Krishnamurthy unpublished results). Only smaller proteins, less than 30 kDa in molecular masses were observed in both of these studies [4,5]. We also demonstrated that simple ultrasonic disruption of the intact cell suspensions for 30 s released proteins, and the lysate can be directly analyzed by electrospray ionization (ESI)-MS methods after a brief liquid chromatographic (LC) separation to identify smaller proteins [6]. However, it has been demonstrated recently that addition of surfactants to the sample aided in the observation of larger proteins by MALDI-TOF mass spectrometry [7]. More recently, *B. anthracis* organism was identified and verified by the isolation of a larger antigenic protein EA1, with molecular mass of 91,362 Da, by affinity chromatography using monoclonal antibody for the organism and analysis by electrospray-ion trap mass spectrometric technique [8]. Beads coated with anti-*Bacillus* antibody was also used to isolate the organism from milk and identify by MALDI-TOF analysis [9]. *Bacillus* spores were identified recently by peptide mapping using a “Tiny TOF” mass spectrometer [10]. Microwave assisted acid hydrolysis and

* Corresponding author at: 10225 Clubhouse Court, Ellicott City, MD 21042, USA. Tel.: +1 410 461 7551; fax: +1 410 436 1912.
E-mail addresses: Thaiya.krishnamurthy@us.army.mil, Thaiyakrishnamurthy@yahoo.com (T. Krishnamurthy).

1387-3806/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.ijms.2006.09.013

SDS-PAGE separation in combination with mass spectrometry has also been applied for the identification of *Bacillus* species [11–13]. Species and two strain specific larger glycoproteins were isolated by the SDS-PAGE separation of *Bacillus* lysates and identified by MALDI-TOF analysis [12]. Atmospheric pressure MALDI-ion trap mass spectral analysis of peptides from *B. anthracis* lysate has also been reported [14].

In most of the above investigations, the identities of proteins were established based on their molecular masses [3–10,12,13]. Since a larger number of proteins were observed for individual bacterial cells, there is a probability for observing different marker proteins with common molecular masses, especially among lower mass proteins. This would complicate the identification of component bacteria especially during mixture analysis since there is a probability for two different marker proteins, belonging to separate organisms, to have identical molecular mass. Hence, the individual bacteria, especially the corresponding strain, in mixtures cannot be established unambiguously using the molecular masses of the bacterial pathogens. Strain recognition leading to the identification of the source of the organism is vital while investigating biological attack under military and bioterrorism conditions. Hence, investigations with the structures (sequences) by tandem mass spectrometry, instead of the molecular masses, would be more appropriate for the identification of pathogens including their corresponding strains.

MALDI-TOF-, ESI-MS/MS spectra of the peptides originating from lysates of *Bacillus* species indicated the biomarkers to be small acid soluble proteins [14–17]. On probe enzymatic digestion of the proteins using immobilized trypsin, during MALDI-MS/MS analysis, has also been applied to reduce the digestion time involved in the conventional enzymatic cleavage procedures [17]. Identification of single and dual components of *Bacillus* species has been accomplished by shotgun proteomics procedures [18,19]. SDS-PAGE separation of vegetative *B. anthracis* cell lysate followed by the enzymatic digestion and ESI-MS/MS analysis identified over one thousand specific proteins [20]. The databases containing proteins of vegetative and sporulated *B. anthracis* were applied for the proteomic analysis of the causative agent of anthrax [20]. Multi-dimensional protein identification technology or MudPIT [21,22] had been used for the identification of marker proteins in anthrax endospores [23]. In all of the above investigations, protein database search has been applied for the identification of proteins in bacterial species [14–23]. Strain distinction among *B. anthracis* organisms has been attributed to the characterization of two large mass glycoproteins [12]. This is not sufficient to distinguish among several known strains of *B. anthracis*. Even though during the above explorations [14–20,23] the ability to identify the bacterial species has been well demonstrated, distinction of the corresponding strains requires more careful studies.

In this investigation, initially we converted the proteins present in *B. anthracis*, Ames into their corresponding peptides by chemical treatment followed by the enzymatic cleavages. The peptide mixture was as such subjected to analysis by multi-dimensional protein identification technique (MudPIT) [21,22]. Over 500 proteins, small and large, of various pIs and functionality were identified based on their sequences, as a result of the *B. anthracis* TIGR database search. Similarly, during this investigation lysates from different strains of *B. anthracis* were analyzed and the proteins identified based on their sequences. Software designated as *Profiler* was generated for further analysis of MudPIT data of various strains of *B. anthracis*, to distinguish marker proteins including several unique strain specific proteins. A database containing the common proteins and strain specific proteins for *B. anthracis* was also developed using Profiler. It was also used to prescreen the MudPIT data of samples and to ascertain whether the sample is *B. anthracis*, followed by identification of the strain. Some unknowns were analyzed and the strains were correctly identified. Thus, we have an unambiguous system to identify *B. anthracis* strains by detecting the strain specific marker proteins of various sizes, pI and sequences. The same approach can also be applied for the identification of other human pathogens and their corresponding strains, which would have enormous application potential in the identification of biological threat agents.

2. Experimental

2.1. Bacterial lysate

Accurately weighed gamma radiated *B. anthracis* cells (1 mg; 2.0×10^8 cells/mg) were treated with trifluoroacetic acid (1%; 400 μl) and subjected to ultrasonic disruption for 30 s. The solution was treated with additional aqueous trifluoroacetic acid (1%; 600 μl) and vortexed for 30 s. The solution was centrifuged for 5 min at 14,000 rpm and stored at −20 °C until use. One hundred microlitre of the solution was used for further investigation.

2.2. Proteolysis of bacterial lysate

Bacterial lysate (100 μl) was treated with ammonium bicarbonate (100 mM; 900 μl) and the pH of the solution was adjusted to 8.5. Solid urea was added to make the solution 8.0 M in urea. Disulfide bonds were reduced using 1 M Tris[2-carboxyethyl]phosphine solution (2 μl) at ambient temperature for 30 min. It was further carboxamidomethylated by 500 mM iodoacetamide solution (10 μl) in the dark at ambient temperature for 30 min. The carboxamidomethylated protein solution was then treated with endoproteinase Lys-C (10 μl) and maintained at 37 °C for 4 h. The solution was then diluted with ammonium bicarbonate solution (3 ml) and the pH was adjusted to 8.5. It was further treated with 1 M calcium chloride (5 μl) and porozyme immobilized trypsin beads (10 μl) and left it overnight at 37 °C. The digestion mixture was centrifuged at 14,000 rpm for 30 min to separate the beads. The supernatant containing the tryptic peptide fragments was used for MudPIT analysis.

2.3. MudPIT [21,22] analysis

Solution containing tryptic peptides was loaded, using a pressure bomb, on a loading column containing strong cation exchange resin (3 cm; SCX; Partisphere 5 μm) and Gemini C18 phase (3 cm; Phenomenex, Torrence, CA) in 250 μm i.d.

capillary closed with a filter union (Upchurch, Oak Harbor, WA). The loading column was connected to an analytical column (100 μm i.d.; 10 cm) filled with C18 particles (3 μm) and with a tip (ca. 10 μm). The analytical column was mounted in front of the LCQ Deca ion trap tandem mass spectrometer (Thermo Electron, San Jose, CA) operated by Xcalibur 1.3 software. The MS/MS spectra were acquired operating the mass spectrometer in the data dependant mode. The data dependant MS/MS spectra of three most intense ions were acquired. The *m*/*z* values of the intense ions were included in an exclusion list. Tandem mass spectra were extracted from the Xcalibur RAW files by RAWExtractor (in-house software written by John Venable, Scripps Research Institute, La Jolla, CA) and the spectra were brought together in ms2 files. Spectra from ms2 files were subjected to a SEQUEST database search using *B. anthracis* database downloaded from TIGR (http://www.tigr.org). Results were stored in sqt files. The identified proteins were filtered based on Xcorr values of (Xcorr > 3.5 > 2.5 > 1.8 and DeltCn of 0.8) using DTASelect v1.9 [24].

2.4. Identification of B. anthracis strains

Individual MudPIT analyzed file was selected by Profiler application for analysis and classification. In the Profiler application, the type of organism was selected to be "unknown" and the "Start" button was pressed to begin the analysis. The total number of unique matches and its corresponding percent of matches observed between the database and sample were displayed by the application on the right hand side of the screen layout. When the percent of match was above the threshold set by the user, the "Detect Strain" button was activated. When the user clicks on this button, the application started to analyze and detect the common matches within the individual unique strain biomarkers and the results were displayed in a tabular format with its percent match. The strain with the highest percent of match (score) was treated for the 100% match and a graph was plotted with relative percent match for the other strains in the database. The identified strain was the one with the 100% match.

3. Results and discussion

The proteins in the *B. anthracis*, Ames lysate were reduced and carboxamidomethylated followed by subjecting the products to enzymatic cleavages using Lys-C and porozyme immobilized trypsin beads. The resulting peptide mixture was resolved by a two-dimensional separation over a strong cationic exchange resin followed by a reverse phase C_{18} substrate. The well resolved peptides were directly introduced into the nanospray ionization source of an ion trap tandem mass spectrometer. The ionized peptides were subjected to collisionally induced dissociation and the corresponding MS/MS spectra were recorded and analyzed by SEQUEST and PepProbe [25] programs using TIGR *B. anthracis* database. The output was analyzed by DTAselect [24] to select and identify the proteins. As a result, the sequences of the identified peptides, total sequences of the corresponding proteins along with their molecular masses, pI and identities were derived and listed. The entire process was

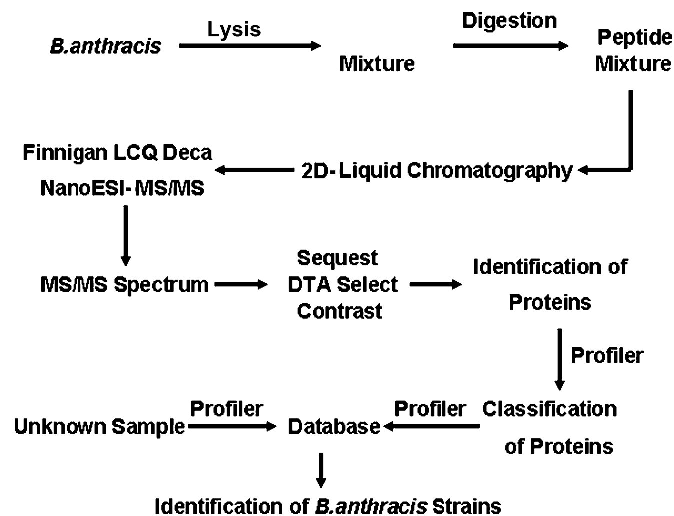

Fig. 1. Identification of *Bacillus anthracis* strains.

repeated three times with different samples of Ames strain to verify the reproducibility of the process. Similarly, experiments were also carried out with Sterne, VNR, Vollum and Zimbabwe strains and the proteins present in the individual lysate were identified in each case.

TIGR or NCBI protein database for *B. anthracis* does not provide the information for distinguishing its strains. Hence, in order to distinguish the individual strains, proteins specific for each strain need to be identified in addition to the protein sequences common to genus *Bacillus* and species *anthracis*. Databases containing marker proteins specific to a strain or species and genus of the bacteria need to be generated, saved and applied for the future identification of the organism present in unknown samples. An algorithm was developed for automated comparison of unknown samples with the database list. All of these had been accomplished as follows. The entire schematics for the investigation are indicated in Fig. 1.

Profiler is a data archiving and analysis software. It was designed and developed in-house for the purpose of archiving MudPIT sample data and to assist the researcher in establishing the classification of bacterial proteins and scoring of the proteins in an automated fashion and thereby simplifying the task of manually shifting through large amounts of data. The Profiler was also developed to automate sample screening which includes the identification of common and strain specific proteins, generation of database and automated identification of the *B. anthracis* strains in samples. Profiler has two (2) modules and is a very small size of just 88 KB, which makes it very compact and easy to load on any computer.

3.1. Module 1—database creation

This module creates bacteria specific databases, which contain proteins with common and unique molecular masses observed in specific bacteria including its strain. The molecular masses, pI and description of these marker proteins were established by the MudPIT analysis, based on their corresponding sequences. So, even though the molecular masses of the marker

proteins were used in the Profiler database, the sequences play a role in the identification of marker proteins, database creation and search. A Microsoft SQL Server relational database management system [http://www.microsoft.com/sql/default.mspx] was used to create the database and the application was written in Microsoft Visual Basic .NET [http://www.microsoft.com/net/default.mspx]. The *n* sample data from the same strain were analyzed to determine the *u* common molecular masses, within a tolerance of 1% of mass value. The researcher would be prompted for archiving the data after the data analysis had determined the list of common molecular masses. This data were displayed on the screen along with the observed number of repetitions. The observed molecular masses derived based on their corresponding sequences, pIs and identities were inserted into the database by the application of Profiler based on the user selection.

3.2. Method

All MudPIT sample data of the same strain of the bacteria were collected over a number of runs. This sample data were used for the application of the data analysis to determine proteins with common molecular masses. The algorithm to determine common masses, reads the sample data files sequentially and updates an array in memory with the molecular masses of the proteins along with their pI and description, and the number of repetitions. After all the selected strain files were read and information was written into the memory array, this array of data was displayed on the screen. The display contained the molecular masses of proteins, found in all the selected files, originally identified from MudPIT data, with the number of repetitions, along with pI and protein description values. The common proteins, obtained from the open reading frame (ORF), and denoted by the BA key, were archived along with pI value and protein description in a genus and species specific database. All of these proteins were evaluated and verified manually to ascertain correct substitution of BA values. In addition, a query was written to determine the unique proteins observed for a specific strain. This query was evaluating individual strain data in order to observe the molecular masses of proteins which were not present in any other strain and thus specific to a particular strain. This query was archived in the Microsoft SQL Server as a stored procedure so that it can be reused for other databases specific to various organisms. Thus, unique biomarkers for all strains were determined and grouped in a strain specific database. While mining the data for common protein masses, we observed 170 common proteins present in all investigated *anthracis* strains from a total of 2669 proteins observed during this study. The common proteins were assigned for *Bacillus* genus and *anthracis* species. In Table 1, different strains of *B. anthracis* and their observed total unique proteins are listed. Since the individual number of distinct proteins observed each strain is numerous, listing of the corresponding sequences and identities in the manuscript was not possible. Several strain specific large mass proteins were also detected. All marker proteins differed in their molecular masses, sequences, pI and functionality.

Table 1
List of unique biomarkers observed for *B. anthracis* strains

Strain	Unique biomarkers
Ames	578
Sterne	402
VNR	284
Vollum	291
Zimbabwe	264

3.3. Module 2—data analysis

Having established a set of unique strain specific proteins and common proteins observed for all *B. anthracis* strains, we designed and developed an algorithm to establish the strain present in samples from the observed MudPIT sample data. The algorithm was developed using Microsoft Visual Basic .NET to design the front end of the graphical user interface and the data processing was written in Microsoft C++ [http://msdn.microsoft.com/visualc/] for faster computations. Common marker proteins from MudPIT data file for a known *B. anthracis* sample were selected and stored in specific database along with the parameters, such as the database name, bacteria name, pI and description of the proteins. The strain specific proteins were stored in separate database. The in-house databases generated for *B. anthracis* and its strains were applied to analyze MudPIT data of the unknown sample. The default tolerance value for the database search was set at 1%. When the processing started, the MudPIT file of the sample was read and the values of masses found in the file were matched with the genus and species database values. The results were represented in a graphical format as well as stored in the memory array along with the number of hits for a particular mass. This process was applied to all the masses observed in the sample file.

3.4. Method

Fig. 2 shows a screen snapshot for the application, where an individual MudPIT analyzed file was selected for classification. To process the data type of organism, i.e., “Known” or “Unknown” was selected. Initially, since we knew the sample from our analysis to be *B. anthracis* strain ‘Vollum’, we selected “Known” check box for the type of sample being processed and selected the in-house ‘*Bacillus anthracis*’ database from the pull down list bar to compare the ‘Vollum’ data (Fig. 3). The results were displayed in the right hand corner with the common data (data in the sample file), total number of matches and percentage of match observed between the database and sample. A grid displayed the data from the sample and when there was a match, the data was shown in blue just for differentiation and visualization purposes.

If the percent of match was above a threshold set by the user, the “Detect Strain” button was enabled. When the user clicked on it, the application started to detect the common matches within individual unique strain biomarkers listed in the in-house database and the results were displayed in a tabular format with percent match. The strain with the highest percentage of matches

Rapid Profiler - [Protein Identifier]

File Help

MudPIT Analyzed File: J:\TK_BA Processed Data Files\Test\test5.csv

Search Parameters: (•) Known Sample ( ) Unknown Sample

Mass Accuracy Tolerance: ± 0

Threshold: 30 Detect Strain

Select Bacteria: Bacillus Anthracis

C:\Projects\RapidProfilerv110\Data\BA_Biomarker.mdb

Results

Total Records in Database: 170

Total Records in File: 333

43.84 % Match

Total Common Match is 146

Compare

Print

Close

Sr.No	ORF	Molecular Wt	pI	Description
1	ORF00006	66330	5	sensor histidine kinase ResE (resE)
2	ORF00015	36162	6	superoxide dismutase
3	ORF00077	35389	6	r specific 2-hydroxyacid dehydrogenase family protein
4	ORF00112	21373	5	conserved hypothetical protein
5	ORF00207	6506	9	small, acid-soluble spore protein, alpha/beta family
6	ORF00237	39520	6	porter, spermidine/putrescine-binding protein (potD)
7	ORF00272	106519	6	oxoglutarate dehydrogenase, E1 component (odhA)
8	ORF00308	19877	7	conserved hypothetical protein
9	ORF00314	17331	5	hypothetical protein
10	ORF00352	70365	5	oligoendopeptidase F (pepF-1)
11	ORF00365	38089	6	oligopeptide ABC transporter, ATP-binding protein
12	ORF00400	18387	9	conserved hypothetical protein
13	ORF00485	26774	6	conserved hypothetical protein
14	ORF00504	39194	6	uroporphyrinogen decarboxylase (hemE)
15	ORF00539	32456	8	protein export protein prsA (prsA-1)
16	ORF00560	31407	6	conserved hypothetical protein
17	ORF00622	62574	5	dihydroxyacetone kinase family protein
18	ORF00735	13592	5	conserved hypothetical protein
19	ORF00741	6966	6	small acid-soluble spore protein (sasP-1)
20	ORF00742	27358	8	amino acid ABC transporter, ATP-binding protein
21	ORF00804	8410	9	cotJA protein (cotJA)
22	ORF00806	21651	5	cotJC protein (cotJC)
23	ORF00855	56195	6	stage V sporulation protein R (spoVR)
24	ORF00913	110947	8	transporter, AcrB/AcrD/AcrF family
25	ORF00930	33336	8	quinol oxidase, subunit II (qoxA)
26	ORF00931	72694	8	quinol oxidase, subunit I
27	ORF00966	23069	6	transaldolase, putative
28	ORF00974	31520	5	ribokinase (rbsK)
29	ORF00997	16700	6	HIT family protein
30	ORF01002	30301	9	nino acid ABC transporter, amino acid-binding protein
31	ORF01015	65203	6	sensory box/GGDEF family protein
32	ORF01039	52597	6	aspartate ammonia-lyase (aspA-1)
33	ORF01040	40959	5	CBS domain protein

Strain Analysis

Sr.No	Name	Tot.Rec	Match	< % >
1	Ames	578	109	18.86
2	Sterne	402	108	26.87
3	vnr	284	92	32.39
4	Vollum	291	36	12.37
5	Zimbabwe	264	183	69.32

Graph

J:\TK_BA Processed Data Files\Test\test5.csv

Rel. %: 0, 20, 40, 60, 80, 100, 120

Strain: 1, 2, 3, 4, 5

2:35 PM | 3/20/2006 | CAPS | NUM | INS | Samir Deshpande & Thaiya Krishnamurthy | Help: Click to detect the strain information

Fig. 2. Snapshot of Profiler.

was designated the 100% match and a graph was plotted with the relative percent match for the strains. Having verified the algorithm was working correctly with multiple analyses of the ‘Vollum’ strains, we tested other strains of *B. anthracis* (Sterne, Ames, VNR and Zimbabwe) to validate the method. Results of the analysis for Vollum strain are shown in Fig. 3.

When MudPIT data of unknown sample, containing *B. anthracis* with unknown strain, were analyzed the “Unknown” option was selected from the parameters list. The MudPIT data from unknown samples were subjected to Profiler analysis to determine if the strain of *B. anthracis* could be determined. Initially, more than 50% of the observed sample proteins matched with that of marker proteins of *B. anthracis* and as a result the procedure for the detection of the strain was prompted. When the button was pressed, it screened through all lists of unique strain specific proteins along with their corresponding molecular masses and matched it with a built-in scoring model. The mathematical equation for computing the closest match is:

$$\text{Score [ps]} = \frac{\text{total number of match } (t)}{\text{total number of records in the database } (t_\text{d})} \quad (1)$$

$$\text{CloseMatch} = [\text{ps}] \times \frac{100}{\text{total number of entries in the sample}} \quad (2)$$

The results are illustrated in a graphical format (Fig. 4) plotting percent matches for all strains keeping the highest score to be relative abundance at 100%. Thus, we were able to detect the specific strain to be Zimbabwe and eliminating the other strains with lower scores. Similarly, we processed four other unknown MudPIT data and the strains were correctly identified in each instance of the unknown sample introduced and none of these data are shown here. Stepwise detection of *B. anthracis* and its corresponding strain in samples had been found to be adequate.

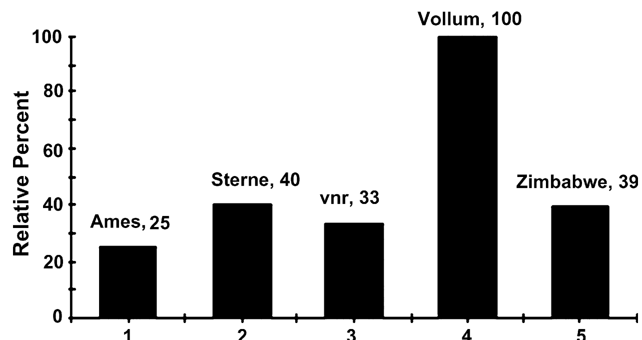

Fig. 3. Profiler analysis of known sample (Vollum).

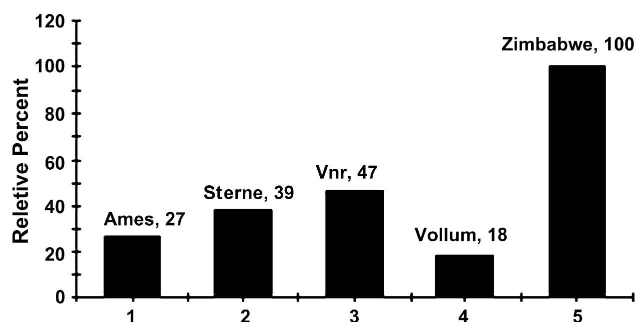

Fig. 4. Profiler analysis of unknown strain (Zimbabwe).

Our observations during other investigations are as follows. Different proteins, less than 25, were reproducibly observed earlier by us during the mass spectrometric analysis of several intact bacterial cells or bacterial lysates by MALDI- and LC/ESI-techniques ([4,5], T. Krishnamurthy unpublished results). The experiments were conducted with various pathogens and non-pathogens of *Bacillus, Brucella*, *Francisella* and *Yersinia pestis* species. Even though, same proteins were consistently detected during MALDI-MS, they deferred from the ones observed during the LC/ESI-MS method [4,5]. However, several distinct marker proteins were consistently observed for specific species and genus of the analyzed bacteria during each of the methods. Pathogenic organisms of a particular genus could easily be distinguished from the corresponding non-pathogenic organisms [4,5]. During MALDI- and ESI-MS analysis of few strains of *B. anthracis*, *Brucella melitensis* and *Y. pestis*, one or two distinct strain specific proteins were also observed ([4,5], T. Krishnamurthy unpublished results). However, the strain specific proteins were not consistently observed for the same strain of the sample from different preparations during any one of these methods. Even though the species identification was quite possible during the MALDI- and ESI-mass spectrometric analysis ([3–10,12,13], T. Krishnamurthy unpublished results) of several bacterial samples containing one to four different organisms, strain distinction in especially in multi-component bacterial mixtures was quite challenging.

However, the MS/MS methods utilizing either MALDI- or ESI-ionization provide the sequences of the proteins instead of only their corresponding molecular masses [14–23]. Hence, MS/MS approach could resolve the problem of identification of bacterial species and strains in mixtures and/or impure samples. Characterization of two exosporium large mass glycoproteins from *B. anthracis*, by SDS-PAGE and tandem mass spectrometry, has been attributed to strain distinction in *B. anthracis* [12]. This is insufficient to distinguish several known *B. anthracis* strains.

Hence, we selected the MudPIT methodology [21,22] for our investigation leading to distinction of *anthracis* strains, since it combines efficient sample preparation, bi-level separation, proficient ionization and automated procedures for MS/MS analysis, data processing including database searches and reporting of results. During this procedure, the presence of each protein was established from the complete sequence of at least three peptide fragments derived from the intact protein [21,22]. MudPIT methodology has also been demonstrated to be rugged and the database searches, using TIGR protein database for *B. anthracis,* resulted in accurate identification of marker proteins for the organism [23]. The approach has been widely applied in biological research including proteomics investigations. In addition, the new software "Profiler", developed by us, has been demonstrated to be applicable in the distinction of species and strain specific marker proteins from the MudPIT results, database development and searches during our investigations. During our present studies, strains of a vital human pathogen, *B. anthracis*, have been studied in detail using the sequences of their marker proteins. All marker proteins were derived, for establishing the in-house database, from the database search using TIGR protein database for *B. anthracis*. As a result of our present research involving *B. anthracis* strains, the MudPIT procedures along with "Profiler" have been demonstrated to be a powerful fully automated approach for the clear distinction of these closely related strains. Since more than 50 marker proteins were observed for individual strains based on their sequence during this MudPIT investigation, the ability to distinguish the specific strains in unknowns is very high. In addition, larger proteins including some hypothetical ones are identified in comparison with the small acid soluble proteins observed during the ESI- and MALDI-MS/MS analysis of *bacillus* species [14–17]. The approach can also be applied for the identification of other bacteria, using protein database generated for the corresponding organism. The in-house database established by Profiler contains only the molecular masses of the marker proteins. However, the identities of these proteins were established based on their sequences during the MudPIT analysis of the standards and samples. Hence, the identification of the organism as a result of the Profiler analysis was based on the molecular masses and the sequences of the corresponding marker proteins. Since numerous large mass strain specific proteins have been observed during our experiments, the methodology has enormous potential for applications in biological warfare and bioterrorism fields. In addition, the proteins are identified during the MudPIT and other MS/MS investigations based on their sequences, and hence the post-translational modification of the individual proteins will have no effect on the identification of specific marker proteins.

4. Conclusions

The software, Profiler, can analyze MudPIT data obtained from bacteria to identify the strain specific proteins and generate and save databases. *B. anthracis* strains in unknown samples can be determined utilizing MudPIT procedures, Profiler and strain specific databases. Profiler with slight modification can also be applied in identifying other bacteria and their corresponding strains.

References

[1] T.V. Inglesby, D.H. Henderson, J.G. Bartlett, JAMA 281 (1999) 1735.
[2] D.B. Jernigan, P.L. Raghunathan, B.P. Bell, R. Brechner, E.A. Bresnitz, J.C. Butler, M. Cetron, M. Cohen, T. Doyle, M. Fischer, C. Greene, K.S. Griffith, J. Guarner, J.L. Hadler, J.A. Hayslett, R. Meyer, L.R. Petersen,

M. Phillips, R. Pinner, T. Popovic, C.P. Quinn, J. Reefhuis, D. Reissman, N. Rosenstein, A. Schuchat, W.J. Shieh, L. Siegal, D.L. Swerdlow, F.C. Tenover, M. Traeger, J.W. Ward, I. Weisfuse, S. Wiersma, K. Yeskey, S. Zaki, D.A. Ashford, B.A. Perkins, S. Ostroff, J. Hughes, D. Fleming, J.P. Koplan, J.L. Gerberding, Emerg. Infect. Dis. 8 (2002) 1019–1028.

[3] C. Fenselau (Ed.), Mass Spectrometry for the Characterization of Microorganisms, vol. 541, ACS Symposium Series, American Chemical Society, 1994.

[4] T. Krishnamurthy, P.L. Ross, U. Rajamani, Rapid Commun. Mass Spectrom. 10 (1996) 883.

[5] T. Krishnamurthy, P.L. Ross, Rapid Commun. Mass Spectrom. 10 (1996) 1992.

[6] T. Krishnamurthy, M.T. Davis, T.D. Lee, Rapid Commun. Mass Spectrom. 13 (1999) 39.

[7] M.A. Meetani, K.J. Voorhees, J. Am. Soc. Mass Spectrom. 16 (2005) 1422.

[8] T. Krishnamurthy, J. Hewel, N. Bonzagni, J. Dabbs, R.L. Bull, J.R. Yates III, Rapid Commun. Mass Spectrom. 20 (2006) 2053.

[9] J. Whiteaker, J. Karns, C. Fenselau, M.L. Perdue, Foodborne Pathogens Dis. 1 (2004) 185.

[10] R.D. English, B. Warscheid, C. Fenselau, R.J. Cotter, Anal. Chem. 75 (2003) 6886.

[11] S. Swatkoski, S.C. Russell, N. Edwards, C. Fenselau, Anal. Chem. 78 (2006) 181.

[12] M.J. Stump, G. Black, A. fox, K.F. Fox, C.E. Turick, M. Matthews, J. Sep. Sci. 28 (2005) 1642.

[13] L.N. Waller, M.J. Stump, K.F. Fox, W.M. Harley, A. Fox, G.C. Stewart, M. Shahgholi, J. Bacteriol. 187 (2005) 4592.

[14] P.A. Pribil, E. Patton, G. Black, V. Doroshenko, C. Fenselau, J. Mass Spectrom. 40 (2005) 464.

[15] K.Y. Lee, E.Y. Kang, S. Park, S.K. Ahn, K.H. Yoo, Y. Jin, H.H. Lee, Proteomics 6 (2006) 1512.

[16] A. Fox, Abstracts of Papers, Computer optical disk for 231st ACS National Meeting, American Chemical Society, Atlanta, GA.

[17] B. Warscheid, C. Fenselau, Proteomics 4 (2004) 2877.

[18] J.P. Dworzanski, A.P. Snyder, R. Chen, H. Zhang, D. Wishart, L. Li, Anal. Chem. 76 (2004) 2355.

[19] N.C. VerBerkmoes, W.J. Hervey, M. Shah, M. Land, L. Hauser, F.W. Larimer, G.J. Van Berkel, D.E. Goeringer, Anal. Chem. 77 (2005) 923.

[20] A.W. Francis, C.E. Ruggiero, A.T. Koppisch, J. Dong, J. Song, T. Brettin, S. Iyer, Biochim. Biophys. Acta, Proteins Proteomics 1748 (2005) 191.

[21] M.P. Washburn, D. Wolters, J.R. Yates III, Nat. Biotechnol. 19 (2001) 242.

[22] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvick, J.R. Yates III, Nat. Biotechnol. 17 (1999) 676.

[23] H. Liu, N.H. Bergman, B. Thomason, S. Shallom, A. Hazen, J. Crossno, D.A. Rasko, J. Ravel, T.D. Read, S.N. Peterson, J.R. Yates III, P.C. Hanna, J. Bacteriol. 186 (2004) 164.

[24] R. Sodygov, J.R. Yates III, Anal. Chem. 75 (2003) 3792.

[25] D. Tab, W. McDonald, J.R. Yates III, J. Proteome Res. 1 (2002) 21.