

OPEN

Elucidation of protein biomarkers for verification of selected biological warfare agents using tandem mass spectrometry

Sakshi Rajoria¹, Sasikumar Sabna¹, Prabhakar Babele¹, Ravi Bhushan Kumar¹,
Dev Vrat Kamboj¹, Subodh Kumar² & Syed Imteyaz Alam^{1*}

Some pathogens and toxins have the potential to be used as weapons of mass destruction and instigate population-based fear. Efforts to mitigate biothreat require development of efficient countermeasures which in turn relies on fast and accurate methods to detect the biological agents in a range of complex matrices including environmental and clinical samples. We report here an mass spectrometry (MS) based methodology, employing both targeted and shot-gun approaches for the verification of biological agents from the environmental samples. Our shot-gun methodology relied on tandem MS analysis of abundant peptides from the spiked samples, whereas, the targeted method was based on an extensive elucidation of marker proteins and unique peptides resulting in the generation of an inclusion list of masses reflecting relevant peptides for the unambiguous identification of nine bacterial species [listed as priority agents of bioterrorism by Centre for Disease Control and Prevention (CDC)] belonging to phylogenetically diverse genera. The marker peptides were elucidated by extensive literature mining, *in silico* analysis, and tandem MS (MS/MS) analysis of abundant proteins of the cultivated bacterial species in our laboratory. A combination of shot-gun MS/MS analysis and the targeted search using a panel of unique peptides is likely to provide unambiguous verification of biological agents at sub-species level, even with limited fractionation of crude protein extracts from environmental samples. The comprehensive list of peptides reflected in the inclusion list, makes a valuable resource for the multiplex analysis of select biothreat agents and further development of targeted MS/MS assays.

The realization that that some pathogenic microbes and toxins of biological origin can be used as weapons of mass destruction has gained prominence especially in the backdrop of growing concerns for bioterrorism. Biological select agents are derived from biological sources, are capable of causing significant damage to human health and safety, and can be overtly or covertly used for ideological, political, or financial gain¹. U.S. Department of Health and Human Services (HHS) and the U.S. Department of Agriculture (USDA) have prioritized these select agents² and CDC has classified the agents of bioterrorism under the categories A through C³. It includes several toxins, bacteria, and viruses that are prioritized based on their high morbidity and mortality, dose of infections, environmental stability, production ease, and lack of an established countermeasure. Our vulnerability to biological warfare agents was realized after the anthrax letter attack of 2001⁴; the mortality rate of inhalational anthrax is 100% when left untreated. The high-consequence bacterial biowarfare (BW) agents encompass wide phylogenetic lineages including the causative agents of plague (*Yersinia pestis*), anthrax (*Bacillus anthracis*), brucellosis (*Brucella* spp), and tularaemia (*Francisella tularensis*). Missile warheads, manned or unmanned spray-tanks coupled to the aircraft, and aircraft capable of delivering biological agents are some of the possible means of delivery of these biological agents for clandestine applications intended to damage human, animal, and plant resources. Aerosol is the most likely and destructive dissemination mode in this regard⁵.

Efforts to mitigate biothreat require development of efficient countermeasures including effective treatment and decontamination. The medical countermeasures in such a scenario will be heavily dependent on rapid methods to detect the biological agents in a range of complex matrices including environmental and clinical samples. In this regard, complexity of matrices, such as environmental soil samples, is a major confounding element

¹Biotechnology Division, Defence Research & Development Establishment, 474002, Gwalior, India. ²Microbiology Division, Defence Research & Development Establishment, Gwalior, India. *email: syimteyaz@gmail.com

resulting in prohibitively significant background signals and false-positive/negative results^{6,7}. In the clinical context and with additional concerns (e.g. resistance to antibiotics, virulence, and emerging pathogens), early detection of the agent and disease diagnosis greatly helps prognosis and selection of a potential therapeutic measure^{8,9}.

Various analytical methods have been used in the recent past to identify agents of bioterrorism; including nucleic acid based assays, such as polymerase chain reaction and protein based immunological assays, for example enzyme-linked immunosorbent assays, the latter applicable for both protein toxins and microorganisms^{10–17}. Although, of immense value for the rapid screening of samples, these methods suffer limitations pertaining to depth of analysis, suitability for unknown and advanced biological agents, and lack of specificity (for example due to cross-reactivity with similar molecules), posing a risk of the false positive and ambiguous results. Recent developments in the MS technology promise un-equivocal identification of microorganisms with remarkable specificity, speed, and sensitivity. The multiplexing capability and suitability for analyzing complex samples such as air, water, culture medium, bodily fluids, and food are of immense value for the utility of MS based techniques in biothreat scenario. MS platforms such as Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF-MS) or liquid chromatography tandem mass spectrometry (LC-MS/MS) are gaining popularity in microbiological research, largely for diagnostic applications, using mass fingerprinting or peptide sequencing approaches¹⁸.

MALDI-TOF MS is emerging as a rapid, robust, and cost-effective technology for a routine bacterial identification method, largely in clinical diagnostic laboratories^{19–22}. For example, MALDI Biotyper and VITEK MS have been approved by the U.S. Food and Drug Administration for bacterial identification in clinical diagnostic laboratories^{23–26}. Both these MALDI-TOF-MS-based platforms ionize intact proteins extracted from whole cell culture without specific protease treatment^{19,27}. However, this approach not only requires pure cultures to produce stable and consistent result^{23,24}, culture conditions may still affect the quality of the spectra and the final identification²⁵. Moreover, most of the MS based detection methodologies reported till now, rely on an in-house database and analysis of MS spectra of intact proteins with very limited or no peptide sequence information; the strength of MS/MS analysis that provides sequence information at the highest resolution is missing in these approaches. Utility of MALDI-TOF-MS method for the species level identification of high-consequence bacterial pathogens was found to be challenging due to mass spectral profile database content and quality²⁸. The single MS approaches also suffer serious limitations for analysis of complex microbial cultures, detection of microorganisms in food and environmental samples, in situations where the organism is non-cultivable or fastidious, and for identifying rare pathogenic bacteria or novel-emerging pathogens²⁹. MS/MS is capable of generating sequence information and therefore can significantly improve the specificity for the identification of biological agents based on the detection of unique tryptic peptides^{30–36}. For example, a combined bottom-up and top-down proteomics methodology was used, employing a high-resolution/high mass accuracy LTQ-Orbitrap instrument to identify specific markers of *B. anthracis* spores, capable of discriminating closely related *Bacillus cereus* biovar anthracis CI, CA, and *Bacillus thuringiensis* 9727³⁷. Proteotypic peptides are classically defined as “the peptides that uniquely identify each protein and are consistently observed when a sample mixture is interrogated by a (tandem) mass spectrometer”³⁸. There are a few software applications available for the *in silico* digestion and predictive elucidation of proteotypic peptides for the targeted proteomic analysis^{39–42}. For example, Skyline software aids selection of proteotypic peptides for the MRM assay development for the quantitative analysis of target proteins (<https://sciex.com/products/software/skyline-software>). Similarly, PeptideManager is a tool for the selection of peptide for the targeted proteomic analysis for mixed samples originating from different species⁴⁰. However, these methodologies select peptides for proteins from a target proteome (e.g., human) and ensure specificity against a background of host proteome (e.g., mouse, rat)⁴³. Mesuere *et al.*⁴³ have recently reported a web application, The Unique Peptide Finder (<http://unipept.ugent.be/peptidefinder>), to screen for taxon-specific tryptic peptides.

Though beginning to be recognized, the reports on application of MS-based methodology in biothreat scenario are scanty: among the first of the applications, Dworzanski *et al.*⁴⁴ demonstrated discrimination and classification of the closely related strains of *B. anthracis*, *B. cereus*, and *B. thuringiensis*. Using tandem MS approach, we have previously reported recovery and analysis of biological agent after bioaerosol exposure in a simulated scenario of biothreat⁶. Using a poly-disperse disseminator in the experimental set-up, the importance of the choice of matrix was highlighted for a post-attack identification after recovery of the biological agent. We recommended to use matrices with minimal interfering substances (e.g. foliage, smooth surfaces, sand) for verification of the bio-agent in a biothreat scenario, retrospectively. Complex environmental matrices (e.g. soil) are challenging for the said purpose as the high background signal hinders assays and lead to false-positive/negative results⁷. We report here a comprehensive elucidation of marker proteins and species specific unique peptides for the targeted tandem MS analysis of bacterial select agents. The targeted approach described here is distinct from the well-known multiple reaction monitoring (MRM) assays and involves selection of several unique peptide masses for the tandem MS analysis after the first MS scan. The MS based methodology employing both targeted and shot-gun approaches for the verification of biological agents from the environmental samples is demonstrated here. The targeted approach is intended to screen for the bacterial agent from a pre-defined list of pathogens using selected unique marker peptides, whereas, the shot-gun methodology aims at generating sequence information for as many peptides as possible and to identify even the unknown or advanced biothreat agents. Our shot-gun methodology relied on tandem MS analysis of abundant peptides from the spiked samples, whereas, the targeted method was based on an extensive elucidation of marker proteins, unique peptides, strain coverage analysis, and generation of an inclusion list of masses reflecting relevant peptides for the unambiguous identification of selected bacterial biowarfare agents. The comprehensive list of peptides reflected in the inclusion list, makes a valuable resource for the multiplex analysis of select biothreat agents and further development of targeted tandem MS assays. The species-specific marker peptides for the nine bacterial select agents were elucidated by extensive literature mining, *in silico* analysis, and MS/MS analysis of abundant proteins from axenically grown bacterial

species in our laboratory. As a proof of concept, the methodology was validated using blind samples spiked with one or two select agents spiked in soil and sand.

Results

Identification of abundant proteins from select bacterial agents. Abundant proteins from *B. anthracis* Sterne strain, *Clostridium perfringens* ATCC13124, *Burkholderia pseudomallei*, *Brucella abortus* strain NCTC 10093, and *Clostridium tetani* strain drde were identified by MS/MS analysis of whole cell lysate. Cells were asexually grown on rich medium and total protein was extracted as described in the methods. The proteins were separated on SDS-PAGE and abundant protein bands (20–40) were excised from each lane. After tryptic digestion of excised bands, proteins were identified by MALDI-TOF-TOF analysis and subsequent MASCOT search against the taxonomic group of bacteria as described in the methods sections. Results obtained from two such independent analysis, were combined together and the proteins identified in the corresponding species with significant MS/MS ion score of peptides are listed in Supplementary Table S1 A–F. On several occasions, more than one proteins from the same species were identified from a single spot, as there could be multiple co-migrating proteins on single dimension of SDS-PAGE. Only for *C. botulinum*, MS/MS data for different botulinum neurotoxin (BoNT) serotypes were compiled from reported literature. This way, 60 non-redundant proteins from *B. anthracis* Sterne strain (184 peptides), 30 from *B. abortus* (176 peptides), 12 from *B. pseudomallei* (73 peptides), 48 *C. perfringens* (211 peptides), and 59 proteins from *C. tetani* (227 peptides), were identified with significant MASCOT score. The peptides identified with significant MS/MS ion score from crude lysate of cells with minimal fractionation indicated their abundance and amenability for MS analysis, in turn their suitability as markers. All peptides observed here were subjected to global protein BLAST against the GeneBank non-redundant protein database with search parameters adjusted for short input sequence and those showing less than 100% sequence identity with any species other than the target species were noted. Representative MS and MS/MS spectra for three proteins have been shown in Supplementary Fig. S1.

Elucidation of putative protein markers for detection. The following parameters were considered for qualifying a protein as putative marker for a pathogenic species: (1) an evidence of expression indicating abundance, supported by the number of scientific reports; (2) identification of the protein in crude cell lysate using MALD-MS/MS analysis with significant ion score for peptides; and (3) a low percent identity of the protein with its closest homolog in any other bacterial species. From the bacterial select agents listed by HHS and USDA as priority agents of bioterrorism, nine bacterial species belonging to diverse genera were selected for the analysis. This included *Brucella suis* [BSU], *B. abortus* [BAB], *Brucella melitensis* [BMT], *B. pseudomallei* [BP], *Burkholderia mallei* [BM], *C. perfringens* [CP], *Clostridium botulinum* [CB], *C. tetani* [CT], and *B. anthracis* [BA].

Evidences for the abundance and expression of different proteins in a given species were screened by extensive literature mining at PUBMED (www.ncbi.nlm.nih.gov) using set of appropriate keywords, one after the other. Any protein for a given species reported in at least three independent publications was selected and its attributes with respect to immunogenicity, immuno-dominance in naturally infected clinical sera, immuno-protection, and role in virulence were noted. The short-listed proteins were subjected to global protein BLAST against the non-redundant protein database at National Center for Biotechnology Information (NCBI) using the FASTA sequences for each of the proteins retrieved from the reference strain of the selected bacterial species (Supplementary Table S2 A–F). Localization of the protein was noted from the experimental reports during literature survey and also predicted by PSORT algorithm at ExPASy Proteomics tools (<http://www.expasy.ch>).

As described in the methods, we used our own ranking scheme to prioritize candidate protein markers for each of the selected species by taking into consideration the three parameters mentioned above: the number of reports for the expression of a given protein and the number of peptides with significant MS/MS ion score in the experimental identification of abundant proteins in the laboratory were assigned 25 points each, whereas the percent identity with the nearest homolog was given 50 points. Low percent identity with the nearest homolog in any other species was given maximum weightage as it entails a better chance of scoring unique peptides for the given species for the targeted MS based assay development. For instance, a sequence showing 100% identity with the nearest homolog in any other species was assigned 0 point (100–100), while 50 points for the identity parameter was given to a sequence with 50% identity (100–50). Abundant proteins identified from the whole cell lysate of selected bacterial species were assigned 25 points: 5 points for every peptide with significant MS/MS ion score, 1 peptide was given 5 points and 5 additional points were given to every additional peptide up to a maximum of 5 peptides (≥ 5 peptides given complete 25 points). These peptides from the tryptic digest of minimally fractionated crude lysates indicated their abundance and amenability for tandem MS analysis, further augmenting their potential as markers for unambiguous detection. The number of times a given protein was reported for a particular species as expressed, in independent studies, was considered a proof of its abundance and putative role in virulence was given priority to improve specificity to discriminate closely related species or to differentiate between pathogenic and saprophytic strains of the species. We assigned 5 points for 3 reports and 5 additional points for every additional report up to 7 (≥ 7 reports for the given protein assigned complete 25 points). This priority protein list was appended with pathogen specific virulence associated proteins from the reported literature, if not selected at this stage. Supplementary Table S2 A–F, lists 15–27 putative proteins markers for each of the selected bacterial species in the decreasing order of the weightage assigned according to the scheme described above. At this stage, the closely related species of *Burkholderia* (*B. pseudomallei* and *B. mallei*) and *Brucella* (*B. melitensis*, *B. suis*, and *B. abortus*) were considered together. Supplementary Table S1 (A–F) also reflects: maximum number of times any peptide of the protein was observed during MS/MS analysis in the present investigation using crude whole cell lysate of the organism; the number of reports for expression; predicted localization (PSORT); percent identity with the nearest homolog; and its role with respect to virulence, immunogenicity, vaccine potential and experimentally determined localization. The number of proteins shortlisted at each of the steps mentioned above

	BMT	BM	BA	CP	CT	CB
Literature mining for evidence of expression	263	524	566	184	56	105
Proteins with ≥ 3 reports of expression	67	24	48	31	6	51
Proteins identified in whole cell lysate (MS/MS)	30	12	60	48	58	6
Overlapping proteins (wet-lab/data mining)	10	7	26	24	6	6
Appended with virulence associated proteins	10	10	10	10	10	10
Priority proteins for targeted analysis	16	16	15	21	15	27

Table 1. Elucidation of protein markers for unambiguous detection of selected bacterial biowarfare agents. BMT represents all three closely related species of *Brucella* (*B. melitensis*, *B. abortus*, and *B. suis*) recognized as agents of bioterrorism by CDC, Atlanta. BM = *B. mallei* and *B. pseudomallei*; BA = *B. anthracis*; CP = *C. perfringens*; CT = *C. tetani*; CB = *C. botulinum*.

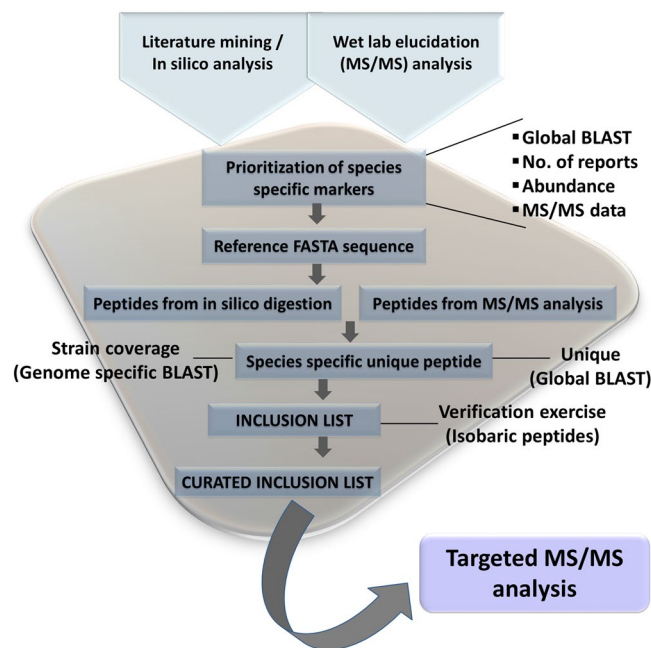


Figure 1. Schematic representation of the workflow for elucidation of marker peptides for targeted analysis of selected bacterial biowarfare agents.

for the selected taxa is summarized in Table 1 and the schematic workflow is reflected in Fig. 1. The predicted molecular function/biological process for the putative marker proteins for the selected bacterial agents indicated a predominance of proteins involved in: post-translational modification, protein turnover and chaperone function; metabolism; and those related to virulence (Fig. 2). Supplementary Table S3 summarizes the selected putative marker proteins for specific detection of target bacterial agents of BW significance with brief molecular function/biological process, predicted localization, and other functional attributes experimentally demonstrated by different groups. These putative markers (19 for *B. anthracis*, 20 for the three closely related species of *Brucella*, 16 for *B. pseudomallei* and *B. mallei*, 27 for *C. botulinum*, and 21 for *C. perfringens*, and 15 for *C. tetani*) are arguably expressed by the selected pathogenic species and are distantly related to their homologs in other bacterial species. The details of the *in silico* methods for selection of the marker proteins is presented in Supplementary Material S6.

Elucidation of unique peptides and generation of inclusion list. The FASTA sequences of the selected putative marker proteins were subjected to *in silico* tryptic digestion and the peptides in the mass range of 1000–3000 Da were subjected to global protein BLAST against the GeneBank non-redundant protein database with search parameters adjusted for short input sequence. Peptides showing less than 100% sequence identity with any other species, indicating a difference of at least one amino acid residue and as consequence unique precursor mass, were listed as unique. Apart from the peptides selected as unique for the target species, peptides were also selected that were specific to the very closely related species within the *B. cereus* sensu lato group, including eleven other species (*B. cereus*, *Bacillus mycoides*, *Bacillus pseudomycooides*, *B. thuringiensis*, *Bacillus weihenstephanensis*, *Bacillus cytotoxicus*, *Bacillus wiedmannii*, *Bacillus toyonensis*, *Bacillus bingmayongensis*, *Bacillus gaemokensis*, and *Bacillus manliponensis*) with a tolerance for one or two species. This way, the final inclusion list for *B. anthracis* included those unique to the species, specific to the *B. cereus* sensu lato group without or with one or two exception, and those abundant in the species to aid a comprehensive identification. Only a few peptides could be scored specific for the closely related species of *B. mallei* and *B. pseudomallei*; the two species were

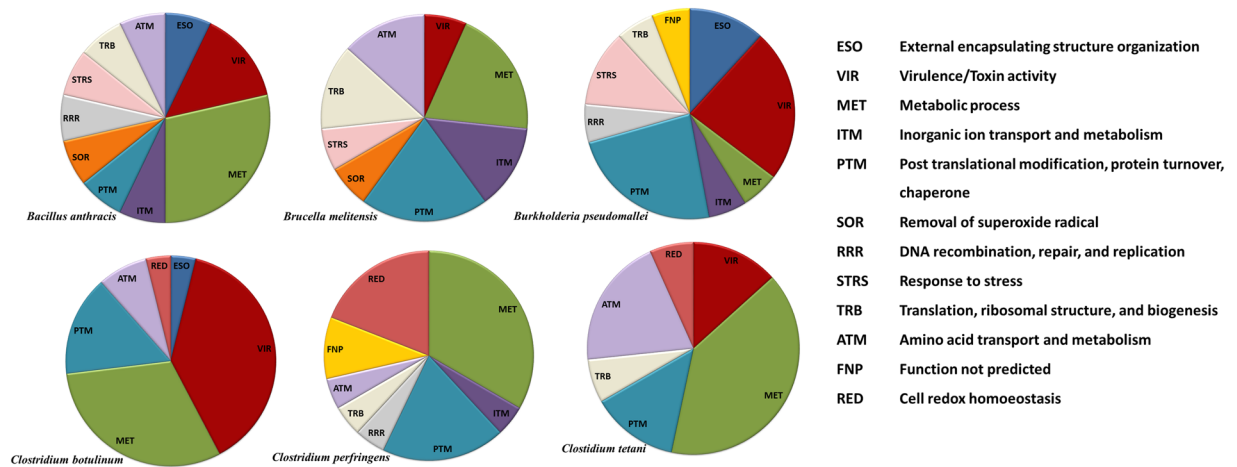


Figure 2. Functional classification of proteins short-listed as putative marker for targeted detection of selected bacterial biowarfare agents.

considered together for the peptide BLAST search and scoring unique peptides. In order to aid a comprehensive identification, the final inclusion list for *B. mallei* and *B. pseudomallei* included those unique to the species, specific to both the target species without or with one or two exception, and those abundant in the species.

All the peptides selected for the three closely related species of *Brucella* (*B. melitensis*, *B. abortus*, and *B. suis*) were further screened for species specific peptides after downloading all the available sequences of putative marker proteins using species specific genome BLAST (keeping specific species as target organism). The FASTA sequences were sorted into three groups, corresponding to each of the taxa (*B. melitensis*, *B. abortus*, and *B. suis*). This resulted in the list of peptides for the three species of *Brucella* including those unique to one, two or all the three species (*B. melitensis*, *B. abortus*, and *B. suis*, as indicated in superscript as M, A, and S) without or with one or two exception, and those abundant in the species to aid a comprehensive identification (Supplementary Table S4 A–G).

In order to avoid false positives for the intended analysis of environmental samples, we deduced isobaric peptides obtained from tryptic digest of bacterial consortium on MS scan using the inclusion list for all the peptides for the selected agents (0.5 Da window) and MS/MS analysis of selected precursors. The curated list of marker peptides and their corresponding masses are shown in Supplementary Table S4 A–G. The unique peptides in the inclusion list were further screened for strain coverage using all the available sequences retrieved by species specific BLAST search and counting the number of strains from the alignment data for the conservation pattern of the selected peptides in diverse strains in the genome data base (Supplementary Table S4 A–G). The details of the in silico methods used for the elucidation of species specific peptides is described in Supplementary Material S6.

Proof-of-concept studies. The shot gun and targeted search methodologies using the inclusion list of peptides described here were used in the proof-of-concept experiment using number coded blind samples. The results of a representative analysis using eight number-coded blind soil and sand samples spiked with either or both of *B. anthracis* Sterne strain and *C. perfringens* ATCC13124 is shown in Supplementary Table S5. Garden soil and sand were either blank (not spiked with any agent) or spiked with cell suspension [10^6 *C. perfringens* ATCC13124 or 10^7 *B. anthracis* Sterne colony forming units per gram (cfu.g⁻¹) of matrix] of one or two bacterial agents and dried overnight at 37 °C in order to simulate field conditions under which the agents are likely to bind to the matrix particles as a result of desiccation or other environmental factors. When both the agents were spiked, *C. perfringens* was added at a concentration one log lower (10^6 cfu.g⁻¹) than *B. anthracis*. The analyst was provided the number coded blind samples to carry out the identification of a pathogenic bacterial species. Total protein was extracted as described in the methods and subjected to minimal resolution on SDS-PAGE. Ten abundant protein bands from each sample (except for the two blank samples) were excised from the SDS-PAGE gel (Supplementary Fig. S2). The gel pieces were subjected to tryptic digestion before loading onto MALDI target plates for tandem MS analysis by MALDI-TOF-TOF using both shot-gun and targeted acquisition from replicate spots. Agents were correctly identified in all the samples, except for the mixture of two bacterial agents where *C. perfringens* (at one log lower concentration), was missed, perhaps due to masking of the proteins on 1D gel and for only a limited fraction of proteins (10 bands) being analyzed (Supplementary Table S5). For *B. anthracis*, other species within the *B. cereus* sensu lato group also exhibited significant match in the MASCOT search. However, the decision regarding *B. anthracis* could easily be taken for the number of places it ranked at the top of the matched species. Limited number of peptides from the inclusion list was seen in the targeted approach due to the limited number of protein bands sampled. The number of peptides matched (MP) to the identified protein during MASCOT search using MS data and the MASCOT scores are also shown in the Supplementary Table S5. MASCOT ion score for the MS/MS analysis of the fragmentation data of peptides showing significant ion score and the RMS error indicate ability of the proposed methodology to correctly identify the biological agent even with the limited sampling of the protein bands from the extracted proteins. MS and MS/MS spectra of the tryptic digest of three representative proteins from this study are shown in Supplementary Fig. S4.

Discussion

Mitigation of bio-threat requires rapid and accurate methods to detect the biological agents in a range of complex environmental and clinical matrices. The study reported here describes a comprehensive MS based methodology employing both targeted and shot-gun approaches for the verification of biological agents from the environmental samples in a biothreat scenario. The shot gun approach aims at generating sequence information for as many peptides as possible and attempting to identify even the unknown or advanced pathogens, taking advantage of the ever expanding protein database and search engines for MS and MS/MS data interrogation. Our shot-gun methodology relies on tandem MS analysis of abundant peptides from the spiked samples, whereas, the targeted method is based on an extensive elucidation of marker proteins and unique peptides. This resulted in the generation of an inclusion list of masses reflecting relevant peptides for the unambiguous identification of bacterial select agents belonging to diverse genera including nine bacterial species. The comprehensive list of peptides reflected in the inclusion list (Supplementary Table S4), makes a valuable resource for the multiplex analysis of select biothreat agents. This marker peptide list was elucidated by extensive literature mining, *in silico* analysis, and MS/MS analysis of abundant proteins of the cultivated select bacterial species in our laboratory. The methodology was validated using blind soil and sand samples spiked with one or two select agents.

MALDI-TOF MS has been used for the detection of microorganisms and toxins in environmental and food samples using both bottom-up and top down approaches after cultivation of microorganisms for enrichment and separation^{45–47}. Shotgun proteomics has shown promising results in describing a large panel of virulence factors in several pathogens including the analysis of exoproteome from *B. cereus*⁴⁸. Lasch *et al.*⁴⁹ analyzed cultivated strains of *Bacillus* and related genera including those of *B. anthracis* and *B. cereus* by MALDI-TOF MS revealing mass spectral peaks specifically belonging to *B. anthracis* species. A rapid MALDI-TOF MS method was developed for identification of *B. anthracis* spores and its differentiation from closely related *B. cereus* in powder samples, with a limit of detection of 2.5×10^6 spores⁴⁶. Ten distinct pathogenic bacterial samples including *B. anthracis* and *Y. pestis* were used in an inter-laboratory ring trial from nine European countries: accurate microbial identification relied on the comprehensive spectral databases used⁵⁰. There are reports where intact protein biomarkers have been defined by proteogenomics in order to differentiate microorganisms and discriminate pathogens of a given genus or clade^{51–53}. However, the single MALDI-MS approach suffers several limitations due to lack of peptide sequence information, need of pure cultures to produce consistent results, and effect of culture conditions on spectral quality and ensuing identifications^{23–25}. The inability to differentiate high-consequence pathogens from related bacterial species can be a serious issue for clinical diagnostic laboratories and more so in a biothreat perspective^{28,50,54}. Incorrect identifications with single MALDI-TOF MS methods have resulted in laboratory exposures with *Francisella tularensis*⁵⁵ and *B. pseudomallei*⁵⁶.

We describe here the use of tandem mass spectrometry (MALDI-TOF-TOF) to seek sequence information for multiple target proteins for the detection of bacterial pathogens of BW significance from diverse phylogenetic clades. The putative marker proteins were judiciously selected to impart selectivity by employing a ranking scheme to prioritize candidate protein markers for each of the selected species by taking into consideration, the three parameters described in the results. Extensive literature mining provided strong evidence of expression indicating abundance and role in virulence at the experimental level as indicated by the number of scientific reports (Supplementary Table S2 A–F). The identification of the abundant proteins, from the selected bacterial species using MALDI-MS/MS analysis in the minimally fractionated crude cell lysate, with significant ion score for peptides indicated abundance and amenability of the peptides for the tandem MS analysis (Supplementary Table S1 A–F). A low percent identity of the protein with its closest homolog in any other bacterial species was given maximum weightage for an increased likelihood of getting species specific unique peptides (Supplementary Table S5). Taking advantage of the sequence information and to improve the specificity, bottom-up analysis using MALDI-TOF(/TOF) MS and electrospray ionization ESI-MS(/MS) methods have been reported for the identification of toxins and microorganisms, based on the detection of unique tryptic peptides^{30–37}. Tandem MS approaches, providing peptide amino acid sequence information, are more often obtained with LC-MS/MS systems for crude proteins: the tryptic digest of whole proteome is separated by HPLC and the ionized peptides are detected and further fragmented to obtain peptide sequences by comparing the tandem spectra to the theoretical peptide fragmentation spectra. For example, a combined bottom-up and top-down proteomics methodology was used, employing a HR/MS LTQ-Orbitrap instrument, to elucidate specific markers of *B. anthracis* spores capable of discriminating closely related *B. cereus* biovar anthracis CI, CA, and *B. thuringiensis* 9727³⁷. Employing a statistical scoring algorithm against a fully sequenced database of 170 bacterial genomes, identification of bacteria has been reported using tandem MS data obtained with bacterial proteins extracts digested with trypsin⁵⁷. Although restricted to the 170 sequenced organisms, the method was shown to be reliable in strain level identification of bacteria when enough peptide information was collected.

Our study represents perhaps the broadest and most comprehensive list of unique peptides for the targeted verification of bacterial select agents including *B. suis*, *B. abortus*, *B. melitensis*, *B. pseudomallei*, *B. mallei*, *C. perfringens*, *C. botulinum*, *C. tetani*, and *B. anthracis*. Using *C. botulinum* as a model, Al-Shahib *et al.*⁵⁸ reported biomarker discovery and validation using mass spectrometry and bioinformatics. However, the validation studies were carried out on the peptides unique to *C. botulinum* strains by searching with phylogenetically related *Clostridium* species and not using a global BLAST analysis. Although, there are a few software applications recently developed for the *in silico* digestion and elucidation of proteotypic peptides for the targeted MS based analysis, our definition of species specific proteotypic peptide is broader in specificity and relies on global identity with all known sequences in the protein database at NCBI (www.ncbi.nlm.nih.gov)^{44–49}. This approach, employing global BLAST search for each peptide from the list of *in silico* digested putative markers is likely to provide subtle advantage in imparting specificity of identification of biowarfare agents in a possible biological emergency. Sensitivity of detection in a biothreat scenario is of lesser importance than the accuracy of agent verification as ambiguity of identification has got serious implications for the execution of medical countermeasures; selection

of a potential therapeutic and prophylactic agent relies on unambiguous agent identification. Mathematical models were used for estimating dosage and casualty that is likely to result from the release of aerosol near ground level from a single point by an aircraft. Notably, it is estimated that in an area of 100 km², several kilograms of bio-agent need to be released for a substantial effect on a population^{5,59}. In the recent past, selective and sensitive quantitation of protein candidates by targeted analysis has been accomplished by selected reaction monitoring (SRM) methods coupled to isotope dilution⁶⁰. The MRM assays on LC-MS/MS are rapid and provide multiplexing capabilities for several target proteins: ion selection and scoring for daughter ions after fragmentation (MRM transition) is used to identify and quantify molecules of interest per sample run. There are reports for the specific and sensitive detection and quantification of toxins and pathogens from environmental and food matrices, including *B. anthracis* spores and *Y. pestis*^{36,37}. Immunocapture of intact cell has often been combined with LC/SRM, to improve sensitivity^{36,46}. However, the SRM acquisition has an intrinsic limitation of selectivity due to low resolution of the mass analyzers (quadrupole) used for mass selection resulting in bias with multiplex targets. Recent developments in high-resolution accurate-mass acquisition (HR/AM) instruments have been used to overcome this limitation with improved selectivity due to enhanced capability for the separation of ions of interest from interfering moieties, especially in complex biological samples⁶¹. A multiplexed and targeted analysis 109 peptides from 27 *Staphylococcus aureus* proteins has been reported using LC-SRM approach for the species level identification and additional information pertaining to antibiotic resistance, virulence factor detection (e.g. toxins involved in toxic shock syndrome), and epidemiological typing (e.g. Protein A peptides)⁶². It is important to note that this SRM assay requires $\geq 10^7$ cfu for identification and a preliminary culture step is required.

The inclusion list peptides described here for the nine bacterial select agents, were derived from rationally selected protein markers. The tryptic peptides of these putative marker proteins were subjected to rigorous screening by global BLAST search against the huge protein data base to ensure maximum representation of unique peptides from multiple target proteins of a given species. The peptides were further screened for the strain coverage and curated to reduce background signal by deducting isobaric peptides from the environmental bacterial consortium. Although reflected together in Supplementary Tables S2 and S3, the putative marker proteins were searched independently for the closely related species of *B. pseudomallei*, *B. mallei* and for the very related three species of *Brucella* (*B. melitensis*, *B. suis* and *B. abortus*). Although, discrimination of these very closely related species is challenging, a combination of shot-gun tandem MS analysis and a targeted search using a panel of unique peptides is likely to provide unambiguous verification of biological agents even at sub-species level, even with limited fractionation of crude protein extracts from the environmental samples. However, the approach based on the species specific unique peptides should be continuously monitored and updated, as sequencing of novel species or strains whose protein sequences differ from those in public databases can potentially erode some of the signatures. As several proteins and peptides are being considered for the targeted analysis here, the signature erosion by discovery of novel taxa are less likely to grossly influence the identification. Although, only unique peptides showing less than 100% sequence identity with any other bacterial species (except the query species) were selected for the inclusion list of unique peptides; it was based on the assumption that a difference of even one single amino acid is most likely to ensure that the selected mass will not come from any other species except the target taxon. Nevertheless, a very low possibility of getting a non-specific isobaric peptide with same amino acid composition but differing in sequence does exist. The proof-of-concept study reported here shows the applicability of the method for the verification of the biological agent. Further studies are needed before the characteristics of the assay reported here are known; rigorous validation study would include many more samples and replicates, include all the agents for which the targeted assay is intended with positive and negative controls and, a limit of detection analysis, and false negative and false positive rates. A detailed validation exercise, when carried out with the state of art LC interface (e.g. two dimensional nano-LC), would likely provide deeper analysis and high resolution data. The data-independent acquisition (DIA) methods, such as SWATH-MS, are emerging technologies that enables deep proteome coverage and consistent quantitation, and accuracy⁶³. The inclusion list of peptides reported here can be of immense use in other platforms of high resolution mass spectrometry (e.g. SWATH-MS) for targeted analysis of bio-threat agents.

Conclusions

The realization that that some pathogenic microbes and toxins of biological origin can be used as weapons of mass destruction has gained prominence especially in the backdrop of growing concerns for bioterrorism. Efforts to mitigate biothreat require development of efficient countermeasures which in turn relies on fast and accurate methods to detect the biological agents in diverse complex matrices including environmental and clinical samples. We report here an MS based methodology, employing both targeted and shot-gun approaches for the verification of biological agents from the environmental samples. Our shot-gun methodology relied on tandem MS analysis of abundant peptides from the spiked samples, whereas, the targeted method was based on an extensive elucidation of marker proteins, unique peptides, strain coverage analysis and generation of an inclusion list of masses reflecting relevant peptides for the unambiguous identification of bacterial select agents belonging to the diverse taxonomic groups. The comprehensive list of peptides reflected in the inclusion list, makes a valuable resource for the multiplex analysis of select biothreat agents, and was elucidated by extensive literature mining, *in silico* analysis, and MS/MS analysis of abundant proteins of the selected cultivated bacterial species in our laboratory. A combination of shot-gun tandem MS analysis and a targeted search using a panel of unique peptides is likely to provide unambiguous verification of biological agents even at sub-species level, even with limited fractionation of crude protein extracts from environmental samples. The validation study reported here is just a proof of concept to show the applicability of the method, as deeper analysis can be carried out using the current HR-MS platforms with high end LC interface (e.g. two dimensional nano-LC). The inclusion list of peptides reported here can be of immense use in other recent platforms of high resolution mass spectrometry (e.g. SWATH-MS) for targeted analysis of bio-threat agents.

Materials and Methods

Bacterial strains and growth conditions. The following strains of bacteria were used for the elucidation of abundant proteins and their peptides by MS/MS analysis of tryptic digest of proteins in the laboratory: *B. anthracis* Sterne strain, *Clostridium perfringens* ATCC13124, *Burkholderia pseudomallei* strain B50, *Brucella abortus* strain NCTC 10093, and *Clostridium tetani* strain drde. A high containment facility (biosafety level 3) was used for the cultivation of bacterial cultures and preparation of whole cell protein lysate and the protocols were approved by institutional biosafety committee of Defence Research and Development Establishment, Gwalior, India. Non-pathogenic *B. anthracis* Sterne strain was obtained from Institute of Veterinary and Preventive Medicine, Ranipet, Vellore, India and grown at 37 °C on Brain Heart Infusion (BHI) broth containing calf brain infusion, 12.5 g; beef heart, 5 g; peptone, 10 g; NaCl, 5 g; D-glucose, 29 g; Na₂HPO₄, 2.5 g; and distilled water 1 L. *C. perfringens* ATCC13124 was obtained from Becton Dickinson India Pvt. Ltd., India and grown anaerobically in modified reinforced clostridial medium (RCM) broth containing 10 g of peptone, 10 g of beef extract, 3 g of yeast extract, 5 g of dextrose, 5 g of NaCl, 3 g of cysteine HCl, 0.5 g of sodium acetate, and 1000 ml distilled water. *C. tetani* drde strain was isolated previously from decaying fish sample at DRDE, Gwalior, India and was anaerobically grown in trypticase peptone yeast extract glucose (TPYG) broth containing pancreatic digest of casein, 50 g; peptone, 5 g; yeast extract, 20 g; glucose, 4 g; sodium thioglycollate, 1 g; NaCl, 2 g; soluble starch, 1 g; cycloserine, 250 mg; sulphamethoxazole, 76 mg and trimethoprim, 4 mg per litre. *B. pseudomallei* strain B50 is a clinical isolate retrieved from our own culture collection at DRDE, Gwalior, India and the bacterium was axenically grown in LB broth at 37 °C. *B. abortus* strain NCTC 10093 was grown in Brucella broth with 5% CO₂ at 37 °C.

Reagents and chemicals. Analytical grade chemicals were procured from Sigma-Aldrich (Sigma-Aldrich Chemicals Pvt. Ltd., New Delhi, India), unless specified otherwise. Tryptic digestion was carried out using sequencing grade modified trypsin from Promega (Promega, Madison, WI, USA). Mass standards for MALDI-TOF-TOF analysis were obtained from Applied Biosystem, USA³⁴. The culture media were procured from; Oxoid Ltd., England, or Difco Laboratories, France, or HiMedia, India.

Extraction of total cellular protein. Bacterial strains were axenically grown in rich media, till the exponential phase, for the preparation of whole cell proteome as described above. Cells were harvested by centrifugation at 8,000 × g for 30 min at 4 °C, followed by a washing step with 20 mM Tris.HCl buffer (pH 7.5). Extraction buffer (2% SDS, 100 mM DTT, 20 mM Tris-HCl pH 8.8) was added to the cell pellets and incubated for 20 min at 95 °C. After centrifugation at 20,000 × g for 10 min at 10 °C, the supernatant was collected and the protein in solution was precipitated with trichloroacetic acid (TCA) at a final concentration of 10% (w/v) in the presence of β-mercaptoethanol (0.07%). After an incubation of 2 h on ice, the protein pellet was collected by centrifugation (10,000 × g, 4 °C, 20 min), washed twice with 500 μL of acetone, and air-dried for 5–10 min, and stored at –80 °C till use³⁴. Using the method described earlier, the protein pellet was solubilised with 8 M urea and 2% w/v CHAPS and loaded onto a discontinuous denaturing SDS PAGE (12%) after an additional heating at 95 °C for 5 min with SDS lysis buffer [12.5 mM Tris-HCl (pH 6.8), 4% glycerol (w/v), 0.4% SDS (w/v), and 1% β-mercaptoethanol (v/v)]³⁴. Quick Start Bradford Protein Assay kit (Bio-Rad) was used for the determination of protein concentration as per manufacturer's instructions using BSA as standard.

For the visualisation of protein bands, gels were stained with Coomassie Brilliant Blue G-250 and gel pieces were excised from the abundant protein bands (20–40) using thin walled microfuge tubes, cut at the tip with an approximate 2 mm diameter.

In-gel protein digestion. The excised gel pieces were destained with three washes of 200 μL 50% ACN/50 mM NH₄HCO₃ for 30 min each, at room temperature. After reduction with 10 mM DTT, the proteins were alkylated using 50 mM iodoacetamide, the gel pieces were dried. Trypsin (100 ng) in 50 mM NH₄HCO₃ was added to the gel pieces and digestion was carried out overnight at 37 °C. Using 60% ACN and 0.1% TFA, the peptides were extracted and dried. MALDI-MS/MS analysis was carried out after resuspending the peptides in 0.5% TFA.

MALDI-TOF-TOF analysis of tryptic protein digests. Abundant proteins from total cellular protein fraction from five bacterial select agents were digested with trypsin and identified by Applied Biosystem 4800 plus MALDI TOF/TOF Analyzer (AB Sciex, USA). As described earlier, the trypsin digested peptides were spotted onto the 384-well MALDI target plate after mixing with equal volume of the CHCA matrix solution (10 mg/mL) and MS mass spectra were recorded in the reflector positive mode using a laser (200 Hz repetition rate with a wavelength of 355 nm) with accelerated voltage of 2 kV³⁴. A six component peptide standard (mass range of 905–3660 Da) was used for the default calibration at 13 calibration points. For the identification of abundant proteins, the MS/MS mass spectra were acquired by the data dependent acquisition method. MS/MS fragment ions were generated by collision induced dissociation (CID) at 1 kV voltage for 20 strongest precursors selected with a signal-to-noise ratio >20, between the m/Z range of 850–4000 Da. MS/MS spectra were acquired as described previously⁶⁴. CID was carried out for the fragmentation of precursor ions, sequentially selected by timed ion selector, using air as collision gas at 1 kV energy and 1.5 × 10^{–6} recharge pressure threshold. All MS and MS/MS spectra were obtained by accumulation of at least 1200 and 1600 laser shots, respectively and peak lists were generated using the 4000 Series Explorer Software v. 3.5 (Applied Biosystems). MS/MS peaks were selected on the basis of a signal-to-noise ratio greater than 10 over a mass range of 60–20 Da below the precursor mass. The data was analyzed using Protein Pilot version 4.0 (Applied Biosystem) employing the MASCOT 2.0 search engine (Matrix Science, London, UK). The peak list was searched against all entries at non-redundant protein sequence database of NCBI with 16338050 sequences. The following search parameters were used: 50 ppm precursor ion mass tolerance; ±0.6 Da fragment ion mass tolerance with +1 charge state, one missed cleavage for

trypsin digestion, and variable modifications (oxidation of methionine and carbamidomethylation of cysteine). For the successful identification of protein by MS/MS, MASCOT score greater than 62 was accepted as significant (p -value < 0.05).

For targeted search approach, precursor ions on a ± 0.5 Da window, with a signal-to-noise ratio greater than 20, were selected for MS/MS analysis from the inclusion list defined in the interpretation method of the instrument containing m/z for all unique specific peptides for the selected pathogenic bacterial species³⁴.

Bioinformatic analysis. Literature mining for the evidence of expression of proteins for a given pathogen was carried out using PUBMED search engine at NCBI (www.ncbi.nlm.nih.gov) using appropriate keywords. All the proteins experimentally shown to be expressed by the selected species (9) were initially ranked in the decreasing order of the number of reports for a given protein with a cut-off of at least three reports for each. Literature mining was also used for the attribution of functional role of expressed proteins with respect to immunogenicity, immuno-dominance in naturally infected clinical sera, immuno-protective nature, role in virulence, and surface localization. ExPASy Proteomics tools (<http://www.expasy.ch>) were also used for the prediction of molecular function/biological process and protein localization sites in cells (PSORT).

Candidate protein markers for each of the selected species were prioritized according to an arbitrarily designed scoring scheme that was aimed at selecting proteins/peptides unique to the species and expressed by the microorganism in abundance, in turn increasing the probability of getting better signal on MS analysis. We have taken three parameters for the prioritization of the proteins: the number of reports for the expression of a given protein (25 points), number of peptides with significant MS/MS ion score in the experimental identification of abundant proteins in the laboratory (25 points), and percent identity with the nearest homolog (50 points). FASTA sequences for each of the short-listed proteins were retrieved from the reference strain of the selected bacterial species and subjected to global protein BLAST against the non-redundant protein database at NCBI. Low percent identity with the nearest homolog in any other species was given maximum weightage for a better probability of getting unique marker peptides for targeted MS based assay development. For example, on the two extremes, 100% sequence identity fetched 0 point (100–100) while $\leq 50\%$ identity gave complete 50 points for the identity parameter. Experimental identification of abundant proteins in the laboratory were assigned 25 points; significant MS/MS ion score for 1 peptide given 5 points and 5 additional points for every additional peptide up to 5 (≥ 5 peptides given complete 25 points). Number of independent reports was assigned 25 points with 5 points for 3 reports and 5 additional points for every additional report up to 7 (≥ 7 reports for the given protein assigned complete 25 points). This way, 15–27 proteins were short-listed as putative marker proteins for each of the selected species.

The FASTA sequences of the selected putative marker proteins were subjected to *in silico* tryptic digestion using the Peptide Mass algorithm at ExPASy Proteomics tools (<http://www.expasy.org>). All the peptides in the mass range of 1000–3000 Da obtained from the *in silico* digestion of putative marker proteins were subjected to global protein BLAST against the GeneBank non-redundant protein database (<http://www.ncbi.nlm.nih.gov>) with search parameters adjusted for short input sequence. Peptides showing less than 100% sequence identity with any other bacterial species (except the query species) were selected for the inclusion list of unique peptides. This selection was based on the assumption that a difference of even one single amino acid is most likely to ensure that the selected mass will not come from any other species except the target taxon. Peptide BLAST was given a tolerance for very closely species within the *Bacillus cereus* sensu lato group, including six other species (*B. cereus*, *B. mycoides*, *B. pseudomycoides*, *B. thuringiensis*, *B. weihenstephanensis*, and *B. cytotoxicus*). The tolerance within the *Bacillus cereus* sensu lato group meant that the peptide queried did not show 100% sequence identity with any other species but a complete match with one or more of the six species within the group was allowed. Similarly, the closely related species of *Burkholderia* (*B. mallei* and *B. pseudomallei*) and *Brucella* (*B. melitensis*, *B. abortus*, and *B. suis*) were considered together for the peptide BLAST search and scoring unique peptides. In order to discriminate between these closely related pathogens, the list of putative markers was further appended with peptides unique to these related species using reported literature. The unique peptides in the inclusion list were further screened for strain coverage using all the available sequences retrieved by taxon specific BLAST search and counting for the presence of a given peptide from the alignment data of these sequences from diverse strains in the genome data base. The details of *in silico* analysis and a summary of results are presented in the Supplementary Material S6.

Proof-of-concept studies. Number coded, eight blind samples of soil or sand were spiked with none, one or two bacterial select agents and subjected to targeted and shotgun analysis using tandem mass spectrometry after extraction of proteins. *B. anthracis* Sterne strain and *C. perfringens* ATCC13124 were used for the spiking experiment. Exponentially growing cultures were harvested and washed with phosphate buffer saline and resuspended in 50 mM Tris.HCl (pH 7.2). Cells were added to garden soil and sand (sieved into particle size range of 1–2 mm) at a concentration of 10^6 or 10^7 cfu.g⁻¹. After addition of the cell suspension (1.2 mL.g⁻¹ soil or sand), the matrix was dried overnight at 37 °C to simulate the possible binding of the cells to matrices in course of drying under field conditions. Total protein was extracted by the method modified from Chourey *et al.*⁶⁵. Briefly, 2 g of spiked sand or soil was resuspended in 1 ml distilled water and 1.5 ml Alkaline-SDS Buffer (5% SDS; 50 mM Tris.HCl; 0.15 M NaCl; 0.1 mM EDTA; 1 mM MgCl₂; 50 mM DTT). After intermittent vortexing for 10 min at room temperature, the mixture was heated at 95 °C for 10 min and allowed to cool at room temperature. Supernatant was collected by centrifugation at 10,000 × g for 30 min at room temperature and protein was precipitated by trichloro-acetic acid (TCA) at a final concentration of 10% (w/v) in the presence of β -mercaptoethanol (0.07%). The protein pellet collected by centrifugation (10,000 × g, 4 °C, 10 min) after 2 h of incubation on ice and washed twice with 500 μ L of acetone. The pellet was air-dried for 5 min and the protein pellet was solubilized in buffer containing 8 M Urea and 2% CHAPS followed by a heating at 95 °C for 5 min in SDS lysis buffer before subjecting

it to discontinuous denaturing SDS-PAGE. After staining with Coomassie Brilliant Blue G-250, abundant proteins from Coomassie-stained SDS-PAGE were excised and subjected to in gel digestion and identification by tandem mass spectrometry using shot-gun and targeted approach as described in the text.

Received: 14 May 2019; Accepted: 22 January 2020;

Published online: 10 February 2020

References

1. Wagar, E. Bioterrorism and the Role of the Clinical Microbiology Laboratory. *Clin. Microbiol. Rev.* **29**(1), 175–189 (2016).
2. <http://www.selectagents.gov>, Accessed on 14th April (2019).
3. <https://emergency.cdc.gov/agent/agentlist-category.asp>, Accessed on 14th April (2019).
4. Hsu, V. P. *et al.* *Emerging Infect. Dis.* **8**, 1039–1043 (2002).
5. World Health Organization Group of Consultants. Health aspects of chemical and biological weapons; WHO: Geneva, p 84–100 (1970).
6. Rajoria, S., Kumar, R. B., Gupta, P. & Alam, S. I. Post-exposure recovery and analysis of biological agent in a simulated biothreat scenario using tandem mass spectrometry. *Anal. Chem.* **89**, 4062–4070 (2017).
7. Radstrom, P., Knutsson, R., Wolffs, P., Lovenklev, M. & Lofstrom, C. Pre-PCR processing: strategies to generate PCR-compatible samples. *Biotechnol.* **26**, 133–146 (2004).
8. Burnett, J. C., Henchal, E. A., Schmaljohn, A. L. & Bavari, S. The evolving field of biodefence: therapeutic developments and diagnostics. *Nat. Rev. Drug Discov.* **4**, 281 (2005).
9. Havlicek, V., Lemr, K. & Schug, K. A. Current trends in microbial diagnostics based on mass spectrometry. *Anal. Chem.* **85**, 790 (2013).
10. Ivnitski, D. *et al.* Nucleic acid approaches for detection and identification of biological warfare and infectious disease agents. *Biotechniques* **35**, 862 (2003).
11. Jones, S. W., Dobson, M. E., Francesconi, S. C., Schoske, R. & Crawford, R. DNA assays for detection, identification, and individualization of select agent microorganisms. *Croat. Med. J.* **46**, 522 (2005).
12. Sharma, S. K. & Whiting, R. C. Methods for detection of Clostridium botulinum toxin in foods. *J. Food Prot.* **68**, 1256 (2005).
13. Mabry, R. *et al.* Detection of anthrax toxin in the serum of animals infected with Bacillus anthracis by using engineered immunoassays. *Clin. Vaccine Immunol.* **13**, 671 (2006).
14. Ler, S. G., Lee, F. K. & Gopalakrishnakone, P. Trends in detection of warfare agents. Detection methods for ricin, staphylococcal enterotoxin B and T-2 toxin. *J. Chromatogr. A* **1133**, 1 (2006).
15. Worbs, S. *et al.* Ricinus communis intoxications in human and veterinary medicine—a summary of real cases. *Toxins (Basel)* **3**, 1332 (2011).
16. Peruski, A. H. & Peruski, L. F. Jr. Immunological methods for detection and identification of infectious disease and biological warfare agents. *Clin. Diagn. Lab. Immunol.* **10**, 506 (2003).
17. Andreotti, P. E. *et al.* Immunoassay of infectious agents. *Biotechniques* **35**, 850 (2003).
18. Cheng, K., Chui, H., Domish, L., Hernandez, D. & Wang, G. Recent development of mass spectrometry and proteomics applications in identification and typing of bacteria. *Proteomics Clin. Appl.* **10**, 346–357 (2016).
19. Doern, C. D. Charting uncharted territory: a review of the verification and implementation process for matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) for organism identification. *Clin. Microbiol. News.* **35**, 69–78 (2013).
20. Clark, A. E., Kaleta, E. J., Arora, A. & Wolk, D. M. Matrix-assisted laser desorption ionization–time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* **26**, 547–603 (2013).
21. Kliem, M. & Sauer, S. The essence on mass spectrometry based microbial diagnostics. *Curr. Opin. Microbiol.* **15**, 397–402 (2012).
22. Tracz, D. M., Antonation, K. S. & Corbett, C. R. Verification of a Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry Method for Diagnostic Identification of High-Consequence Bacterial Pathogens. *J. Clin. Microbiol.* **54**(3), 764–7 (2016).
23. Loff, M., Mare, L., de Kwaadsteniet, M. & Khan, W. Molecular detection system versus MALDI-TOF mass spectrometry and molecular techniques for the identification of Escherichia coli 0157:H7, Salmonella spp. & Listeria spp. *J. Microbiol. Methods* **101**, 33–43 (2014).
24. Richter, S. S. *et al.* Identification of Enterobacteriaceae by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using the VITEK MS system. *Eur. J. Clin. Microbiol. Infect. Dis* **32**, 1571–1578 (2013).
25. Balazova, T. *et al.* The influence of culture conditions on the identification of Mycobacterium species by MALDI-TOF MS profiling. *FEMS Microbiol. Lett.* **353**, 77–84 (2014).
26. Raemy, A. *et al.* Phenotypic and genotypic identification of streptococci and related bacteria isolated from bovine intramammary infections. *Acta. Vet. Scand.* **55**, 53 (2013).
27. Patel, R. MALDI-TOF MS for the diagnosis of infectious diseases. *Clin. Chem.* **61**, 100–111 (2015).
28. Tracz, D. M., Antonation, K. S. & Corbett, C. R. Verification of a matrix assisted laser desorption ionization–time of flight mass spectrometry method for diagnostic identification of high-consequence bacterial pathogens. *J. Clin. Microbiol.* **54**, 764–767 (2016).
29. Seng, P. *et al.* Identification of rare pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser desorption ionization–time of flight mass spectrometry. *J. Clin. Microbiol.* **51**, 2182 (2013).
30. Duriez, E. *et al.* Detection of ricin in complex samples by immunocapture and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Proteome Res.* **7**, 4154 (2008).
31. Brinkworth, C. S. Identification of ricin in crude and purified extracts from castor beans using on-target tryptic digestion and MALDI mass spectrometry. *Anal. Chem.* **82**, 5246 (2010).
32. Kull, S. *et al.* Multiplex detection of microbial and plant toxins by immunoaffinity enrichment and matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **82**, 2916 (2010).
33. Fox, K., Fox, A., Rose, J. & Walla, M. Speciation of coagulase negative staphylococci, isolated from indoor air, using SDS PAGE gel bands of expressed proteins followed by MALDI TOF MS and MALDI TOF-TOF MS-MS analysis of tryptic peptides. *J. Microbiol. Methods* **84**, 243 (2011).
34. Alam, S. I., Kumar, B. & Kamboj, D. V. Multiplex detection of protein toxins using MALDI-TOF-TOF tandem mass spectrometry: application in unambiguous toxin detection from bioaerosol. *Anal. Chem.* **84**, 10500 (2012).
35. Gekenidis, M. T., Studer, P., Wuthrich, S., Brunisholz, R. & Drissner, D. Beyond the matrix-assisted laser desorption ionization (MALDI) biotyping workflow: in search of microorganism-specific tryptic peptides enabling discrimination of subspecies. *Appl. Environ. Microbiol.* **80**, 4234–4241 (2014).
36. Chenau, J. *et al.* Detection of Yersinia pestis in environmental and food samples by intact cell immunocapture and liquid chromatography tandem mass spectrometry. *Anal. Chem.* **86**, 6144 (2014).
37. Chenau, J. *et al.* Identification and validation of specific markers of Bacillus anthracis spores by proteomics and genomics approaches. *Mol. Cell. Proteomics* **13**, 716 (2014).

38. Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol* **25**(1), 125–31 (2007).
39. Colangelo, C. M., Chung, L., Bruce, C. & Cheung, K. H. Review of software tools for design and analysis of large scale MRM proteomic datasets. *Methods* **61**(3), 287–298 (2013).
40. Demeure, K., Duriez, E., Domon, B. & Niclou, S. P. PeptideManager: a peptide selection tool for targeted proteomic studies involving mixed samples from different species. *Front Genet.* **5**, 305 (2014).
41. Vandembrouck, Y., Christy D., Combes, F., Loux, V. & Brun, V. Bioinformatics Tools and Workflow to Select Blood Biomarkers for Early Cancer Diagnosis: An Application to Pancreatic Cancer. *Proteomics*, <https://doi.org/10.1002/pmic.201800489> (2019)
42. Keerthikumar, S. & Mathivanan, S. Proteotypic Peptides and Their Applications. *Methods Mol Biol.* **1549**, 101–107 (2017).
43. Mesuere, B., Van der Jeugt, F., Devreese, B., Vandamme, P. & Dawyndt, P. The unique peptidome: Taxon-specific tryptic peptides as biomarkers for targeted metaproteomics. *Proteomics*. **16**(17), 2313–8 (2016).
44. Dworzanski, J. P., Dickinson, D. N., Deshpande, S. V., Snyder, A. P. & Eckenrode, B. A. Discrimination and phylogenomic classification of *Bacillus anthracis-cereus-thuringiensis* strains based on LC-MS/MS analysis of whole cell protein digests. *Anal. Chem.* **82**, 145 (2010).
45. Seibold, E., Maier, T., Kostrzewa, M., Zeman, E. & Splettstoesser, W. Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. *J. Clin. Microbiol.* **48**, 1061 (2010).
46. Dybwad, M., van der Laaken, A. L., Blatny, J. M. & Paauw, A. Rapid identification of *Bacillus anthracis* spores in suspicious powder samples by using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS). *Appl. Environ. Microbiol.* **79**, 5372 (2013).
47. Duriez, E., Armengaud, J., Fenaille, F. & Ezan, E. Mass spectrometry for the detection of bioterrorism agents: from environmental to clinical applications. *J. Mass Spectrom.* **51**, 183–199 (2016).
48. Madeira, J. P., Alpha-Bazin, B., Armengaud, J. & Duport, C. Time dynamics of the *Bacillus cereus* exoproteome are shaped by cellular oxidation. *Front Microbiol.* **6**, 342 (2015).
49. Lasch, P. *et al.* Identification of *Bacillus anthracis* by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. *Appl. Environ. Microbiol.* **75**, 7229 (2009).
50. Lasch, P. *et al.* Identification of highly pathogenic microorganisms by matrix-assisted laser desorption ionization-time of flight mass spectrometry: results of an inter-laboratory ring trial. *J. Clin. Microbiol.* **53**, 2632–2640 (2015).
51. Christie-Oleza, J. A., Miotello, G. & Armengaud, J. Proteogenomic definition of biomarkers for the large *Roseobacter* clade and application for a quick screening of new environmental isolates. *J. Proteome Res.* **12**, 5331 (2013).
52. Christie-Oleza, J. A. *et al.* Shotgun nano LC-MS/MS proteogenomics to document MALDI-TOF biomarkers for screening new members of the *Ruegeria* genus. *Environ. Microbiol.* **15**, 133 (2013).
53. Durighello, E., Bellanger, L., Ezan, E. & Armengaud, J. Proteogenomic biomarkers for identification of *Francisella* species and subspecies by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. *Anal. Chem.* **86**, 9394 (2014).
54. Cunningham, S. A. & Patel, R. Importance of using Bruker's security-relevant library for Biotyper identification of *Burkholderia pseudomallei*, *Brucella* species, and *Francisella tularensis*. *J. Clin. Microbiol.* **51**, 1639–1640 (2013).
55. Keller, P. M., Bruderer, V. & Muller, F. Restricted identification of clinical pathogens categorized as biothreats by MALDI-TOF mass spectrometry. *J. Clin. Microbiol.* **54**, 816 (2016).
56. Dingle, T. C., Butler-Wu, S. M. & Abbott, A. N. Accidental exposure to *Burkholderia pseudomallei* in the laboratory in the era of matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* **52**, 3490–3491 (2014).
57. Dworzanski, J. P. *et al.* Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. *J. Proteome Res.* **5**, 76 (2006).
58. Al-Shahib, A. *et al.* Coherent pipeline for biomarker discovery using mass spectrometry and bioinformatics. *BMC Bioinformatics* **11**, 437 (2010).
59. USAMRIID. Medical Management of Biological Casualties Handbook, <http://www.usamriid.army.mil/education/bluebookpdf/> (2011).
60. Marx, V. Targeted proteomics. *Nat. Methods* **10**, 19 (2013).
61. Gallien, S. *et al.* Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol. Cell. Proteomics* **11**, 1709 (2012).
62. Charretier, Y. *et al.* Rapid bacterial identification, resistance, virulence and type profiling using selected reaction monitoring mass spectrometry. *Sci. Rep.* **5**, 13944 (2015).
63. Ludwig, C. *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**(8), e8126 (2018).
64. Chaubey, K. *et al.* From the Cover: Proteome Profile of Different Rat Brain Regions After Sarin Intoxication. *Toxicol Sci.* **160**(1), 136–149 (2017).
65. Chourey, K. *et al.* Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J. Proteome Res.* **9**(12), 6615–22 (2010).

Acknowledgements

The authors are thankful to the Director, DRDE, Gwalior, India, for providing facilities for the study. The work has been funded by the Ministry of Defence, Government of India. Ms Sakshi Rajoria received fellowship from Department of Biotechnology, India during the period of study.

Author contributions

S.R. carried out major part of the experiment and compiled data. S.S., P.B., D.V.K., S.K., R.B.K., helped in validation experiments. S.K. was involved in experiments carried out in BSL3 facility. S.I.A. designed the experiment, analyzed data, carried out mass spectrometry work and wrote paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-59156-3>.

Correspondence and requests for materials should be addressed to S.I.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020