Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Rapid Detection & Identification of *Bacillus* Species using MALDI-TOF/TOF and Biomarker Database

A Strategic Plan

Nora W.C. Chan and William E. Lee
Defence R&D Canada – Suffield

Zoltán Mester
Institute for National Measurement Standards, National Research Council

Technical Memorandum
DRDC Suffield TM 2006-078
June 2006

Canada

Rapid Detection & Identification of *Bacillus* Species using MALDI-TOF/TOF and Biomarker Database

A Strategic Plan

Nora W.C. Chan and William E. Lee
Defence R&D Canada – Suffield

Zoltán Mester
Institute for National Measurement Standards, National Research Council

Defence R&D Canada – Suffield

Technical Memorandum

DRDC Suffield TM 2006-078

June 2006

Author

Dr. N.W.C. Chan

Approved by

Dr. L. P. Nagata

Head, Chemical & Biological Defence Section

Approved for release by

Dr. P. A. D'Agostino

Chair, DRDC Suffield Document Review Panel

Abstract

A strategic plan to develop a matrix-assisted laser desorption/ionization with time-of-flight (MALDI-TOF/TOF) tandem mass spectrometry (MS) method in conjunction with a biomarker database for rapid microbial identification is described in this technical memorandum. The objective is to develop rapid, highly sensitive and definitive protocols for detection and identification of BW agents such as *Bacillus* species and demonstrate their effectiveness. The protocols will be based on MALDI-TOF MS technology applicable to intact microorganisms. MALDI can directly analyze biological samples for biomarkers (proteins, lipids, etc.). Forensic biomarkers can be used to identify microorganisms at species and strain levels. In addition to method development, a new mass spectral biomarker database could be created and populated for selected species and strains, starting with selected *Bacillus* species. This technology would create a new area for comprehensive forensic analysis of biological materials that are not addressed by current methods (PCR and DNA sequencing, and immunoassays). This memorandum contains a review of *Bacillus* species and current detection and identification methods, and a review of MALDI-TOF/TOF MS methods and biomarker databases.

Résumé

Ce document technique décrit un plan stratégique consistant à mettre au point une méthode de spectrométrie de masse (SM) avec temps de vol en tandem avec ionisation – désorption par impact laser assistée par matrice (MALDI -TV/ TV) qui utilise parallèlement une base de données de dépistage biologique permettant une identification microbienne rapide. L'objectif est de mettre au point des protocoles rapides, hautement sensibles et définitifs de détection et d'identification d'agents de guerre biologiques tels que les espèces *Bacillus* et d'en démontrer leur efficacité. Les protocoles seront basés sur la technologie de SM MALDI-TV applicable à des microorganismes intacts. La technique MALDI peut directement analyser des échantillons biologiques pour des biomarqueurs (protéines, lipides, etc.). Les biomarqueurs médico-légaux peuvent être utilisés pour identifier des microorganismes au niveau des espèces et des souches. En plus de mettre au point une méthode, on pourrait créer une nouvelle base de données de dépistage biologique par spectre de masse et la peupler pour des espèces et des souches sélectionnées, en commençant avec des espèces *Bacillus* sélectionnées. Cette technologie créerait ainsi une nouvelle sphère pour les analyses médico-légales détaillées de matériaux biologiques dont ne traitent pas les méthodes actuelles (séquençage et immuno-essais PCR et ADN). Ce document comprend une étude des espèces *Bacillus* et des méthodes actuelles de détection et d'identification ainsi qu'une étude des méthodes MS TV /MALDI-TV et des bases de données de dépistage biologique.

This page intentionally left blank.

Executive summary

Introduction: Traditional bacterial taxonomy to characterize microorganisms is time-consuming, labour-intensive, and largely inaccurate. Recent advances in molecular biological methods can classify and identify bacteria with better accuracy, but still suffer drawbacks such as lack of viability (live/dead) information, and requires the specificity and availability of nucleic acid primers. Newer molecular sub-typing methods to classify bacterial strains and isolates are gaining interest in biodefence, yet each primer set experiment is specific to a certain bacterium and therefore cannot become a rapid, broad-based method for bacterial detection and identification.

Significance: This article describes a rapid bacterial detection and identification method using a combination of matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry (MALDI-TOF/TOF MS) and development of an in-house biomarker database. The proposed method uses a two-dimensional protein identification (mass-fingerprinting and proteomic-based sequencing technique) to correlate bacterial genotypic and phenotypic characteristics. As the function of a cell or a bacterium is based on proteins, rapid identification of specific proteins associated with specific functions, such as bacterial toxicity or viability would provide more in-depth information than nucleic acid based identification methods alone.

MALDI-TOF mass spectrometry has been a prominent technique in detection of biomolecules. With tandem mass spectrometry, MALDI-TOF/TOF MS, one can identify biomolecules such as proteins or peptides with high accuracy. In recent years, MALDI-TOF has been used expansively to profile bacterial proteins and glycoproteins, peptides, carbohydrates, and lipids. MS has the potential to become a broad-based biodetector with direct interrogation of whole bacterial cells, without much sample processing required.

A review of currently used bacterial detection and identification methods and tandem mass spectrometry of MALDI-TOF/TOF, and development of biomarker databases in public and private domains is included in this memorandum.

Future Research: Initially, the proposed work would focus on the *Bacillus cereus* group, which consisted of *B. anthracis*, *B. cereus*, and *B. thuringiensis*. Specific data relating to protein markers and live/dead identification in MS would be included in the development of the in-house biomarker database. Once the method is developed and proven with *Bacillus* species, the method could be expanded to other bioagents, viruses, and other bacteria implicated in human infectious diseases.

Chan, N.W.C., Lee, W.E., Mester, Z. 2006. Rapid detection & identification of *Bacillus* species using MALDI-TOF and biomarker database. DRDC Suffield TM 2006-078. Defence R&D Canada – Suffield.

Sommaire

Introduction: La taxonomie bactérienne classique caractérisant les microorganismes absorbe du temps, est exigeante en main-d'œuvre et souvent erronée. Les découvertes récentes en matière des méthodes en biologie moléculaire permettent de classifier et d'identifier des bactéries avec plus de précision mais elles ont encore des inconvénients tels que le manque d'information en viabilité (vivant/mort) et nécessitent la spécificité et la disponibilité d'initiateurs à l'acide nucléique. Des méthodes moléculaires de sous-typage plus récentes consistant à classifier les souches bactériennes et isolats sont actuellement examinées en bio-défense, cependant chaque ensemble d'expériences d'initiateurs est spécifique à une certaine bactérie ce qui ne permet pas de développer une méthode rapide de portée générale pour la détection et l'identification bactérienne.

La portée des résultats : Cet article décrit une méthode rapide de détection et d'identification bactérienne qui utilise une combinaison de la technique de spectrométrie de masse en tandem de ionisation – désorption par impact laser assistée par matrice avec temps de vol (SM TV / MALDI-TV) avec la mise au point d'une base de donnée de dépistage biologique interne. La méthode proposée utilise une identification de protéines bidimensionnelle (technique de cartographie peptidique de masse et de séquençage protéomique) pour corréler les caractéristiques génotypiques et phénotypiques bactériennes. La fonction d'une cellule ou d'une bactérie étant basée sur les protéines, l'identification rapide de protéines spécifiques associées à des fonctions spécifiques telles que la toxicité bactérienne ou la viabilité fournirait une information plus approfondie que des seules méthodes d'identification à base d'acide nucléique.

La spectrométrie de masse MALDI-TV a été une technique proéminente en matière de détection de biomolécules. Avec la spectrométrie de masse en tandem, MS TV /MALDI-TV, on peut identifier des biomolécules telles que les protéines ou les peptides avec haute précision. La méthode MALDI-TV a récemment été largement utilisée pour profiler des protéines bactériennes, des glycoprotéines, peptides, glucides et lipides. La SM a le potentiel de devenir un bio-détecteur à large assise interrogeant directement des cellules bactériennes entières sans avoir besoin de traiter beaucoup d'échantillons.

On inclut, dans ce document, un examen des méthodes de détection et d'identification bactériennes actuellement utilisées et de spectrométrie de masse en tandem de MALDI-TV /TV ainsi que le développement de bases de donnée de dépistage biologique dans des domaines privés et publics.

Travaux de recherche futurs : Les travaux proposés devraient initialement être axés sur le groupe *Bacillus cereus,* comprenant *B. anthracis*, *B. cereus*, and *B. thuringiensis*. Les données spécifiques reliées aux marqueurs de protéines et à l'identification vivant/mort dans la SM devraient être inclues dans le développement de la base de données interne de dépistage biologique. Une fois que la méthode sera mise au point et prouvée avec des espèces *Bacillus,* la méthode pourra être étendue à d'autres bio-agents, virus et autres bactéries impliquées dans les maladies infectieuses humaines.

Chan, N.W.C., Lee, W.E., Mester, Z. 2006. Rapid detection & identification of *Bacillus* species using MALDI-TOF and biomarker database. DRDC Suffield TM 2006-078. R&D pour la défense Canada – Suffield.

Table of contents

Abstract i

Résumé i

Executive summary iii

Sommaire iv

Table of contents v

List of figures vi

List of tables vi

Acknowledgements vii

Introduction 1

The Genus *Bacillus* 2

Traditional bacterial taxonomy 2

Genetic identification methods 3

Overview of MALDI-TOF 6

Matrix-assisted laser desorption ionization 7

Time-of-flight mass analyzer 8

MS/MS detection & identification 9

Biomarker database 12

Future Research 13

Discussion & Conclusions 15

References 16

List of symbols/abbreviations/acronyms/initialisms 24

List of figures

Figure 1. Compare tradition taxonomy, genetic analysis and proposed MALDI-TOF MS method for detection and identification of *Bacillus* species 5

Figure 2. Principle of the matrix-assisted laser desorption/ionization (MALDI) process.. 7

Figure 3. Principle of the time-of-flight mass analyzer. 9

Figure 4. MALDI-TOF/TOF instrument 10

Figure 5. Biemann nomenclature for peptide CID fragment ions 10

Figure 6. The scheme of a rapid MALDI TOF/TOF MS-based microorganism identification using a combination of mass fingerprinting, peptide sequence information, and proteomic database search engine 13

List of tables

Table 1. Basic characteristics for identification of selected *Bacillus* species. 3

Acknowledgements

The authors would like to thank Mr. Doug Bader for his professional advice and knowledge of all the new molecular subtyping identification techniques.

This page intentionally left blank.

Introduction

The detection and identification of microorganisms using current methods is an arduous task. Traditionally, clinical diagnostics using morphological/ physiological, biochemical, and chemotaxonomic characteristics of microorganisms have been used for species classification and identification. This type of characterization is time-consuming, labour-intensive, and largely inconclusive. However, recent advances in molecular biology suggest that classification and identification of microorganisms reflecting relationships encoded in DNA sequences are much more reliable [1].

PCR (polymerase chain reaction) is increasingly being used for diagnostic and forensic purposes. However, the technique is limited by the specificity and availability of the PCR primers that are used to amplify targeted sequences. As more DNA sequence and experimental data become available, it is increasingly clear that many sequences thought to be target-specific are shared by other, sometimes unrelated, species. This problem is compounded because portions of particular virulence genes are shared among pathogenic and non-pathogenic species; they sometimes encode a gene with similar functions. It therefore becomes critical to understand whether the targeted DNA sequences are, indeed, species- or strain-specific. Additionally there is only a single copy of DNA per cell, therefore amplification is necessary in any genetic analyses.

Mass spectrometry (MS) has been applied to many areas of microbiology. Laser desorption, plasma desorption, and fast atom bombardment (FAB)-mass spectrometry have been used to analyze lipids from lyophilized membranes and cells in microorganisms [2, 3]. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has become a prominent technique in biological mass spectrometry since it was invented in 1987 [4, 5]. The development of MALDI-TOF MS provided high performance mass spectrometry for the study of proteins and glycoproteins [6], peptides, lipids [2, 7 - 9], lectins and carbohydrates [10] in microorganisms; and extensive studies of small, acid-soluble proteins (SASPs) in bacterial spores [11 - 17]. The method has been used expansively to profile bacterial proteins from cell extracts and intact cells [18 - 20]; to identify Gram-positive and Gram-negative bacteria [20]; to rapidly identify bacterial strains via a compilation of MALDI-TOF mass spectral database [21 - 29]; and rapid virus identification [6, 30], and quantitative determination of bacillus spores for forensic characterization [31].

MS has the potential to be the most broadband detector because everything (proteins, lipids, nucleic acids, carbohydrates, metabolites, etc) has a specific mass [32]. But for MS to become widely accepted as a bioagent detector, the method must overcome some challenges. In this article, we propose a rapid bacterial detection and identification method using MALDI-TOF/TOF tandem mass spectrometry in conjunction to an in-house biomarker database.

The Genus *Bacillus*

Bacillus species are characterized as Gram-positive, rod-shaped, aerobic, usually catalase-positive, endospore-forming bacteria [33, 34]. This genus is one of the largest and most ubiquitous, and has gained notoriety with taxonomists for its extreme phenotypic diversity and heterogeneity. Most *Bacillus* species are regularly encountered and cultivated from soil samples, their primary habitat. The *Bacillus cereus* group, a very homogeneous cluster within the *Bacillus* genus (recently termed *B. cereus sensu lato*), comprises four recognized species: *B. cereus, B. thuringiensis, B. anthracis,* and *B. mycoides* [35 - 37]. These *Bacillus* species can contaminate almost any environment including food and clean rooms, causing various infections and foodborne illnesses. For example, *B. anthracis*, *B. cereus*, and *B. thuringiensis* are genetically closely related, but phenotypically very diverse. *B. anthracis* produces a toxin that is the cause of the acute and often lethal disease anthrax; *B. cereus* is a ubiquitous soil bacterium and plays a role in food-transmitted gastroenteritis; *B. thuringiensis* is a very useful source of insecticidal toxins, and is widely used in agriculture and horticulture.

Bacillus species form endospores in the presence of oxygen when under stress. The spores are resistant to heat, cold, radiation, desiccation, and disinfectants, and are ubiquitous in soil. The spores may survive in the environment for decades in certain soil conditions [38].

Traditional bacterial taxonomy

Taxonomy is defined as the study of the general principles of scientific classification [39]. Bacterial classification is mainly based on morphology and staining methods in several categories, morphology (cocci, bacilli, spiral and pleomorphic), motility, stains (Gram positive or Gram negative, acid fast), oxygen requirements (aerobic or anaerobic), spore-forming capabilities, culture conditions and requirements, antigenic properties, biochemical reactions (catalase or oxidase activity), and mol% G + C content. Table 1 lists the identifying characteristics of some of the more common *Bacillus* species. One can see ambiguities arise if identification is based solely on morphological taxonomy.

In the case of *B. anthracis*, an isolate with the characteristic colonial morphology on nutrient of blood agar (mat appearance, fairly flat, markedly tacky, white or grey-white and non-haemolytic on blood agar, and often having curly tailing at the edge), which is aerobic, non-motile, non-hemolytic, sensitive to penicillin and the diagnostic "gamma" phage, and able to produce the capsule in blood or on nutrient agar containing 0.7% bicarbonate following incubation in a 5-20% CO_2 atmosphere, is thus identified as virulent *B. anthracis* [40]. Dragon *et al.* reported a shortened procedure of isolating colonies from a selective media, and then testing with blood agar and penicillin; such non-haemolytic, penicillin-sensitive isolates resulted in 78% of samples being proven positive by PCR [41]. This shows that relying solely on taxonomical methods will result in false positives (22% in [41]) or in worst cases result in false negatives (missed identification). Additionally, this reported process is labour-intensive and tedious, requiring days before positive identification.

Table 1. Basic Characteristics for Identification of Selected Bacillus Species (excerpt from [42])

SPECIES	MOTILITY	CATALASE PRODUCTION	PARASPORAL BODIES	LIPID GLOBULES IN PROTOPLASM	LECITHOVITELLIN REACTION	CITRATE UTILIZATION	ANAEROBIC GROWTH	V-P REACTION	pH (IN V-P MEDIUM) <6.0	GROWTH AT 50C	GROWTH AT 60C	GROWTH IN 7% NaCl	ACID FROM AS GLUCOSE	ACID + GAS FROM AS GLUCOSE	NITRATE REDUCTION	CASEIN HYDROLYSIS	STARCH HYDROLYSIS	PROPRIONATE UTILIZATION
B. megaterium	v	+	-	+	-	+	-	-	v	-	-	+	+	-	v	+	+	n
B. cereus	+	+	-	+	+	+	+	+	+	-	-	+	+	-	+	+	+	n
B. cereus subsp mycoides	-	+	-	+	+	+	+	+	+	-	-	+	+	-	+	+	+	n
B. anthracis	-	+	-	+	+	v	+	+	+	-	-	+	+	-	+	+	+	n
B. thuringiensis	+	+	+	+	+	+	+	+	+	-	-	+	+	-	+	+	+	n
B. licheniformis	+	+	-	-	-	+	+	+	+	+	-	+	+	-	+	+	+	+
B. subtilis	+	+	-	-	-	+	-	+	v	v	-	+	+	-	+	+	+	-
B. pumilus	+	+	-	-	-	+	-	+	+	v	-	+	+	-	-	+	-	-
B. firmus	v	+	-	-	-	-	-	-	-	-	-	+	+	-	+	+	+	-
B. coagulans	+	+	-	-	-	v	+	+	+	+	v	-	+	-	v	v	+	-

V-P, Voges-Proskauer; AS, ammonium salt; +. More than 85% of strains examined by Gordon et al. [43] were positive; -, more than 85% of strains negative; v, variable; n, test not applicable.

Bacterial taxonomical classification is more difficult than elemental molecular analyses.

Genetic identification methods

Bacillus anthracis is a member of the *B. cereus*/*B. thuringiensis* phylogenetic group [35, 40 - 48]. The members of this group are virtually indistinguishable by 16S and 23S rRNA sequence analysis. Multilocus enzyme electrophoresis (MEE) and comparative DNA sequence analysis suggest that they may represent a single species [46]. *B. anthracis* isolates are extremely uniform in chromosome composition and are limited to the presence of the two virulence plasmids, pX01 and pX02. This conserved state reflects the relatively recent

evolution of *B. anthracis* from a parental *B. cereus* subgroup. DNA samples must be extracted from suspicious samples prior to any genetic analyses. PCR targeting a combination of lethal factor gene or protective antigen gene on pX01, capsule A, B or C gene on pX02, plus a chromosomal gene (*rpoB* or *gyrA*), followed by gel electrophoresis and DNA sequencing, can differentiate *B. anthracis* from other *Bacillus* species. Yet there are potential mis-classification because these assays are based on single-nucleotide differences between *B. anthracis* and other *Bacillus* species. Unwarranted assumptions are often made about the specificity of the DNA sequences targeted. It was once thought that targeting selected virulence genes would provide an unambiguous identification of a microbial pathogen within a complex sample. False positives were detected for non-suspicious soil samples using a PCR-ELISA technique with primers targeting the B or C gene of pX02; four isolated strains belonging to the *B. cereus* group were identified [48].

Further strain identification requires sequencing the complete genome of the *Bacillus spp.* which takes weeks or even months for positive results [49 - 53]. The complete sequencing of the protective antigen (PA) gene sequence from 26 different *Bacillus anthracis* strains revealed only five point mutations confirming the lack of genetic diversity [46, 54 - 57] in this bacterium.

Many researchers in the genetic analysis field have realized that positive identification requires detection of nucleic acid sequences at multiple locations (i.e. multilocus) rather than just one or two specific genes. Amplified fragment length polymorphism (AFLP) is a multilocus method that can be used to rapidly identify bacterial species and phylogenetic relationship to other microorganisms [57]. AFLP is simple and provides quick analysis to differentiate bacterial species, but lacks the further differentiation power to separate *B. anthracis* strains and isolates [58].

Molecular sub-typing methods, such as multiple-locus variable-number tandem repeat analysis (MLVA), were used to determine the strain of *B. anthracis* used in the 2001 bioterrorism-associated outbreak [56, 59, 60] and to identify other outbreaks [58, 61, 62]. The work reported by Hoffmaster et al. required sub-typing 135 isolates and sequencing the protective antigen gene from 42 representative isolates for confirmation of the identity of the *B. anthracis* strain. MLVA which uses variable number of tandem repeats (VNTRs) has proven to be a suitable method for assessing genetic polymorphisms within bacterial species. This method still suffers drawbacks such as availability and selection of VNTR loci as suitable PCR priming sites. Multiloci sequence typing (MLST) is another recent sub-typing method based on sequencing a number of essential or housekeeping genes spread around the bacterial chromosome, a direct adaptation of the MEE method [63] but with greater discrimination [64]. All of these new PCR-based subtyping methods are certainly superior and more sophisticated than standard PCR which targets only a few gene segments.

The marked absence of reliable characteristics for rapidly delineating these species makes MALDI-TOF/TOF MS an attractive alternative method. Moreover, genetic analysis does not reveal the virulence of the identified *B. anthracis*, due to the fact that the technique involves only detection of nucleic acid sequences and cannot determine whether the bacteria are live or dead. Thus, it is necessary that bacterial colonies isolated from culture plates be used in genetic analyses. Any protein biomarkers that can indicate live/dead status of *B. anthracis* rapidly would be an asset to emergency response teams. Genetic analysis also requires use of

specific primer sets to target specific groups of bacteria, therefore each bacterium requires its own set of experiments. MADLI-TOF/TOF MS method in combination with a biomarker database search system will be a universal detection and identification method for microorganisms.

Figure 1 shows a comparison of time required for proper bacterial identification using traditional taxonomy, genetic analysis, and the proposed MALDI-TOF/TOF MS method. The significantly short time expected in the proposed method is due to minimal sample handling and preparation, no need for protein separation or bacteria growth for confirmation of viability.

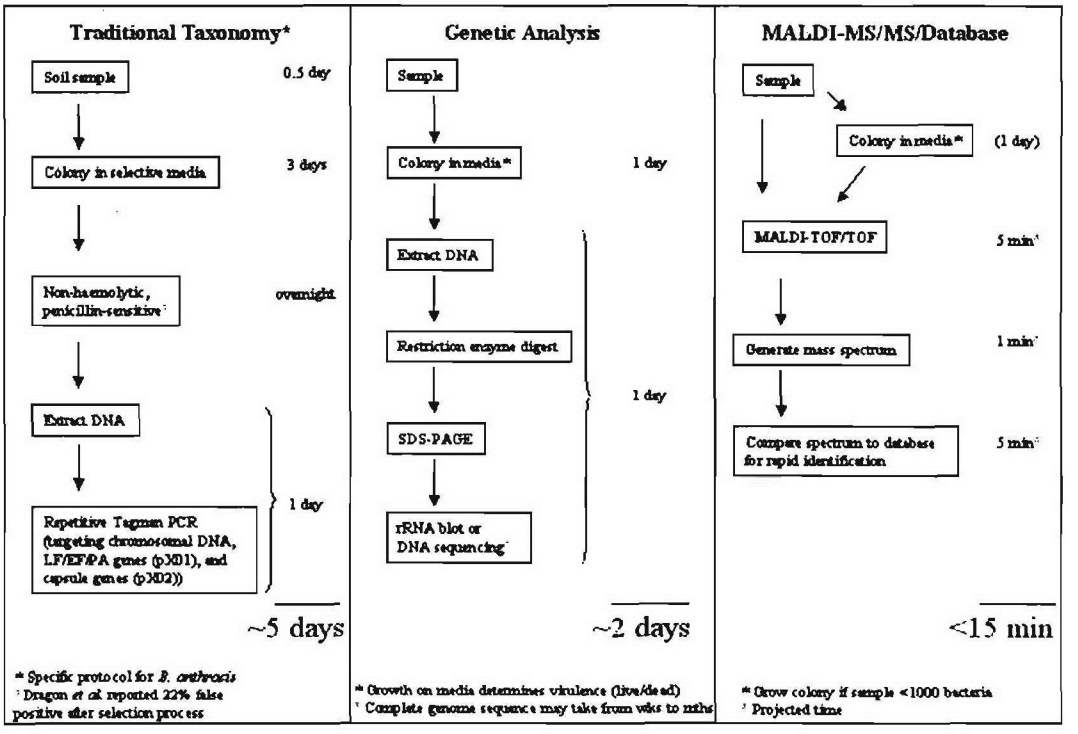

Figure 1. Compare traditional taxonomy, genetic analysis and proposed MALDI-TOF MS method for detection and identification of Bacillus species.

Overview of MALDI-TOF

Mass spectrometry has emerged as an important tool for analyzing and characterizing biomolecules of varying complexity. The matrix assisted laser desorption/ionization (MALDI) technique [65], first introduced in 1985, has increased the upper mass limit for mass spectrometric analyses of biomolecules to over 300,000 Da, and has enabled the analysis of a wide range of biomolecules by mass spectrometry.

MALDI-time-of-flight (TOF) mass spectrometry is now a mature technique offering extremely fast and accurate identification of organic molecules. MALDI requires an ultraviolet absorbing matrix which is mixed with the analyte leading to the co-crystallization of analyte with the matrix. When a pulsed laser beam tuned to the appropriate frequency hits the crystalline structure, the energy is initially transferred to the matrix (present in large excess) which is then partially vaporized. Intact target molecules are transformed into ions and then entered into the gas phase enabling mass spectrometry detection. In some cases a full molecular mass spectrum can be collected from a single laser shot, offering an extremely fast and information-rich sample interrogation method.

Microbial taxonomy based on mass spectrometry has been investigated for over 30 years. During this time the pool of accessible biomarkers has constantly increased, paralleling the development of ionization methods and other mass spectrometric technologies. Currently the most intensively studied potential biomarkers are the proteins. In order to correlate cell function to unique biomarkers, proteins are prime candidates because almost all function is governed by proteins. Proteins contribute about 50% of the dry weight of bacteria and are distributed among 200-6000 molecules. Though DNA constitutes the most unique biomarker, there is only one copy per cell requiring an intense amplification process for analytical use. With this proposed MALDI-TOF/TOF MS technology, proteins provide the most characteristic biomarkers accessible in the analysis of intact organisms without extraction, separation, or amplification.

MS-based bacterial identification methods using protein biomarkers based on the comparison of a mass spectrum is a relatively simple procedure [21 - 24]. These methods require rigorous controls of experimental conditions such as cell growth media or temperature, otherwise results can be uncertain and erroneous. Bacterial identification based on comparison of a database generated from experiments will reflect a plethora of information that is otherwise not available from public databases due to insufficient information on complete genome sequence information and posttranslational modifications.

Mass spectrometry is an important technique for chemical and biological analyses and for the study of ion chemistry. There are five components in a mass spectrometer: sample introduction, ionization, mass analyzer, ion detector, and computer data system. The MALDI ionization component, mass analyzer and ion detector are operated under vacuum. Vacuum is required to reduce collisional defocusing (or scattering), electrical discharge and high background from air, all of which result in high sensitivity and resolution in detection. The ionization methods create ions in the gas phase. Soft (non-destructive) ionization methods such as electrospray ionization (ESI) and MALDI are popular choices for analysis of

biological samples. The mass analyzer functions to separate ions according to their mass-to-charge (*m/z*) ratio, and to maximize the resolved ion intensity. The ion detector (electron multiplier detector or multichannel plate detector) detects the individual ions after the mass analyzer and converts to a readable mass spectrum (ion intensity vs. *m/z*).

Matrix-assisted laser desorption ionization

MALDI as a soft ionization method for ionization of proteins was introduced by Karas and Hillenkamp in 1988 [66]. Koichi Tanaka first published MALDI spectra of proteins under 100,000 Da [67] and went on to win the 2002 Nobel Prize in chemistry for his work. In a MALDI experiment, a low molecular weight photoabsorbing organic compound (matrix) is added to the sample prior to irradiation with ultraviolet (UV) nanosecond laser pulses to desorb high molecular weight biomolecular ions (figure 2). The function of the matrix is to absorb the energy from the laser beam, and transfer energy to sample molecules via vibrational or electronic excitation. The matrix protects the sample molecules and promotes formation of intact molecular ions (M^+). Typical choices of matrix materials are α-cyano-hydroxycinnamic acid (CHCA), sinapinic acid (SA) and dihydroxybenzoic acid (DHB).

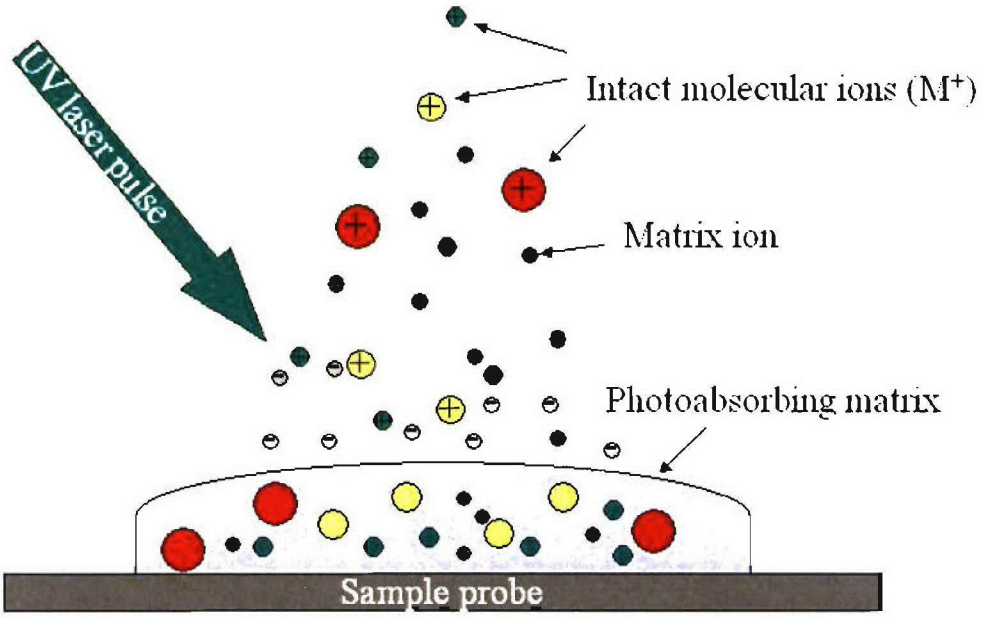

Figure 2. Principle of the matrix assisted laser desorption ionization (MALDI) process. Sample molecules are mixed with photoabsorbing matrix materials and loaded onto a sample probe prior to an experiment. UV laser pulses irradiate the sample/matrix mixture. Intact molecular ions (M^+) along with matrix ions are desorbed into the vacuum as a result of laser photon-matrix molecule interactions. All ions acquire kinetic energy proportional to their charge (z).

Time-of-flight mass analyzer

All ions generated from the MALDI process acquire kinetic energy proportional to their charges (z). The intact M^+ desorbs into a vacuum drift region (typically between 0.5 to 2 m long tube) where they are separated according to their charges (z) and masses (m). Ions with the same charge but different masses have different times of flight (t) through the drift region of the instrument. Calibration procedures correlate an ion's time of flight, measured in an experiment, with its mass. Equation 1 describes the relationships of time of flight, mass and charge of an ion.

$$zV = \frac{1}{2}mv^2 \qquad \text{where} \qquad v = \frac{L}{t}$$

$$t = \sqrt{\frac{1}{2}\frac{m}{zV}} \cdot L$$

Therefore, $$t \propto \sqrt{\frac{m}{z}} \qquad \text{Eq.1}$$

L is the length of the drift region, V is the voltage applied in the ionization region (typically in kV), t is the flight time through the drift region to the detector, m and z are the molecular mass and charge of the ion, respectively. The length, voltage and time can be obtained experimentally to determine the mass-to-charge ration (m/z) of the ion. Figure 3 shows a cartoon of the TOF mass analyzer separating ions with same charge but different masses.

All ions can be made to proceed from ionization source to detection and thus no upper theoretical mass limitation. Therefore, TOF mass analyzer is commonly used in the biological field for detection of large molecules up to 500 kDa. Most TOF spectrometers employ multichannel plate (MCP) detectors which have a rapid time response (<1 ns) and a high sensitivity. The large planar (2 cm diameter with a smooth, flat surface) detection area results in a large acceptance volume of the spectrometer. It is possible to detect many ions at the same time which is important for MALDI where hundreds of ions can be created within a few nanoseconds.

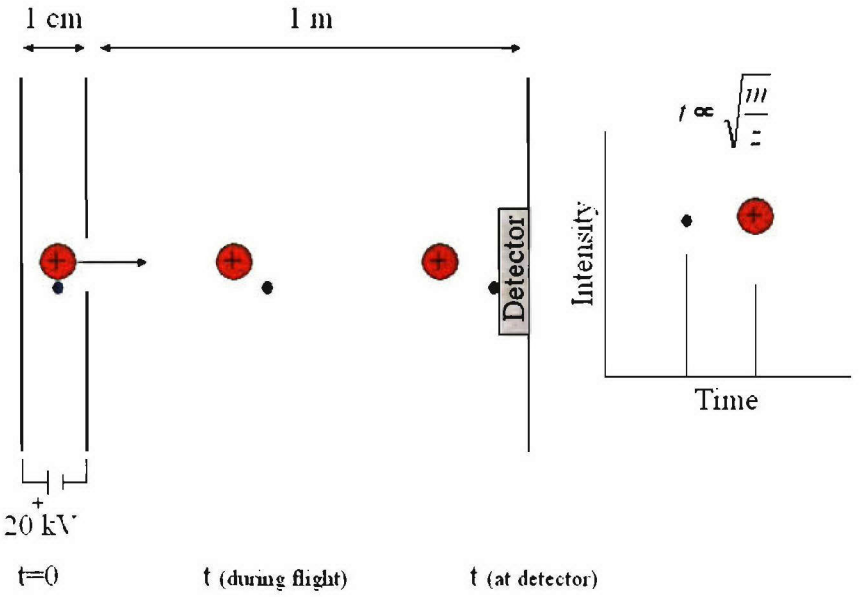

Figure 3. Principle of the time-of-flight mass analyzer. Ions formed in an ion source are extracted and accelerated to a high velocity by an electric field (typically 15-20 kV), then entered into an analyzer consisting of a long straight drift tube. The time taken for an ion to traverse the drift tube is proportional to the square root of its mass-to-charge ratio. If two ions both carry the same charge, the large ion (red) will travel down the tube at a lower velocity than the small ion (blue). The small ion reaches the detector before the large ion, and thus reflected on the spectrum (right) as shorter time.

MS/MS detection & identification

Despite all the advantages, MALDI-TOF has a major limitation, it reports molecular weight without any structural information. Thus it cannot be used to determine protein sequence – data critical to a protein chemist. To overcome this limitation, tandem mass spectrometry (MS/MS) is used to obtain sequence information for peptide and protein identification. MALDI-TOF/TOF MS combines the advantages of high sensitivity for peptide analysis associated with MALDI and comprehensive fragmentation information from high-energy collision-induced dissociation (CID) [68, 69] (Fig. 4). A TOF/TOF tandem mass spectrometer permits high resolution precursor-ion selection and high-energy collisional activation of the selected ion and recording of full fragment spectra. High-energy CID fragmentation pathways are usually rationalized by direct bond cleavages. The nature and the extent of fragmentation can be controlled by the choice of matrix (CHCA, DHB, etc.) and collision gas (He, Ar, Xe, etc.). The resulting fragmentation spectra provide comprehensive sequence and structural characterizations. Figure 5 shows the Biemann nomenclature [69] for peptide CID fragment ions. High-energy peptide CID fragmentation produces mostly side chain cleavage (*d*, *v*, *w* ions), as opposed to low-energy CID spectra exhibiting mostly peptide chain cleavages (*b* and *y* ions).

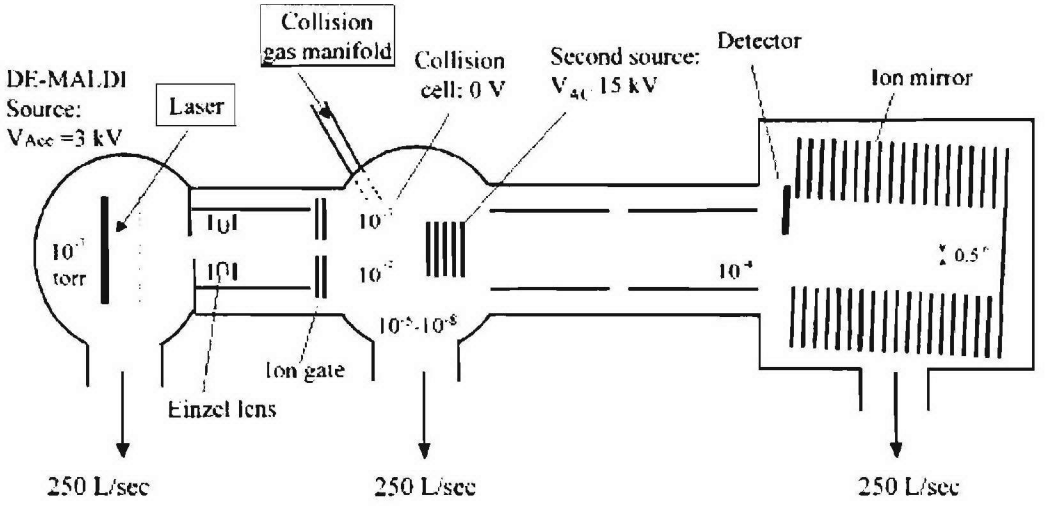

Figure 4. MALDI-TOF/TOF instrument (after [68]). The tandem TOF-TOF has a high-energy collision cell placed between two TOF mass analyzers. The sample is irradiated by laser in the ionization chamber, then accelerated into the first TOF drift tube. The precursor ion of interest is selected at the TOF1 stage. In the collision cell, mostly single collision events are observed and ion excitation is mainly electronic. Daughter ions resulting from fragmentation enter the second TOF stage and separated by mass and charge before reaching the detector.

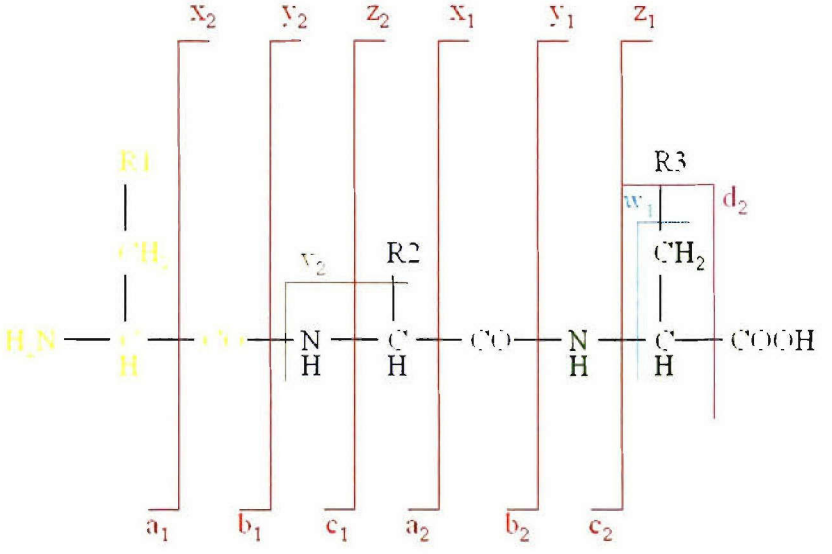

Figure 5. Biemann nomenclature for peptide CID fragment ions. N-terminal fragment ions (those containing the N-terminus of the peptide) are labeled with the first few letters of the alphabet (a, b, c, d), whereas C-terminal fragment ions (those containing the C-terminus of the peptide) are labeled with the last few letters of the alphabet (v, w, x, y, z). High-energy CID of peptide often provides additional fragmentation from the cleavage of the bond between amino acid β- and γ-carbon atoms resulting in the loss of part of the side chain of the amino acid (d, v, w ions).

There are two distinct approaches in research related to MALDI-based bacterial identification. The most commonly used approach is protein mass fingerprinting, i.e. a mass spectrum reporting molecular weights of a collection of proteins [14 - 16, 18]. Characteristic mass spectra of intact biomolecules (mostly proteins) are collected from bacterial surfaces; databases are generated from known species and, using these databases for comparison, unknown bacteria can be identified. The main advantage of this approach is its simplicity and speed. Samples can be analyzed at a rate of one sample per minute, unmatched by any other technique. No sample preparation is required, and the sample requirement is very low (a few microliters containing a few thousand cells). Using this approach a one-dimensional (1-D) protein mass fingerprint can be collected from the bacterial surfaces. The information content of this 1-D picture is often sufficient to yield reliable identification for species (eg. *B. anthracis* vs. *B. cereus*), but not for strains (eg. *B. anthracis* Ames vs. *B. anthracis* Sterne). Analysis of spores is commonly viewed as difficult because of low levels of proteins markers and requirement of strong conditions to lyse spores, but these disadvantages become an asset to rapid characterization in intact spore MALDI-TOF MS because of the specificity of a small group of proteins, the small acid-soluble proteins (SASPs). In specific cases of dealing with bacillus spores, Fenselau and colleagues have reported differentiation at the strain levels [15].

The other commonly employed approach is based on proteomics involving lots of sample preparations and using MALDI-TOF for peptide sequencing. The proteins in the cell lysate are separated by two-dimensional gel electrophoresis based on molecular weight seaparation in SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) and pKa (the negative logarithm of the acid dissociation constant) separation in IEF (isoelectric focusing). Then each protein is treated with a proteolytic enzyme (e.g., trypsin) and the partially or fully digested protein is interrogated by MALDI to give a mass fingerprint of the tryptic peptides [70]. If the protein sequence of the target microbe is available, a search can be performed against the peptide map predicted by theoretical enzyme-specific digestion of the protein. This proteomic approach can increase the reliability of the identification by increasing the content of the information derived from the MALDI experiment, but it requires a known protein sequence in a public database for a successful match and also requires several time-consuming sample preparation and separation steps. This proteomic approach provides a 1-D identification of a protein by its sequence information, but a combination of protein sequences is required for positive identification of bacterial species or strains.

Our proposed approach combines protein mass fingerprinting and proteomic sequencing. This 2-D approach reports both molecular weight and sequence information of all characteristic proteins from intact bacteria, and makes two important advances: more information and minimal sample preparation. Using a MALDI TOF/TOF mass analyzer, peptide sequences could be created by fragmenting the characteristic proteins observed in the protein mass fingerprint spectrum and subsequently recording the fragment spectra. Mass spectrometric fragmentation of ions from even small intact proteins, to give partial information on amino acid sequence that can be used as an identification tag, presents a unique challenge which can be overcome by using high-energy fragmentation approaches available with TOF/TOF instrumentation. The result is a plethora of information starting with a spectrum of protein mass fingerprint accompanied by a collection of spectra with sequence information to identify the corresponding protein chosen from the fingerprint. This 2-D approach does not require the considerable sample treatment steps that are necessary with the 1-D proteomic approach with enzymatic proteolysis and protein separation, mentioned above.

Biomarker database

The 2-D fingerprint database will be built and populated for selected species (*Bacillus* species). The database will be built open-ended so that the data can be refreshed and new data added by the end-users when new microbes are mapped. The current technology based on 1-D fingerprinting is potentially capable of identifying bacteria down to the strain level. However, the reliability of this technology is often questioned even at theoretical levels. Use of advanced TOF/TOF mass spectrometry for the generation of 2-D fingerprints significantly increases the amount of information accessed and will address this issue. Another potential advantage of this approach is that the fragment spectra generated by the TOF/TOF instrument provides sequence tags that are searchable against protein/DNA databases, further enhancing reliability of identification.

The resources to build biomarker databases based on protein and peptide information have increased in recent years. Public databases and algorithms such as Blast, MSBlast, [71 - 74], Open Mass Spectrometry Search Algorithm (OMSSA) at NCBI [75], ProteinProspector [76], ExPASy proteomic server and Swissprot by the Swiss Institute of Bioinformatics (SIB) [77], and Open Proteomics Database (OPD) [78], private databases such as MAXspec and MicrobeLynx™ equipped in a M@LDI™ Linear Time of Flight Mass Spectrometer from Waters Corporation [21] or PROWL developed in collaboration between ProteoMetrics and Rockefeller University, and numerous in-house databases [24] indicate the need to expand search capabilities to accommodate the vast amount of data and information generated from peptide and protein analyses in MS. Though large numbers of protein databases are available either in the public domain or in private instrumentation, there is a requirement for a specific protein database to benefit identification and detection of biological agents, similar to those for chemical agents. A chemical agent database provided by the Organization for the Prohibition of Chemical Weapons (OPCW) is based on specific fragmentation patterns of chemical warfare agents, precursors and degradation products obtained using electron impact (EI) ionization or chemical ionization (CI) methods. At DRDC Suffield, D'Agostino et al. developed a mass spectral database of chemical warfare agents, degradation products and related compounds using liquid chromatography-electrospray ionization mass spectrometry (LC-ESI-MS) [79].

With more advance instrumentation in MS and better knowledge in bacterial protein fingerprinting in the form of MALDI-TOF/TOF MS, a mass spectral database compiled with *Bacillus* species will provide a rapid method for detection and identification of virulent *Bacillus* strains.

Future Research

The proposed research plan, described in figure 6, will assess the ability of MALDI-TOF/TOF MS to rapidly identify intact bacterial cells and to generate species- and strain-specific protein fingerprint spectra for *Bacillus anthracis*, *B. cereus*, *B. subtilis*, *B. thuringiensis*, and *B. globigii*. The database of spectra will be assembled from a selection of relevant species grown in pure culture. Chemometric (statistical and computer) approaches will be applied and evaluated to assess the performance, limits and ruggedness of the technique in providing accurate identification of bacterial species in complex test samples. Chemometrics will evaluate the procedures, and assist in extraction of valuable information, and most importantly in pattern recognition for spectral library matching and comparison

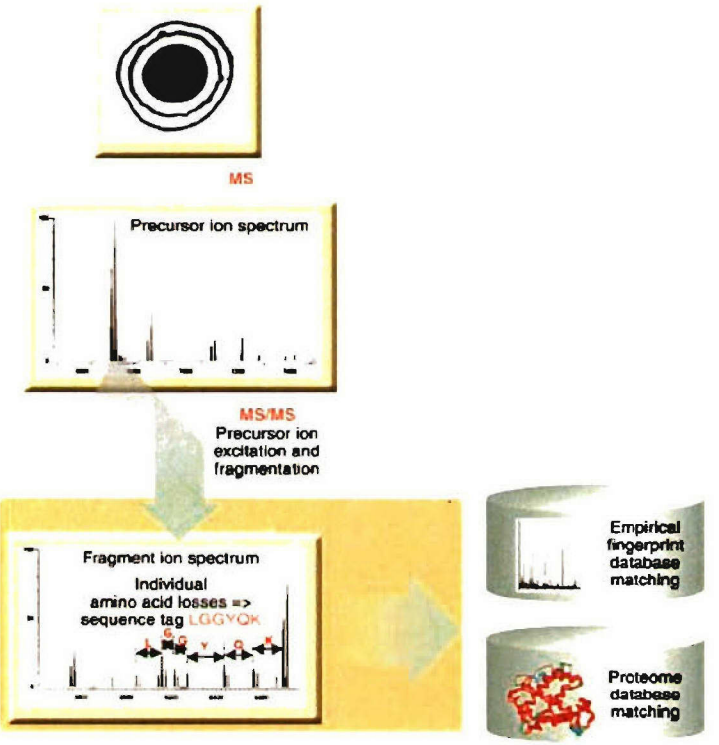

Figure 6. The scheme of a rapid MALDI-TOF/TOF MS-based microorganism identification using a combination of mass fingerprinting, peptide sequence information, and proteomic database search engine. Intact bacteria (or spores) are mixed with a matrix and then loaded onto a MALDI probe. The sample is irradiated and a precursor-ion spectrum (mass fingerprinting) is produced in the first TOF stage. Then, precursor-ion collision-induced dissociation and fragmentation result in a fragment ion spectrum for peptide sequence information. A 2-D fingerprinting database system is developed for future rapid identification.

Additionally, evaluation of various common culturing protocols for *Bacillus* species with respect to the consistency of MALDI information will be obtained. The culturing conditions are probably the single most important factor in the consistency of the mass fingerprint. Various growth media in plate and broth formulations will be evaluated. Growth stage, vegetative *vs.* spore states, will also be studied. The application of MALDI-MS to bacterial spores presents some challenges because spore coats are difficult to lyse in order to release biomarkers for MS analysis. Several techniques will be evaluated to study the release of biomarkers from spores including corona discharge, sonication, strong acid treatment.

In order to safely handle pathogenic strains it is crucial to evaluate various sterilization methods and their effect on the mass fingerprints generated by MALDI. Study of various sterilization methods with respect to the consistency of MALDI information will be obtained. Heat treatment, gamma irradiation, and various chemical treatments will be evaluated in respect to their effect on the MALDI 2D spectra. *B. anthracis* or any other virulent bacteria will be studied only after suitable sterilization methods have been developed. A quick live/dead identification using intact cell MALDI-TOF/TOF MS will prove advantages over current virulence test by growing bacteria in culture.

A mass spectral database compiled with MALDI-TOF/TOF data could be used to rapidly identify and differentiate virulent strains of *Bacillus* from non-virulent strains. Initially the database would be populated by *Bacillus* species, then it could be expanded to include other bioagents (*Brucella* species, *Clostridium botulinum*), viruses (alphaviruses) and various bacteria implicated in human infectious diseases (*Escherichia coli* O157:H7, *Salmonella* species, etc.).

Discussion & Conclusions

Traditional bacterial taxonomy classifies bacteria according to their morphological, physiological, biochemical, and chemical characteristics. This type of characterization is arduous and results are usually inconclusive. Bacterial classification is more difficult than elemental molecular anaylsis. Genetic identification methods, such as 16S and 23S ribosomal RNA sequence analysis, are commonly used to identify bacteria and are often sufficient to yield reliable identification for genus (eg. *Bacillus* vs. *Escherichia*) and species (*Bacillus anthracis* vs. *Bacillus megaterium*), but not strains (*B. anthracis* Ames vs. *B. anthracis* Sterne). Among the *Bacillus cereus* group, which consists of *B. anthracis*, *B. cereus*, and *B. thuringiensis*, the 16S and 23S rRNA sequences are 100% identical, so more rigorous molecular typing methods are required for identification. Moreover, genetic analysis does not reveal the virulence of the identified bacteria, due to the fact that the technique involves detection of nucleic acid sequences without live/dead information. Since cell functions are associated with proteins, identification of protein biomarkers which could indicate live/dead status of bacteria would be an asset.

Bacterial identification based on protein biomarkers detected by mass spectrometry is a simple procedure. This method allows direct interrogation of intact bacterial cells, thereby eliminating laborious preparation and sample processing steps. Mass spectrometry detection with a soft ionization method, such as matrix-assisted laser desorption/ionization (MALDI), has the potential to be the most generic detector because it detects everything (protein, lipids, carbohydrates, metabolites, etc.) that has a specific mass. The proposed research project is to develop a tandem MALDI-TOF method to rapidly identify bacteria and to generate a species- and strain-specific protein biomarker database. The potential for rapid analysis due to minimal sample preparation, no need for protein separation or bacterial growth for confirmation of viability would prove advantageous over current methods.

A mass spectral database compiled with MALDI-TOF/TOF data could be used to rapidly identify and differentiate virulent strains of *Bacillus* from non-virulent strains. Specific data relating to protein biomarkers and live/dead identification in MS would be included in this in-house database. Initially the database would be populated by *Bacillus cereus* group, then the project could be expanded to include other bioagents (*Brucella* species, *Clostridium botulinum*), viruses (alphaviruses) and various bacteria (*Escherichia coli* O157:H7, *Salmonella* species, etc.) implicated in human infectious diseases.

References

1. Ezaki, T. (1994). Genetic identification of bacterial pathogens. *Nippon rinsho. Japanese Journal of Clinical Medicine*, 52, 355-358.

2. Heller, D.N., Fenselau, C., Cotter, R.J., Demirev, P., Olthoff, J.K., Honovich, J., Uy, M., Tanaka, T., and Kishimoto, Y. (1987). Mass spectral analysis of complex lipids desorbed directly from lyophilized membranes and cells. *Biochemical and Biophysical Research Communications*, 142, 194-199.

3. Heller, D.N., Cotter, R.J., Fenselau, C., and Uy, O.M. (1987). Profiling of bacteria by fast atom bombardment mass spectrometry. *Analytical Chemistry*, 59, 2806-2809.

4. van Baar, B.L.M. (2000). Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiology Reviews*, 24, 193-219.

5. Lay Jr., J.O. (2001). MALDI-TOF mass spectrometry of bacteria. *Mass Spectrometry Reviews*, 20, 172-194.

6. Kim, Y. J., Freas, A., and Fenselau, C. (2001). Analysis of viral glycoproteins by MALDI-TOF mass spectrometry. *Analytical Chemistry*, 73, 1544-1548.

7. Hettick, J.M., Kashon, M.L., Simpson, J.P., Siegel, P.D., Mazurek, G.H., and Weissman, D.N. (2004). Proteomic profiling of intact mycobacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analytical Chemistry*, 76, 5769-5776.

8. Jones, J.J., Stump, M.J., Fleming, R.C., Lay, J.O., and Wilkins, C.L. (2003). Investigation of MALDI-TOF and FT-MS techniques for analysis of *Escherichia coli* whole cells. *Analytical Chemistry*, 75, 1340-1347.

9. Silipo, A., Lanzetta, R., Amoresano, A., Parrilli, M., and Molinaro, A. (2002). Ammonium hydroxide hydrolysis: a valuable support in the MALDI-TOF mass spectrometry analysis of Lipid A fatty acid distribution. *Journal of Lipid research*, 43, 2188-2195.

10. Bundy, J.L, and Fenselau, C. (2001). Lectin and carbohydrate affinity capture surfaces for mass spectrometric analysis of microorganisms. *Analytical Chemistry*, 73, 751-757.

11. English, R.D., Warscheid, B., Fenselau, C., and Cotter, R.J. (2003). *Bacillus* spore identification via proteolytic peptide mapping with a miniaturized MALDI TOF mass spectrometer. *Analytical Chemistry*, 75, 6886-6893.

12. Warscheid, B., Jackson, K., Sutton, C., and Fenselau, C. (2003). MALDI analysis of *Bacilli* in spore mixtures by applying a quadrupole ion trap time-of-flight tandem mass spectrometer. *Analytical Chemistry*, 75, 5608-5617.

13. Warscheid, B., and Fenselau, C. (2003). Characterization of *Bacillus* spore species and their mixtures using postsource decay with a curved-field reflectron. *Analytical Chemistry*, 75, 5618-5627.

14. Hathout, Y., Demirev, P.A., Ho, Y., Bundy, J.L., Ryzhov, V., Sapp, L., Stutler, J., Jackman, J., and Fenselau, C. (1999). Identification of *Bacillus* spores by matrix-assisted laser desorption ionization-mass spectrometry. *Applied and Environmental Microbiology*, 65, 4313-4319.

15. Ryzhov, V., Hathout, Y., and Fenselau, C. (2000). Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. *Applied and Environmental Microbiology*, 66, 3828-3834.

16. Hathout, Y., Setlow, B., Cabrera-Martinez, R., Fenselau, C., and Setlow, P. (2003). Small, acid-soluble proteins as biomarkers in mass spectrometry analysis of *Bacillus* spores. *Applied and Environmental Microbiology*, 69, 1100-1107.

17. Pribil, P.A., Patton, E., Black, G., Doroshenko, V., and Fenselau, C. (2005). Rapid characterization of *Bacillus* spores targeting species-unique peptides produced with an atomospheric pressure matrix-assisted laser desorption/ionization source. *Journal of Mass Spectrometry*, 40, 464-474.

18. Ryzhov, V., and Fenselau, C. (2001). Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Analytical Chemistry*, 73, 746-750.

19. Dieckmann, R., Graeber, I., Kaesler, I., Szewzyk, U., and von Döhren, H. (2005). Rapid screening and dereplication of bacterial isolates from marine sponges of the Sula Ridge by intact-cell-MALDI-TOF mass spectrometry (ICM-MS). *Applied Microbiology and Biotechnology*, 67, 539-548.

20. Claydon, M.A., Davey, S.N., Edwards-Jones, V., and Gordon, D.B. (1996). The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotechnology*, 14, 1584-1586.

21. Keys, C.J., Dare, D.J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M., and Shah, H.N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterization of bacteria implicated in human infectious diseases. *Infection, Genetics and Evolution*, 4, 221-242.

22. Wang, Z., Dunlop K., Long, S.R., and Li, L. (2002). Mass spectrometric methods for generation of protein mass database used for bacterial identification. *Analytical Chemistry*, 74, 3174-3182.

23. Tao, L., Yu, X., Snyder, A.P., and Li, L. (2004). Bacterial identification by protein mass mapping combined with an experimentally derived protein mass database. *Analytical Chemistry*, 76, 6609-6617.

24. Dworzanski, J. P., and Snyder, A.P. (2001). Classification and identification of bacteria using mass spectrometry-based proteomics. *Expert Review of Proteomics*, 2, 863-878.

25. Arnold, R.J., and Reilly, J.P. (1998). Fingerprint matching of *E. coli* strains with matrix-assisted laser desportion/ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. *Rapid Communications in Mass Spectrometry*, 12, 630-636.

26. Demirev, P.A., Lin, J.S., Pineda, F.J., and Fenselau, C. (2001). Bioinformatics and mass spectrometry for microorganism identification: proteome-wide post-translational modifications and database search algorithms for characterization of intact *H. pylori*. *Analytical Chemistry*, 73, 4566-4573.

27. Demirev, P.A., Feldman, A.B., and Lin, J.S. (2004). Bioinformatics-based strategies for rapid microorganism identification by mass spectrometry. *Johns Hopkins APL Technical Digest*, 25, 27-37.

28. Ochoa, M.L., and Harrington, P.B. (2005). Immunomagnetic isolation of enterohemorragic *Escherichia coli* O157:H7 from ground beef and identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and database searches. *Analytical Chemistry*, 77, 5258-5267.

29. Cain, T.C., Lubman, D.M., Weber Jr., W.J., and Vertes, A. (2005). Differentiation of bacteria using protein profiles from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 8, 1026-1030.

30. Yao, Z., Demirev, P.A., and Fenselau, C. (2002), Mass spectrometry-based proteolytic mapping for rapid virus identification. *Analytical Chemistry*, 74, 2529-2534.

31. Whiteaker, J.R., Fenselau, C, Fetterolf, D., Steele, D., and Wilson, D. (2004). Quantitative determination of heme for forensic characterization of *Bacillus* spores using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analytical Chemistry*, 76, 2836-2841.

32. Cottingham, K. (2006). MS on the bioterror front lines. *Analytical Chemistry*, 78, 18A-23A.

33. Holt, J.G. (1986). Bergey's Manual of Systematic Bacteriology, vol. 2. Gram-positive bacteria other than Actinomycetes. 1st ed. Baltimore: Williams & Wilkins, section 13, p 1104-1139.

34. Claus, D., and Berkeley, R.C.W. (1986) Genus Bacillus Cohn 1872, 174^{AL}. In Sneath, P.H.A., Mair, N.S., Sharpe, M.E., Holt, J.G. (Eds.), *Bergey's Manual of Systematic Bacteriology*. Vol. 2., pp.1105-1139. Williams & Wilkins, Baltimore.

35. Turnbull, P.C.B. (1999). Definitive identification of *Bacillus anthracis* – a review. *Journal of Applied Microbiology*, 87, 237-240.

36. Ash, C., Farrow, J.A.E., Wallbanks, S., and Collins, M.D. (1991). Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. *Letters in Applied Microbiology*, 13, 202-206.

37. Priest, F.G. (1993). Systematics and ecology of *Bacillus*. In Sonenshein, A.L., Hoch, J.A., and Losick, R. (Eds.), *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Biology*, pp.3-16. American Society for Microbiology, Washington, DC.

38. Friedlander, A.M. (1997). Anthrax. In Sidell, F.R., Takafuji, E.T., and Franz, D.R. (Eds.), *Medical Aspects of Chemical and Biological Warfare*, pp. 467-478. Washington, DC, USA: TMM Publications.

39. Merriam-Webster Online Dictionary. (Online) Merriam-Webster Inc. http://www.merriam-webster.com/ (31 Jan. 2006).

40. Papaparaskevas, J, Houhoula, D.P., Papadimitriou, M., Saroglou, G., Legakis, N.J., and Zerba, L. (2004). Ruling out *Bacillus anthracis*. *Emerging Infectious Diseases*, 10, 732-735.

41. Dragon, D.C., Bader, D.E., Mitchell, J., and Woollen, N. (2005). Natural dissemination of *Bacillus anthracis* spores in northern Canada. *Applied and Environmental Microbiology,* 71, 1610-1615.

42. Norris, J.R., Berkeley, R.C., Logan, N.A., and O'Donnell, A.G. (1981). The Genera *Bacillus* and *Sporolactobacillus*. In Starr, M.P., Stolp, H., Trüper, H.G., Balows, A., and Schlegel, H.G. (Eds), *The Prokaryotes. A Hanbook on Habitats, Isolation and Identification*, pp.1711-1742. Springer-Verlag, Berlin.

43. Gordon, R.E., Haynes, W.C., and Pang, C.H. (1973) The genus Bacillus. *U.S. Department of Agriculture Agricultural Handbook no. 427.* U.S. Department of Agriculture, Washington DC.

44. Ash, C., Farrow, J.A., Dorsch, M., Stackebrandt, E., and Collins, M.D. (1991). Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequenceing of 16S rRNA. *International Journal of Systematic Bacteriology*, 41, 343-346.

45. Carlson, C.R., Caugant, D., and Kolsto, A.-B. (1994). Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. *Applied Environmental Microbiology*, 60, 1719-1725.

46. Helgason, E., Økstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., and Kolstø, A.-B. (2000). *Bacillus anthracis*, *Bacillus cereus*, *Bacillus thuringiensis* – one species on the basis of genetic evidence. *Applied and Environmental Microbiology,* 66, 2627-2630.

47. Jensen, G.B., Hansen, B.M., Eilenberg, J., and Mahillon, J. (2003). The hidden lifestyles of *Bacillus cereus* and relatives. *Environmental Microbiology*, 5, 631.

48. Beyer, W., Pocivalsek, S., and Böhm, R. (1999). Polymerase chain reaction-ELISA to detect *Bacillus anthracis* from soil samples – limitations of present published primers. *Journal of Applied Microbiology,* 87, 229-236.

49. Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R., Holtzapple, E.K., Økstad, O.A., Helgason, E., Rilstone, J., Wu, M., Kolonay, J.F., Beanan, M.J., Dodson, R.J., Brinkac, L.M., Gwinn, M., Deboy, R.T., Madpu, R., Daugherty, S.C., Durkin, A.S., Haft, D.H., Nelson, W.C., Peterson, J.D., Pop, M., Khouri, H.M., Radune, D., Benton, J.L., Mahamoud, Y., Jiang, L., Hance, I.R., Weidman, J.F., Berry, K.J., Plaut, R.D., Wolf, A.M., Watkins, K.L., Nierman, W.C., Hazen, A., Cline, R., Redmond, C., Thwaite, J.W., White, O., Salzberg, S.L., Thomason, B., Friedlander, A.M., Koehler, T.M., Hanna, P.C., Kolstø, A., and Fraser, C.M. (2003). The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature*, 423, 81-86.

50. Rasko, D.A., Ravel, J.R., Økstad, O.A., Helgason, E., Cer, R.Z., Jiang, L., Shores, K.A., Fouts, D.E., Tourasse, N.J., Angiuoli, S.V., Kolonay, J., Nelson, W.C., Kolstø, A., Fraser, C.M., and Read, T.D. (2004). The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nucleic Acids Research*, 32, 977-988.

51. Price, L.B., Hugh-Jones, M., Jackson, P.J., and Keim, P. (1999). Genetic diversity in the protective antigen gene of *Bacillus anthracis*. *Journal of Bacteriology*, 181, 2358-2362.

52. Bavykin, S.G., Lysov, Y.P, Zakhariev, V., Kelly, J.J., Jackman, J., Stahl, D.A., and Cherni, A. (2004). Use of 16S rRNA, 23S rRNA, and *gyr*B gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *Journal of Cminical Microbiology*, 42, 3711-3730.

53. Sacchi, C.T., Whitney, A.M., Mayer, L.W., Morey, R., Steigerwalt, A., Boras, A., Weyant, R.S., and Popovic, T. (2002). Sequencing of 16S rRNA gene: a rapid tool for identification of *Bacillus anthracis*. *Emerging Infectious Diseases*, 8, 1117-1123.

54. Harrell, L.J., Anderson, G.L., and Wilson, K.H. (1999). Genetic variability of *Bacillus anthracis* and related species. *Journal of Clinical Microbiology*, 33, 1847-1850.

55. Henderson, E., Yu, D., and Turnbull, P.C. (1995). Differentiation of *Bacillus anthracis* and other '*Bacillus cereus* group' bacteria using IS231-derived sequences. *FEMS Microbiology Letters*, 128, 113-118.

56. Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P.J., and Hugh-Jones, M.E. (2000). Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *Journal of Bacteriology*, 182, 2928-2936.

57. Vos, P., Hogers, R., Bleeker, M., Reijans, M., van der Lee, T., Hornes, M, Friters, A., Pot, J., Paleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research*, 23, 4407-4424.

58. Ryu, C., Lee, K., Hawng, H., Yoo, C., Seong, W., and Oh, H. (2005). Molecular Characterization of Korean *Bacillus anthracis* isolates by amplified fragment length polymorphism analysis and multilocus variable-number tandem repeat analysis. *Applied and Environmental Microbiology*, 71, 4664-4671.

59. Hoffmaster, A.R., Fitzgerald, C.C., Ribot, E., Mayer, L.W., and Popovic, T. (2002). Molecular subtyping of *Bacillus anthracis* and the 2001 bioterrorism-associated anthrax outbreak, United States. *Emerging Infectious Diseases,* 8, 1111-1116.

60. Lindstedt, B. (2005). Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis*, 26, 2567-2582.

61. Cheung, D.T.L., Kam, K.M., Hau, K.L., Au, T.K., Marston, C.K., Gee, J.E., Popovic, T., van Ert, M.N., Kenetic, L., Keim, P., and Hoffmaster, A.R. (2005). Characterization of a *Bacillus anthracis* isolate causing a rare case of fatal anthrax in a 2-year-old boy from Hong Kong. *Journal of Clinical Microbiology*, 43, 1992-1994.

62. Smith, K.L., de Vos, V., Bryden, H.B., Hugh-Jones, M.E., Klevytska, A., Price, L.B., Keim, P., and Scholl, D.T. (1999). Meso-scale exology of anthrax in southern Africa: a pilot study of diversity and clustering. *Journal of Applied Microbiology*, 87, 204-207.

63. Selander, R. K., and Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N., and Whittam, T.S. (1986). Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* for bacterial population genetics and systematics. *Applied Environmental Microbiology*, 51, 873-884.

64. Helgason, E., Tourasse, N.J., Meisal, R., Caugant, D.A., and Kolstø, A. (2004). Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. *Applied and Environmental Microbiology*, 70, 191-201.

65. Karas, M., Bachmann, D., and Hillenkamp, F. (1985). Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry*, 57, 2935-2939.

66. Karas, M., and Hillenkamp, F. (1988). Laser Desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, 60, 2299-2301.

67. Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988). Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2, 151-153.

68. Medzihradszky, K.F., Campbell, J.M., Baldwin, M.A., Falick, A.M., Juhasz, P., Vestal, M.L., and Burlingame, A.L. (2000). The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical Chemistry*, 72, 552-558.

69. Biemann, K. (1992). Mass spectrometry of peptides and proteins. *Annual Review of Biochemistry*, 61, 977-1010.

70. Henzel, W.J., Watanabe, C., and Stults, J.T. (2003). Protein identification: the origins of peptide mass fingerprinting. *Journal of the American Society for Mass Spectrometry*, 14, 931-942.

71. Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A., and Wagner, L. (2003). Database resources of the national center for biotechnology. *Nucleic Acids Research*, 31, 28-33.

72. Habermann, B., Oegema, J., Sunyaev, S., Shevchenko, A. (2004). The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. *Molecular and Cellular Proteomics*, 3, 238-249.

73. Shevchenko, A, Sunyaev, S., Liska, A., Bork, P. (2002). Nanoelectrospray tendem mass spectrometry and sequence similarity searching for identification of proteins from organisms with unknown genomes. *Methods in Molecular Biology*, 211, 221-234.

74. Zhong, H., and Li, L. (2005). An algorithm for interpretation of lonw-energy collision-induced dissociation product ion spectra for *de novo* sequencing of peptides. *Rapid Communications in Mass Spectrometry*, 19, 1084-1096.

75. Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W., and Bryant, S.H. (2004). Open mass spectrometry search algorithm. *Journal of Proteome Research*, 3, 958-964.

76. Clauser K.R., Baker P.R. and Burlingame A.L. (1999). Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical Chemistry*, 71, 2871-2882.

77. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 21, 3784-3788.

78. Prince, J.T., Carlson, M.W., Wang, R., Lu, P., and Marcotte, E.M. (2004). The need for a public proteomics repository. *Nature Biotechnology*, 22, 471-472.

79. D'Agnostino, P., Chenier, C., and Hancock, J.R. (2002). Electrospray mass spectrometry of chemical warfare agents, degradation products and related compounds. (DRDC Suffield TR 2002-028). Defence R&D Canada Suffield.

List of symbols/abbreviations/acronyms/initialisms

AFLP	Amplified fragment length polymorphism analysis
Ar	Argon gas
B.	*Bacillus*
CB	Chemical Biological
CF	Canadian Forces
CHCA	α-cyano-hydroxycinnamic acid
CI	Chemical ionization
CID	Collision-induced dissociation
CO_2	Carbon dioxide
DHB	Dihydroxybenzoic acid
DNA	Deoxy-nucleic acid
DND	Department of National Defence
DRDC	Defence Research & Development Canada
ELISA	Enzyme-linked immunosorbent assay
EI	Electron impact
ESI	Electrospray ionization
ExPASy	Expert Protein Analysis System
G + C	Guanine and cytosine
He	Helium gas
kDa	Kilodalton
kV	kilovolt

L	Length of drift tube in a mass spectrometer
LC	Liquid chromatography
m	Mass of an ion
M^+	Molecular ion
m/z	Mass-to-charge
MALDI	Matrix-assisted laser desorption/ionization
MCP	Multichannel plate
MEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
MLVA	Multilocus variable-number tandem repeat (VNTR) analysis
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
OMSSA	Open Mass Spectrometry Search Algorithm
OPCW	The Organization for the Prohibition of Chemical Weapons
OPD	Open Proteomics Database
PA	Protective antigent
PCR	Polymerase chain reaction
PLET	Polymyxin, lysozyme, EDTA, thallium acetate medium
rRNA	Ribosomal RNA
SA	Sinapinic acid
SASPs	Small, acid-soluble proteins
SDS-PAGE	Sodium dodecyl (lauryl) sulphate – polyacrylamide gel electrophoresis
SIB	Swiss Institute of Bioinformatics
spp.	Species

t	Flight time of an ion
TOF	Time-of-flight
TOF/TOF	Tandem time-of-flight mass analyzers
UV	ultraviolet
V	voltage
VNTRs	Variable-number tandem repeats
Xe	Xenon gas
z	Charge of an ion

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for who the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in Section 8.)

   Defence R&D Canada – Suffield
   PO Box 4000, Station Main
   Medicine Hat, AB T1A 8K6

2. SECURITY CLASSIFICATION
   (overall security classification of the document, including special warning terms if applicable)

   Unclassified

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title).

   Rapid Detection & Identification of *Bacillus* Species using MALDI-TOF/TOF and Biomarker Database (U)

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

   Chan, Nora. W.C.; Lee, William E.; and Mester, Zoltán

5. DATE OF PUBLICATION (month and year of publication of document)

   June 2006

6a. NO. OF PAGES (total containing information, include Annexes, Appendices, etc) 36

6b. NO. OF REFS (total cited in document) 79

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

   Technical Memorandum

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

   Defence R&D Canada – Suffield

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

   DRDC Suffield TM 2006-078

10b. OTHER DOCUMENT NOs. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

   (x) Unlimited distribution
   ( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
   ( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
   ( ) Distribution limited to government departments and agencies; further distribution only as approved
   ( ) Distribution limited to defence departments; further distribution only as approved
   ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally corresponded to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected).

   Unlimited

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C) or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

A strategic plan to develop a MALDI-TOF/TOF tandem MS method in conjunction with a forensic biomarker database for rapid microbial identification is described in this technical memorandum. The objective is to develop rapid, highly sensitive and definitive protocols for detection and identification of BW agents such as *Bacillus* species and demonstrate their effectiveness. These protocols will be based on matrix-assisted laser desorption/ionization with time-of-flight (MALDI-TOF) mass spectrometry (MS) technology applicable to intact microorganisms. MALDI can directly analyze biological samples for biomarkers (proteins, lipids, etc.). These biomarkers can be used to identify microorganisms at species and strain levels. In addition to method development, a new forensic mass spectral biomarker database could be created and populated for selected species and strains, starting with selected *Bacillus* species. This technology would create a new area for comprehensive forensic analysis of biological materials that are not addressed by current methods (PCR and DNA sequencing, and immunoassays). This memorandum contains a review of *Bacillus* species and current detection and identification methods, and a review of MALDI-TOF/TOF MS methods and biomarker databases.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifies, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Biological detection & Identification, *Bacillus* species, mass spectrometry (MS), matrix-assisted laser desorption/ionization (MALDI), protein biomarker database.

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

www.drdc-rddc.gc.ca