Plamen Demirev
Todd R. Sandrin *Editors*

Applications of Mass Spectrometry in Microbiology

From Strain Characterization to Rapid Screening for Antibiotic Resistance

Springer

Applications of Mass Spectrometry in Microbiology

Plamen Demirev • Todd R. Sandrin
Editors

Applications of Mass Spectrometry in Microbiology

From Strain Characterization to Rapid Screening for Antibiotic Resistance

Springer

Editors
Plamen Demirev
Applied Physics Laboratory
Johns Hopkins University
Laurel, MD
USA

Todd R. Sandrin
School of Mathematical and Natural Sciences
New College of Interdisciplinary
Arts & Sciences
Arizona State University
Phoenix, AZ
USA

ISBN 978-3-319-26068-6
ISBN 978-3-319-26070-9 (eBook)
DOI 10.1007/978-3-319-26070-9

Library of Congress Control Number: 2015959597

Springer
Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Contents

List of Contributors

Denis Andrzejewski Spectroscopy and Mass Spectrometry Branch, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, College Park, MD, USA

Franco Basile Department of Chemistry, Biodetection and Mass Spectrometry Laboratory, University of Wyoming, Laramie, WY, USA

René Brunisholz Functional Genomics Center Zurich, ETH Zurich and University of Zurich, Zurich, Switzerland

John H. Callahan Spectroscopy and Mass Spectrometry Branch, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, College Park, MD, USA

Plamen Demirev Applied Physics Laboratory Johns Hopkins University, Laurel, MD, USA

Shobha Devi Department of Chemistry, National Dong Hwa University, Hualien, Taiwan

David Drissner Agroscope Institute for Food Sciences, Food Microbiology and Analytics, Waedenswil, Switzerland

Jacek P. Dworzanski Leidos, Inc., Bel Air, MD, USA

Catherine Fenselau Department of Chemistry & Biochemistry, University of Maryland, College Park, MD, USA

Maria-Theresia Gekenidis Agroscope Institute for Food Sciences, Food Microbiology and Analytics, Waedenswil, Switzerland

Vladimír Havlicek RCPTM, Department of Analytical Chemistry, Palacky University, Olomouc, Czech Republic

Yen-Peng Ho Department of Chemistry, National Dong Hwa University, Hualien, Taiwan

Jaroslav Hrabak Biomedical Center and Department of Microbiology, Charles University in Prague, Plzen, Czech Republic

Anren Hu Institute of Medical Biotechnology, Tsu-Chi University, Hualien, Taiwan

Daniela Jacob ZBS 2, Highly Pathogenic Microorganisms, Robert Koch-Institute, Berlin, Germany

Silke R. Klee ZBS 2, Highly Pathogenic Microorganisms, Robert Koch-Institute, Berlin, Germany

Peter Lasch Proteomics and Spectroscopy at the Centre for Biological Threats and Special Pathogens (ZBS6), Robert Koch-Institute, Berlin, Germany

Melinda A. McFarland Spectroscopy and Mass Spectrometry Branch, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, College Park, MD, USA

Rudolph K. Mignon Department of Chemistry, Biodetection and Mass Spectrometry Laboratory, University of Wyoming, Laramie, WY, USA

Costas C. Papagiannitsis Biomedical Center and Department of Microbiology, Charles University in Prague, Plzen, Czech Republic

Todd R. Sandrin School of Mathematical and Natural Sciences, New College of Interdisciplinary Arts & Sciences, Arizona State University, Phoenix, AZ, USA

Ralph Schlapbach Functional Genomics Center Zurich, ETH Zurich and University of Zurich, Zurich, Switzerland

Ondrej Šedo Research Group Proteomics, CEITEC—Central European Institute of Technology, Masaryk University, Brno, Czech Republic

Freek Spitaels Laboratory of Microbiology, Faculty of Sciences, Ghent University, Ghent, Belgium

Peter Vandamme Laboratory of Microbiology, Faculty of Sciences, Ghent University, Ghent, Belgium

Guido Werner FG 13, Nosocomial Pathogens and Antibiotic Resistances, Robert Koch-Institute, Wernigerode, Germany

Anneleen D. Wieme Laboratory of Microbiology, Faculty of Sciences, Ghent University, Ghent, Belgium

Laboratory of Biochemistry and Brewing, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

Zbyněk Zdráhal Research Group Proteomics, CEITEC—Central European Institute of Technology, Masaryk University, Brno, Czech Republic

Lin Zhang School of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ, USA

Chapter 1
Introduction

Plamen Demirev and Todd R. Sandrin

Mass Spectrometry and Microbiology

Mass spectrometry (MS) is a physical method for analysis introduced more than 100 years ago. During that period, MS applications have successfully proliferated in almost all areas of science and technology—from early studies of the structure of atoms and molecules culminating with the discovery of isotopes to characterization of planetary atmospheres and surfaces and search for extraterrestrial life. MS is an indispensable tool in organic chemistry and biochemistry for structural elucidation of various classes of natural products and synthetic compounds. In the last quarter century, advances in MS methods and instrumentation have been at the forefront of efforts to map complex biological systems, including the human metabolome, proteome, and microbiome.

MS was first successfully applied to analysis of intact microorganisms more than 40 years ago (Anhalt and Fenselau 1975). These efforts have expanded and have been particularly significant after the introduction of the soft ionization MS techniques—matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) (Fenn et al. 1989; Tanaka 2003; Karas and Hillenkamp 1988). Both techniques (recognized by the Nobel Prize in Chemistry in 2002) allow the ionization and transfer into vacuum of large, intact, nonvolatile biomolecules, such as proteins. Various types of mass analyzers—quadrupole, ion trap, time-of-flight (TOF)—have been coupled to both MALDI and ESI ion sources, allowing multiple stages (tandem) MS to be performed for structure elucidation of analytes of interest. All these instrumental developments have allowed MS to become a well-established

P. Demirev (✉)
Applied Physics Laboratory Johns Hopkins University Laurel, MD, USA
e-mail: plamen.demirev@jhuapl.edu

T. R. Sandrin
School of Mathematical and Natural Sciences, New College of Interdisciplinary Arts & Sciences, Arizona State University, Phoenix, AZ, USA
e-mail: Todd.Sandrin@asu.edu

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_1

method for microorganism characterization. MS has demonstrated considerable advantage as a rapid, precise, and cost-effective method for identification, compared to conventional phenotypic techniques. The method is ultimately based on detection of organism-specific "fingerprints" (or "signatures", i.e., biomarker molecules, from either intact and/or lysed cells (Fenselau and Demirev 2001; Wilkins et al. 2005; Demirev and Fenselau 2008a, 2008b; Seng et al. 2009; Freivald and Sauer 2009; Shah and Gharbia 2010; Ho and Reddy 2010; Bizzini and Greub 2010; Sauer and Kliem 2010; Cliff et al. 2011; Welker 2011; Fenselau and Demirev 2011; Croxatto et al. 2012; Havlicek et al. 2013; Sandrin et al. 2013; DeMarco and Ford 2013; Fenselau 2013; Clark et al. 2013; Fagerquist 2013; Calderaro et al. 2014)). Different organisms exhibit different MS signatures allowing differentiation between organisms to be made. Examples of microorganism-specific biomarkers include highly expressed intact proteins, their proteolytic products, nonribosomal peptides, polar and nonpolar lipids, RNA, and DNA. Sequence/structure-specific fragments for biomarker identification are generated by tandem MS. In top-down proteomics, these biomarkers are intact proteins, while proteolytic peptides (obtained after enzymatic or chemical hydrolysis) are mapped to their precursor proteins in bottom-up/middle-down approaches. Ultimately microorganism identification relies on mapping between spectra of unknowns with signatures of known microorganisms in MS signature libraries. Such libraries are compiled either by experimentally acquiring mass spectra of reference organisms and/or by generating in silico signatures from information in genomic or proteomic databases (Pineda et al. 2000; Demirev et al. 2004).

Thousands of reports on applications of MS for microorganism characterization in research, clinical microbiology, counter-bioterrorism, food safety, environmental monitoring, and quality have been published (Havlicek et al. 2013). Regulatory bodies in Europe, the US (FDA), and elsewhere have approved MS-based assays for infectious disease diagnostics. As of mid-2015, more than 3300 commercial MALDI TOF MS systems have been deployed worldwide in hospitals and clinical laboratories. As interest has increased in this technology, the pace of discovery and development of new applications has accelerated. The technology has been shown repeatedly to be effective at rapidly discriminating, identifying, and characterizing microorganisms at the species level and above. Some of the most promising yet challenging applications of this technology require microorganism characterization at the subspecies and strain levels. Categorization of strains sharing similar traits, differentiation of closely related strains, and/or identification of a single strain by MS techniques is desired. For example, there is tremendous need in expanding this approach to rapidly identify strains of antibiotic-resistant microorganisms.

Chapters Included in This Book

While previous work has covered broader approaches to using MS to characterize microorganisms at the species level or above, this book focuses on strain-level and subtyping applications. Innovators, leaders, and practitioners in the field from

around the world have contributed to this comprehensive overview of current and next-generation approaches for MS-based microbial characterization at the subspecies and strain levels. Research and developments into novel MS-based assays for antibiotic resistance determination are reviewed as well.

As an introduction to the field, Basile and Mignon present in Chap. 2 a general overview of MS ionization techniques, instrumentation, and methodology currently used for the analysis of closely related bacteria. Specific properties and parameters of the types of mass analyzers used in modern MS are listed. Important factors determining the specificity in target microorganism identification and the ability to differentiate among closely related microorganisms (i.e., selectivity) are discussed in the context of strain differentiation and antibiotic resistance determination.

Sample preparation is arguably one of the most crucial steps in efforts to identify microorganisms by MS. In Chap. 3, Ho and coworkers review different sample preparation steps currently used in the context of rapid MS analysis of microorganisms. Approaches that might eliminate the need for culturing of the target organism (currently, the key rate-limiting step), while maximizing biosafety to obtain detectable signals, are emphasized. These include protocols for intact microbial cell and/or biomarker enrichment through various affinity techniques as well as cell lysis combined with biomarker solubilization. Separation techniques (e.g., liquid chromatography) may facilitate more accurate and efficient identification of strain-specific biomolecules in microbial mixtures or complex biological samples. Since MALDI-MS is the method of choice for the rapid identification of microorganisms, a discussion on the selection of MALDI matrices and matrix solvents is included as well.

In Chap. 4, Fenselau, a pioneer in the application of MS to microbiology, stresses the overriding importance of modern proteomics and bioinformatics tools in MS approaches for microorganism identification of bacteria. Utilizing genomic database information is usually faster, more efficient, and more reliable than matching to a library of experimentally collected spectra alone. Identifications can be made without controlling sample preparation or instrumental conditions, e.g., ionization. In addition, specific biomarkers can be identified for strain identification and forensic science applications. The advantages of these proteomic strategies are illustrated in the analysis of components in mixtures, genetic engineering in bacteria, and bacteria with unsequenced genomes.

Dworzanski provides in Chap. 5 an extensive overview of bottom-up shotgun proteomics for MS-based microorganism characterization. This peptide-centric technique matches product ion mass spectra of tryptic peptides against a comprehensive database of protein sequences translated from protein-encoding open reading frames found in bacterial genomes. Phylogenomic profiles of sequenced peptides are then analyzed using numerical taxonomy tools to reveal strain identities up to the subspecies level. Bottom-up proteomics also allows sequence-based subtyping of microbial strains based on identification of proteins associated with virulence, antibiotic resistance, or used in other serotyping methods.

Methods to enhance the taxonomic resolution of MALDI TOF MS to characterize bacteria to the subspecies and strain levels are reviewed in Chap. 6 by Zhang and Sandrin. They focus on several experimental factors that will improve strain-level

characterization efforts. These factors include culture medium, sample preparation, data acquisition, and data analysis. Specific examples illustrating both successes and challenges of this approach are presented.

Sedo and Zdrahal provide in Chap. 7 specific examples of MALDI TOF MS profiling for successful differentiation between strains of the *Lactobacillus acidophilus* group and selected *Mycobacterium* spp. In these two examples, careful optimization of the culture protocols contributed to the method robustness. In addition, strains within the *Acinetobacter calcoaceticus–Acinetobacter baumannii* complex, *Staphylococcus aureus,* and *Bacillus subtilis* ecotypes can be successfully typed by utilizing two alternative sample preparation protocols: alternative MALDI matrix solution or microwave-assisted tryptic digestion of the intact cells.

Lasch and coworkers describe in Chap. 8 their group's efforts to improve taxonomic resolution without compromising the simplicity and the speed of MALDI TOF MS. Such improvements may be achieved by signature database expansion with novel and diverse strains, optimization, and standardization of sample preparation and data-acquisition protocols. Further enhancement in data analysis pipelines including more advanced spectral preprocessing, feature selection, and supervised methods of multivariate classification analysis also contribute to taxonomic resolution enhancements. Strains of *Staphylococcus aureus, Enterococcus faecium,* and *Bacillus cereus* are selected to illustrate aspects of that strategy.

Efficient methods based on MALDI TOF MS to dereplicate (i.e., group together) bacterial isolates with highly similar properties have been developed and are discussed by Vandamme and coworkers in Chap. 9. The high throughput capability and low running costs for dereplication by MALDI TOF MS allow direct microorganism identification at the species level in a large number of samples and obviate the need for more labor-intensive characterization. While isolates cultured in different media under varying conditions can be identified at the species level, isolates from the same species should be carefully re-grown in standardized conditions in order to eventually select individual peaks as strain-specific markers.

In Chap. 10, McFarlane et al. utilize liquid chromatography (LC)-MS to generate intact protein expression profiles as a snapshot of expressed proteins in a wide range of bacterial samples. Subsequent top-down proteomic analysis by LC-tandem MS allows identification of expressed serovar-specific proteins, resulting from nonsynonymous single-nucleotide polymorphisms (SNPs). Closely related, unsequenced or bacterial strains with newly acquired SNPs and plasmid proteins can be successfully differentiated by this multiplexed approach.

In Chap. 11, Drissner and coworkers provide an overview of MALDI TOF and off-line LC MALDI TOF/TOF (tandem MS) methods for typing applications. They describe further a rapid procedure for tryptic peptide generation from a simple whole-cell extract. Within minutes and without the need for further sample processing they are able to differentiate each of three different *Salmonella enterica* subspecies based on the detection of strain-specific peptide biomarkers.

Drug-resistant strains of pathogenic organisms are some of the most persistent and difficult to eradicate clinical infections, substantially increasing patient mortality as well as healthcare costs. Novel MALDI TOF MS methods for fast and reli-

able detection of the presence of β-lactamases in drug-resistant bacterial strains are discussed by Hrabak et al. in Chap. 12. One method involves direct detection of β-lactam hydrolysis by monitoring the molecular mass of carbapenem antibiotics. Software tools for spectral interpretation to discern drug hydrolysis will allow assay automation and high throughput. Direct detection of β-lactamases (an enzyme with a molecular weight (MW) of around 29 kDa) by MALDI TOF MS (e.g., in clinical isolates of Enterobacteriaceae) provides a complementary tool for establishing drug resistance.

Functional assays that involve the combination of MS and stable-isotope labeling for establishing drug resistance are reviewed by Demirev in the final book chapter (Chap. 13). These include global or local labeling of growth media with C, N, or H isotopes in abundance ratios differing from the natural isotope abundances of these elements. Drug resistance is determined by observing characteristic mass shifts of one or more microorganism-specific biomarkers. A similar approach involves the amplification of organism-specific bacteriophages in targeted microorganisms. In this approach, the shift in biomarker masses for phages, initially proliferated in isotopically manipulated growth medium, is monitored. The advantages of these methods as well as tools for automating the data analysis are also discussed.

Emerging MS Methods and Technologies Not Covered Here

This book has focused on MS methods and applications that rely on generation/analysis of protein and protein-related biomarkers for subspecies typing and strain differentiation of bacteria. These applications have matured significantly as reflected in the dominant number of MALDI TOF MS instruments installed worldwide. Several MS methods not covered here but with potential to impact future clinical applications in microbiology are pointed below.

Peptide-based MS strategies for rapid virus characterization have been developed in the last 15 years. In an early proof of concept (Yao et al. 2002), the Sindbis virus AR 339 was unambiguously identified by mapping the masses of proteolytic products to a database of tryptic peptides generated *in silico* from a set of viruses with sequenced genomes. Animal (swine, avian) and human flu viruses have been rapidly and reliably typed by high-resolution MS mapping of peptide digests of the isolated matrix M1 protein as well as whole-virus digests (Schwahn et al. 2010; Nguyen and Downard 2013). With the development of a phylogenetics algorithm, the method has been expanded to chart the evolutionary history of the influenza virus based on spectra produced from the proteolytic digestion of hemagglutinin (a viral coat protein; Lun et al. 2013). A high degree of overlap is observed between the mass tree (i.e., generated from MS data) when compared to trees generated from the respective viral genome sequences.

A method combining nucleic acid amplification with high-resolution MS detection relies on very accurate measurement of masses of polymerase chain reaction (PCR) products to infer the base composition (Hofstadler et al. 2005; Ecker et al.

2005, 2008). In it, "intelligent" PCR primers target broadly conserved regions between 80 and 140 base pairs that flank the variable microorganism-specific genome regions. The PCR-amplified variable regions (both forward and reverse strands) are analyzed by ESI high-resolution and high mass accuracy MS. The accurate mass information allows unambiguous base composition determination of the amplified regions. A broad set of organisms, including the major families of human and animal viruses, bacteria, and fungi, can be identified by comparison with available genome sequences in databases. The sample preparation procedure, including PCR, currently takes more than an hour. The high degree of multiplexing (more than 1500 PCR reactions per day) facilitates surveillance of a large number of clinical samples for pathogenic microorganisms as well as virulence factors and antibiotic resistance markers.

Nonprotein (including small molecule) biomarker approaches for microorganism characterization rely predominantly on the detection of lipids or lipid constituents (Heller et al. 1987, 1988; Claydon et al. 1996; Krasny et al. 2013), e.g., fatty acids (Hendricker et al. 1999; Voorhees et al. 2006), comprising up to 10% of dry cell weight. Carbohydrates (Fox et al. 2003) and heme (Demirev et al. 2002) have also been identified as biomarkers for microorganism identification by MS. Unlike proteins, correlated directly to the genome, all secondary biomarkers exhibit much higher dependence on environmental conditions, e.g., growth medium.

A laser ablation TOF mass spectrometer has been developed to identify individual airborne micrometer-sized particles, comprising a single cell or a small number of clumped cells (Tobias et al. 2005). This approach is reagent-less, and it relies on laser ablation and detection of lower mass (less than m/z 200) positive and negative ions. MS signatures for aerosolized *Mycobacterium tuberculosis* particles are distinct from *M. smegmatis, Bacillus atrophaeus,* and *B. cereus* particles. This technique is tested as a stand-alone airborne *M. tuberculosis* detector in bioaerosols from an infected patient at airborne concentrations of 1 particle/liter.

Atmospheric pressure ionization (API) techniques are among the emerging tools and approaches developed recently that allow samples, including individual colonies, to be interrogated in ambient conditions (Song et al. 2007; Meetani et al. 2007; Pierce et al. 2007; Watrous et al. 2013; Rath et al. 2013; Strittmatter et al. 2014; Hamid et al. 2014; Fang and Dorrestein 2014; Hayes and Murray 2014; Luzzatto-Knaan et al. 2015). Lipids and other secondary metabolites are the predominant biomarkers detected by desorption electrospray ionization (DESI) in MS profiling of intact untreated bacteria (Song et al. 2007; Meetani et al. 2007). Nano-DESI MS analysis of individual bacterial colonies directly from the Petri dish without any sample preparation has provided unique information on the chemical constituents of each species in vivo and in real time (Watrous et al. 2013). Strains of 28 clinically relevant bacterial species were recently analyzed by rapid evaporative ionization MS (REIMS; Strittmatter et al. 2014). In blind tests, strains cultured on different culture media have been correctly identified more than 97% of the time. Bacterial colonies, smeared onto filter paper, can be rapidly analyzed by paper spray MS without sample preparation (Hamid et al. 2014). Phospholipids—the major bio-

markers observed in both the negative and positive ion mode spectra—allow successful bacterial discrimination at the species level by this API technique.

Perspective

Continuing proliferation of robust MALDI TOF MS systems in clinical laboratories in hospitals is envisioned within the next 5 years. Hardware improvements—miniaturization of the TOF mass analyzer and the laser, and sample preparation modules and associated electronics—are also expected. These will be in parallel with improved instrumental parameters—mass resolving power, mass accuracy, sensitivity, as well as reduction in instrumental and analysis costs. Introduction of new types of mass analyzers (e.g., miniature ion traps) and/or ionization sources (e.g., for API) would further expand the applications of MS in clinical microbiological diagnostics and environmental monitoring. Developments that can accelerate the environmental applications of MS include smaller, commercially available, and less-expensive MS systems with efficient on-line aerosol collectors. Additional research in lab-on-a-chip (microfluidics) devices will result in novel sample preparation protocols. Further improvement of methods for analysis of microbial mixtures, specifically of closely related strains/subspecies, and compiling of "standard" instrument-independent spectral libraries would propel the entire field forward. Software improvements including novel computer bioinformatics algorithms for rapid and automated pathogen identification will be combined with further expansion of available genomic/proteomic information. MS will play an expanded role in the development of novel, rapid, reliable, and efficient methods for detection of hard-to-confirm pathogens in bodily fluids, e.g., *Borrelia,* the causative agent of Lyme disease. The transformation of MS into a viable and widespread tool for biomedical diagnostics at point-of-care has been a long-standing goal of researchers (Mann 2002). With the improvement of current and the advent of new MS methods for pathogen detection, we are coming closer to realizing that goal.

Disclaimer

Mention of commercial products and/or trademarks throughout this book does not imply recommendation or endorsement and is included for information purposes only. Approved regulatory and safety procedures (e.g., microorganism inactivation, work in appropriate biosafety lab, etc.) should be followed when handling pathogens.

Acknowledgments During the preparation of this text P.A.D. was supported in part by a Stuart Janney Sabbatical Fellowship, administered by Johns Hopkins University/Applied Physics Laboratory.

References

Anhalt JP, Fenselau C. Identification of bacteria using mass spectrometry. Anal Chem. 1975;47:219–25.

Bizzini A., Greub G. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. Clin Microbiol Infect. 2010;16:1614–9.

Calderaro A, Arcangeletti M, Rodighiero I, Buttrini M, Gorrini C, Motta F, Germini D, Medici M, Chezzi C, De Conto F Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry applied to virus identification. Sci Rep. 2014;4:6803.

Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. 2013;26:547–603.

Claydon MA, Davey SN, Edwards-Jones V, Gordon DB. The rapid identification of intact micro-organisms using mass spectrometry. Nat Biotechnol. 1996;14:1584–6.

Cliff JB, Kreuzer HW, Ehrhardt CJ, Wunschel DE, editors. Chemical and physical signatures for microbial forensics. New York: Springer; 2011.

Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012;36:380–407.

DeMarco ML, Ford BA. Beyond identification emerging and future uses for MALDI-TOF mass spectrometry in the clinical microbiology laboratory. Clin Lab Med. 2013;33:611–28.

Demirev P, Fenselau C. Mass spectrometry for rapid characterization of microorganisms. Annu Rev Anal Chem. 2008a;1:71–94.

Demirev P, Fenselau C Mass spectrometry in biodefense. J Mass Spectrom. 2008b;43:1441–57.

Demirev PA, Feldman AB, Kongkasuriyachai D, Scholl P, Sullivan D Jr, Kumar N. Detection of malaria parasites by laser desorption mass spectrometry. Anal Chem. 2002;74:3262.

Demirev PA, Feldman AB, Lin JS. Bioinformatics-based strategies for rapid microorganism identification by mass spectrometry. Johns Hopkins APL Tech Digest. 2004;25:27–37.

Ecker DF, Sampath R, Blyn LB, Eshoo MW, Ivy C, Ecker JA, Libby B, Samant V, Sannes-Lowery KA, Melton RE, Russell K, Freed N, Barrozo C, Wu J, Rudnick K, Desai A, Moradi E, Knize DJ, Robbins DW, Hannis JC, Harrell PM, Massire C, Hall TA, Jiang Y, Ranken R, Drader JJ, White N, McNeil JA, Crooke ST, Hofstadler SA. Rapid identification and strain-typing of respiratory pathogens for epidemic surveillance. Proc Natl Acad Sci U S A. 2005;102:8012.

Ecker DF, Sampath R, Massire C, Blyn LB, Hall TA, Eshoo MW, Hofstadler SA. Innovation-Ibis T5000: a universal biosensor approach for microbiology. Nat Rev Microbiol. 2008;6:553–8.

Fagerquist CK. Top-down proteomic identification of bacterial protein biomarkers and toxins using MALDI-TOF-TOF-MS/MS and post-source decay. Rev Anal Chem. 2013;32:127–33.

Fang JS, Dorrestein PC. Emerging mass spectrometry techniques for the direct analysis of microbial colonies. Curr Opin Microbiol. 2014;19:120–9.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass-spectrometry of large biomolecules. Science. 1989;246:64–71.

Fenselau C. Rapid characterization of microorganisms by mass spectrometry-what can be learned and how? J Am Soc Mass Spectrom. 2013;24:1161–6.

Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev. 2001;20:157–71.

Fenselau C, Demirev PA, editors. Rapid characterization of microorganisms by mass spectrometry. Washington DC: ACS; 2011.

Fox A, Stewart GC, Waller LN, Fox KF, Harley WM, Price RL. Carbohydrates and glycoproteins of *Bacillus anthracis* and related Bacilli: targets for biodetection. J Microbiol Methods. 2003;54:143.

Freivald A, Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry. Nat Protoc. 2009;4:732–42.

Hamid AM, Jarmusch AK, Pirro V, Pincus DH, Clay BG, Gervasi G, Cooks RG. Rapid discrimination of bacteria by paper spray mass spectrometry. Anal Chem. 2014;86:7500–7.

Havlicek V, Lemr K, Schug KA. Current trends in microbial diagnostics based on mass spectrometry. Anal Chem. 2013;85:790–797.
Hayes JM, Murray KK. Ambient laser ablation sample transfer with nanostructure-assisted laser desorption ionization mass spectrometry for bacteria analysis. Rapid Commun Mass Spectrom. 2014;28:2382–4.
Heller DN, Fenselau C, Cotter RJ, Demirev P, Olthoff JK, Honovich J, Uy M, Tanaka T, Kishimoto Y. Mass-spectral analysis of complex lipids desorbed directly from lyophilized membranes and cells. Biochem Biophys Res Commun. 1987;142:194–9.
Heller DN, Murphy CM, Cotter RJ, Fenselau C, Uy OM. Constant neutral loss scanning for the characterization of bacterial phospholipids desorbed by fast atom bombardment. Anal Chem. 1988;60:2787.
Hendricker AD, Abbas-Hawks C, Basile F, Voorhees KJ, Hadfield TL. Rapid chemotaxonomy of pathogenic bacteria using in situ thermal hydrolysis and methylation as a sample preparation step coupled with a field-portable membrane-inlet quadrupole ion trap mass spectrometer. Int J Mass Spectrom. 1999;191:331.
Ho YP, Reddy PM. Advances in mass spectrometry for the identification of pathogens. Mass Spectrom Rev. 2010;30:1203–24.
Hofstadler SA, Sampath R, Blyn LB, Eshoo MW, Hall TA, Jiang Y, Drader JJ, Hannis JC, Sannes-Lowery KA, Cummins LL, Libby B, Walcott DJ, Schink A, Massire C, Ranken R, Gutierrez J, Manalili S, Ivy C, Melton R, Levene H, Barrett-Wilt G, Li F, Zapp V, White N, Samant V, McNeil JA, Knize D, Robbins D, Rudnick K, Desai A, Moradi E, Ecker DJ. TIGER: the universal biosensor. Int J Mass Spectrom. 2005;242:23.
Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons. Anal Chem. 1988;60:2299–301.
Krasny L, Hynek R, Hochel I. Identification of bacteria using mass spectrometry techniques. Int J Mass Spectrom. 2013;353:67–79.
Lun ATL, Swaminathan K, Wong JWH, Downard KM. Mass trees: a new phylogenetic approach and algorithm to chart evolutionary history with mass spectrometry. Anal Chem. 2013;85:5475–82.
Luzzatto-Knaan T, Melnik AV, Dorrestein PC. Mass spectrometry tools and workflows for revealing microbial chemistry. Analyst. 2015;140:4949–66.
Mann M. Mass tool for diagnosis. Nature. 2002;418:731–2.
Meetani MA, Shin YS, Zhang SF, Mayer R, Basile F. Desorption electrospray ionization mass spectrometry of intact bacteria. J Mass Spectrom. 2007;42:1186–93.
Nguyen AP, Downard KM. Proteotyping of the parainfluenza virus with high-resolution mass spectrometry. Anal Chem. 2013;85:1097–105.
Pierce CY, Barr JR, Cody RB, Massung RF, Woolfitt AR, Moura H, Thompson HA, Fernandez FM. Ambient generation of fatty acid methyl ester ions from bacterial whole cells by direct analysis in real time (DART) mass spectrometry. Chem Commun. 2007;8:807–9.
Pineda FJ, Lin JS, Fenselau C, Demirev P. Testing the significance of microorganism identification by mass spectrometry and proteome database search. Anal Chem. 2000;72:3739–44.
Rath CM, Yang JY, Alexandrov T, Dorrestein PC. Data-independent microbial metabolomics with ambient ionization mass spectrometry. J Am Soc Mass Spectrom. 2013;24:1167–76.
Sandrin TR, Goldstein JE, Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom Rev. 2013;32:188–217.
Sauer S, Kliem M. Mass spectrometry tools for the classification and identification of bacteria. Nat Rev Microbiol. 2010;8:74–82.
Schwahn AB, Wong JWH, Downard KM. Typing of human and animal strains of influenza virus with conserved signature peptides of matrix M1 protein by high resolution mass spectrometry. J Virol Methods. 2010;165:178–85.
Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49:543–51.

Shah HN, Gharbia SE, editors. Mass spectrometry for microbial proteomic. New York: Wiley-Interscience; 2010.

Song YS, Talaty N, Tao WA, Pan ZZ, Cooks RG. Rapid ambient mass spectrometric profiling of intact, untreated bacteria using desorption electrospray ionization. Chem Commun. 2007;1:61–3.

Strittmatter N, Rebec M, Jones EA, Golf O, Abdolrasouli A, Balog J, Behrends V, Veselkov KA, Takats Z. Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry. Anal Chem. 2014;86:6555–62.

Tanaka K. The origin of macromolecule ionization by laser irradiation (Nobel lecture). Angewandte Chemie Int Ed. 2003;42:3860–70.

Tobias HJ, Schafer MP, Pitesky M, Fergenson DP, Horn J, Frank M, Gard EE. Bioaerosol mass spectrometry for rapid detection of individual airborne *Mycobacterium tuberculosis* H37Ra particles. Appl Environ Microbiol. 2005;71:6086–95.

Voorhees KJ, Miketova P, Abbas-Hawks C, Hadfield TL. Identification of lipid-based biomarkers in the high-resolution pyrolysis/mass spectrum of *Brucella neotomae*. J Anal Appl Pyrolysis. 2006;75:33.

Watrous J, Roach P, Heath B, Alexandrov T, Laskin J, Dorrestein PC. Metabolic profiling directly from the petri dish using nanospray desorption electrospray ionization imaging mass spectrometry. Anal Chem. 2013;85:10835–391.

Welker M. Proteomics for routine identification of microorganisms. Proteomics. 2011;11:3143–53.

Wilkins CL, Jackson O, Lay JO, editors. Identification of microorganisms by mass spectrometry. New York; Wiley; 2005.

Yao Z-P, Demirev PA, Fenselau C. Mass spectrometry-based proteolytic mapping for rapid virus identification. Anal Chem. 2002;74:2529–34.

Part I
Methodology and Techniques

Chapter 2
Methods and Instrumentation in Mass Spectrometry for the Differentiation of Closely Related Microorganisms

Franco Basile and Rudolph K. Mignon

Introduction

Mass spectrometer instruments can be considered as a complex chemical reaction vessel, and as such, the resulting mass spectrum (i.e., the "product" of these reactions) is directly related to all experimental parameters, including, but not limited to, sample preparation, instrument settings, and environmental conditions. Because of its highly informative data output, mass spectrometry (MS) has found many applications in the analysis and quantitation of small to large molecular weight (MW) compounds in areas of energy, environment, forensics, space exploration, and in clinical and biological laboratories, to name just a few. To this list of applications, the analysis of microorganisms has proven to be an accurate and cost-effective approach in clinical settings. Because microorganisms can be considered as a complex chemical sample, its preparation is closely related to the information being sought, and this in turn will determine the type of MS instrumentation to be used. Unfortunately, a single sample preparation protocol will not provide a compatible sample state for all types of mass spectrometers (and vice versa). This relationship between methodology and instrumentation is illustrated (albeit simplified) in Fig. 2.1, where the final sample state prior to analysis is matched with the type of sample preparation required, instrumentation(s), and required data processing.

This relationship between the final state of the sample and MS instrumentation is mainly a consequence of the type of sample inlet and ionization technique used in a particular mass spectrometer. Referring to Fig. 2.1, the analysis of intact cells by matrix-assisted laser desorption/ionization-MS (MALDI-MS), (Jaskolla and Karas 2011) one of the simplest approaches for microorganism analysis by MS, (Holland et al. 1996) requires the isolation of a pure microbial colony, which is then deposited directly onto the MALDI plate. The subsequent mass spectral profiles,

F. Basile (✉) · R. K. Mignon
Department of Chemistry, Biodetection and Mass Spectrometry Laboratory, University of Wyoming, 1000 E. University Ave (Department 3838), Laramie, WY 82071, USA
e-mail: Basile@uwyo.edu

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_2

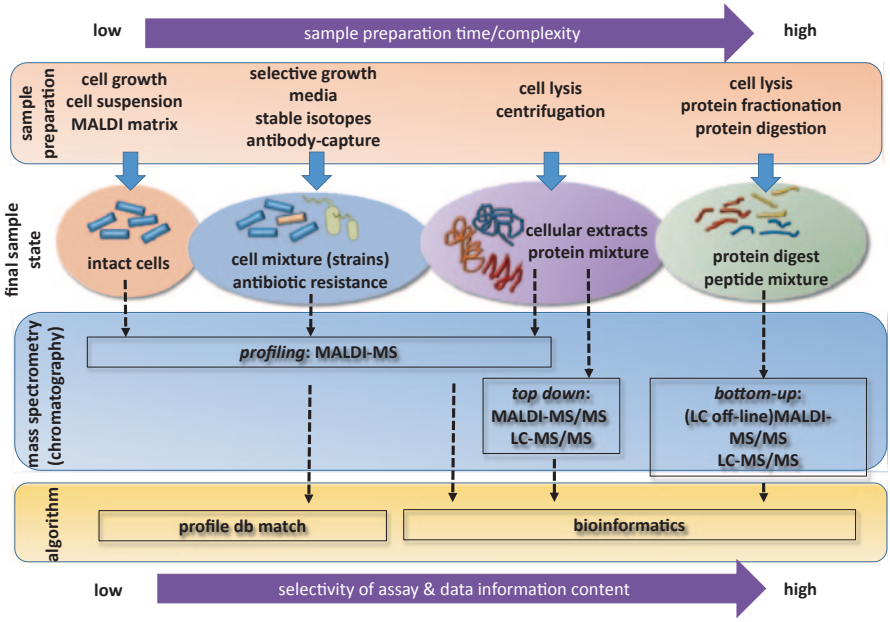


Fig. 2.1 Relationship between sample preparation time and complexity for several MS-based methods for the analysis of microorganisms

consisting mostly of ribosomal proteins, (Holland et al. 1999; Ryzhov and Fenselau 2001) are then used to classify, differentiate, and identify the microorganism. This approach requires the use of standard mass spectral databases of known microorganisms that have been acquired using the same experimental conditions. If on the other hand, one does not possess such standard mass spectral databases, a bioinformatics approach can be used. In one approach requiring a pure microbial sample, the experimentally obtained protein masses are matched to a proteome database (Demirev et al. 2001). In a second approach, a protein signal is selected for gas-phase fragmentation (tandem MS or MS/MS) and the observed ions in the tandem mass spectrum are then matched to expected fragmentation patterns of proteins contained in a proteome database (Fagerquist et al. 2010). This top-down proteomics approach can also be enhanced (i.e., more proteins detected) by the use of a liquid chromatography (LC) separation/fractionation step followed by MS/MS of the intact protein ions (McFarland et al. 2014). However, this enhancement in selectivity comes with additional sample preparation steps to extract proteins and remove other cell components incompatible with the LC step. Because of the unique ion chemistry of the protein fragmentation process and the large mass-to-charge ratios (m/z's) of the resulting fragment ions, these top-down analyses require the use of specialized MS instrumentation that allow for the fragmentation of large protein ions and the analysis of their fragment ions with sufficient mass accuracy to provide meaningful database search results. Lastly, this bioinformatics approach can be per-

formed in a bottom-up mode where the sample preparation includes protein extraction followed by site-specific enzymatic digestion (e.g., with trypsin). The resulting complex mixture of peptides is analyzed by LC-MS/MS, and the acquired tandem mass spectrum for each peptide is matched, via a database search, to the protein originating the peptide, and if possible, its biological origin (i.e., the microorganism). The increased level of complexity for the sample preparation and/or analysis steps for both top-down and bottom-up proteomic approaches results in the highest degree of selectivity of all MS-based methods as a mixture of microorganisms can, in principle, be identified, regardless of growth conditions. Finally, the ability of MS to detect isotopologues allows the use of stable isotopes (e.g., ^{13}C, ^{15}N, or ^{18}O) to differentiate and/or quantitate biomolecules between two different cell states, as in the detection of antibiotic-resistant strains of microorganisms.

The following discussions will focus on factors affecting the ability of MS-based methods to achieve high levels of specificity and selectivity that are required in the detection of closely related bacteria and the detection of antibiotic-resistant strains, followed by a description of MS instrumentation, and examples from the current scientific literature.

Selectivity and Specificity in the Analysis of Microorganisms with MS

In the differentiation or identification of microorganisms, several factors are influential in determining the *specificity of a technique for* a target microorganism or the ability of a technique to *select among* several closely related microorganisms (i.e., selectivity). General strategies to achieve these goals include:

1. *Increasing the selectivity of the measurement to differentiate among unique features that define a certain microorganism*. This strategy may include the addition of a chromatographic step and/or increasing the mass resolution and mass accuracy of the mass spectrometer (time-of-flight (TOF), Fourier transform (FT) orbitrap or FT-ion cyclotron resonance (ICR) mass analyzers).
2. *Decreasing the overall variance of the measurement in order to detect subtle differences in traits common to all samples*. In this instance, the goal is to detect subtle differences in the pattern between two mass spectra, each obtained from different species and/or strain. Thus, the differentiation of two closely related microorganisms depends on the quantitative (relative) detection of small differences in signal strength common to both samples. Factors affecting the overall measurement variance (s^2, where s is the standard deviation) are additive and ideally independent of each other, with the total variance of an analysis being the sum of the individual steps in the analysis

$$s^2_{analysis} = \sum_{i=1}^{n} s_i^2,$$

where *i* is the individual step (e.g., sampling, sample preparation, measurement, data processing) in the overall analysis. In general, it has been recognized that the individual variances in the analysis follow the trend:

$$s^2(\text{sampling}) > s^2(\text{sample prep}) >> s^2(\text{measurement})$$

Therefore, it is usually the case for most analytical protocols to focus on decreasing the variance contributions of the sampling and sample preparation steps. Manufacturers of modern chemical instrumentation, with the availability of advanced electronic components and signal processing, have considerably decreased the contribution of the measurement to the overall analysis variance. The use of automation in both sample preparation and data acquisition is key in a strategy to reduce the overall variance of the analysis. The contribution due to sampling can be reduced by increasing the number of biological samples analyzed (replicate samples).

3. *Increasing the specificity of the measurement for a target microorganism.* Factors that may increase the specificity for a target microorganism include the incorporation of a selective growth media step (antibiotic resistant), DNA amplification, antibody capture/enrichment, stable-isotope labeling, and multistage mass analyses (e.g., tandem MS or MS/MS, selective reaction monitoring or SRM, vide infra).

Approaches involving these strategies will be addressed in subsequent sections of this chapter with examples from the recent literature. However, a brief review of the MS instrumentation involved in these measurements will be presented first.

MS Instrumentation

The analysis of microorganisms with MS-based techniques involves a wide range of instrumentation, and knowledge of their capabilities and limitations is key in extracting the most information from the analysis. Two components are fundamental in defining the capabilities of any MS instrument and include the type of (i) ionization and (ii) mass analyzer used. For the techniques relevant to the characterization of biomarkers in microorganisms being discussed here, only MALDI and electrospray ionization (ESI), with the TOF, quadrupole(s), and orbitrap mass analyzers, will be described in detail. However, regardless of the type of MS instrument being used, a common operational requirement is that the final state of the sample, prior to mass analysis, be gas-phase ions of either positive or negative polarity. These gas-phase ions are then separated or sorted based on their mass-to-charge ratio or *m*/*z,* a dimensionless quantity (Price 1991; Gross 2011). (For convenience, mass, *m,* is expressed in terms of the unified atomic mass unit, which is defined as 1/12 the mass in kilograms of one atom of ^{12}C, u or $m_u = 1.66054 \times 10^{-27}$ kg. Thus, the quantity *m* is the ratio of the mass in kilograms of the ionized molecule divided by m_u or

$m = m(kg)/m_u(kg)$. The quantity z represents the number of elementary charges on the ion, which is also a dimensionless number) (Boyd 2008). All mass analyzers are operated under vacuum ($\sim 10^{-4}$–10^{-12} Torr), their magnitude depends on the mode of operation, and are required in order to avoid collisions of the analyte gas-phase ion with neutral molecules present in air (as well as avoiding arcing within components in the mass analyzer held at high voltages). This increases signal sensitivity and avoids unwanted ion–molecule reactions between the analyte ion and reactive gaseous species (e.g., oxygen).

Both MALDI and ESI are unique in their ability to form gas-phase ions from large MW molecules, biological or synthetic, without inducing fragmentations, and are thus considered to be "soft" ionization techniques (unlike "hard" ionization techniques like electron ionization (EI) which induce fragmentations *during* the ionization step) (McLafferty and Tureek 1993). ESI is considered an atmospheric pressure (AP) ionization technique since ions are generated outside the mass analyzer vacuum manifold. Although MALDI is usually conducted under vacuum in TOF-MS instruments used for bacteria identification, MALDI can also be performed under AP conditions, (Laiko et al. 2000; Madonna et al. 2003) allowing its use with instruments originally setup to use ESI, like the triple quadrupole MS.

MALDI and MALDI-MS Instrumentation

The development of MALDI by Hillenkamp and coworkers (Karas and Hillenkamp 1988) allowed for the analysis of high MW biological (e.g., proteins) and synthetic (e.g., polymers) samples without inducing fragmentation. The MALDI process relies on mixing an organic compound, termed the matrix, with the biological sample, the former in a 100:1 to 1000:1 molar excess. When the mixture is dried, the organic compound forms a heterogeneous crystalline matrix (Fig. 2.2) that surrounds and isolates individual analyte molecules in the original biological sample. Upon irradiation by a pulsed laser (UV laser in most commercial instruments), the photon energy is absorbed predominantly by the matrix compound and this electronic excitation is converted into thermal (vibrational) and translational energy, ablating (i.e., desorbing) matrix molecules as well as intact and ionized analyte molecules into the gas phase (Zenobi and Knochenmuss 1998). As such, the MALDI process is considered a pulsed ion source as it generates discrete packets of ions.

Because of the heterogeneous nature of the MALDI matrix when dry, ion yields at different locations within a MALDI matrix are not the same, leading to the description of these locations within the sample as "hot" or "cold" spots to refer to locations yielding intense or weak signals, respectively. The presence of these hot and cold signal spots within the MALDI matrix limits the usefulness of the MALDI process as a quantitative tool, imposing the need to acquire, on average, several hundred mass spectra from different locations within a sample in order to obtain a representative (average) mass spectrum.

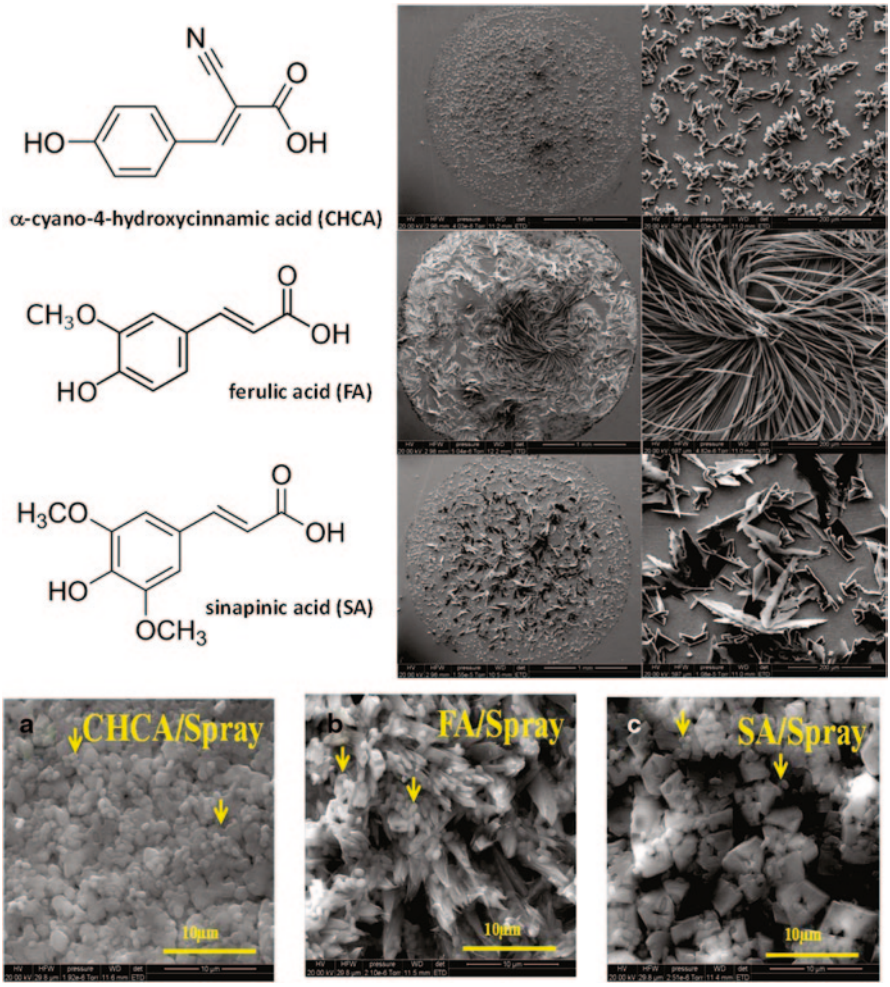


Fig. 2.2 Structures and scanning electron microscopy *(SEM)* photographs of different MALDI matrices deposited onto a stainless steel plate. *Lower* SEM photographs show *E. coli* cells co-crystallized with different MALDI matrices (matrix applied with a spray deposition technique). *Arrows point* to intact cells within the crystalline matrix. (Adapted from Toh-Boyo et al. 2012, copyright American Chemical Society)

The MALDI ionization process is very complex and depends heavily on the type of analyte molecule, matrix used, and laser fluence, but a recent study (Jaskolla and Karas 2011) suggests that two ionization models are mainly at play: (1) charge separation during the desorption step of *preformed ions* embedded in the crystalline matrix (a.k.a., the "Lucky Survivor" model), and (2) gas-phase protonation via ion–molecule reactions during the desorption step. In the analysis of biological molecules, the MALDI process yields primarily single-charged ions, either due to protonation or cation adduct formation (e.g., $[M+H]^+$ or $[M+Na]^+$, where M is the neutral molecule) in positive ion mode or deprotonation in negative ion mode (e.g.,

$[M-H]^-$). This fact is particularly useful when analyzing a mixture of proteins as it yields a simplified mass spectrum without overlapping signals. Important to note when analyzing complex mixtures of biomolecules with MALDI is the signal suppression effect, which takes place during the ionization process. For example, in positive ion mode the signal from a highly abundant, but acidic protein may be suppressed by the presence of a low abundant, but basic protein which yields an intense signal. As a result, what is seen in the mass spectrum is neither a quantitative nor a qualitative reflection of the composition of the sample. This effect is clearly exemplified in the MALDI-MS analysis of intact bacterial cells, which mostly yields signals due to ribosomal proteins while DNA, metabolites, lipids and other high MW proteins remain undetected.

Part of the success of MALDI-MS for the analysis of microorganisms derives from the simplicity and robustness of the methodology, and in its simplest form, intact or whole cells can be deposited directly onto the MALDI plate or mixed with the matrix solution and analyzed directly. Many methods have been published describing this process, but it is believed that bacterial cells are lysed and proteins extracted into the matrix solution in the minutes before crystallization (i.e., during solvent evaporation on the plate, ~1–2 min), even though preserved cell integrity has been observed in microphotographs of the co-crystallized bacteria-matrix sample (Fig. 2.2) (Toh-Boyo et al. 2012; Madonna et al. 2000). This is backed by the fact that protocols using either solvent extraction or intact cells are both effective in producing similar protein signals, albeit with different profiles (i.e., relative peak intensities) (Basile 2011).

As mentioned earlier, it is generally agreed that the majority of the proteins observed in the analysis of bacterial cells with MALDI-MS are ribosomal proteins in the molecular mass range of 2–20 kDa (Holland et al. 1996, 1999; McFarland et al. 2014; Suarez et al. 2013). This is the case since they are abundant (almost half of the mass of growing cells), basic ($pI > 9$, easily ionized under mild acid conditions), and slightly hydrophilic in nature (easily solubilized when mixed with the matrix solution) (Ryzhov and Fenselau 2001). These facts highlight the importance of solvent composition and control of every step (i.e., exact sequence of events) (Cohen and Chait 1996) in the sample preparation protocol for MALDI-MS of bacteria, as they dictate the range of proteins detected, their observed signal strength, and overall signal pattern.

TOF Mass Analyzer

The TOF mass analyzer is suitable to measure the m/z distribution of discrete pulsed ion sources, unlike a continuous stream of ions, and for this reason it is usually coupled with MALDI, a pulsed ion source. In a TOF-MS, a discrete packet of ions with different m/z's (generated via MALDI) are accelerated to the same kinetic energy by applying a voltage (U~10–25 kV; direct current, DC) to the stainless steel sample plate. These ions enter a field-free region (no voltage or magnetic fields

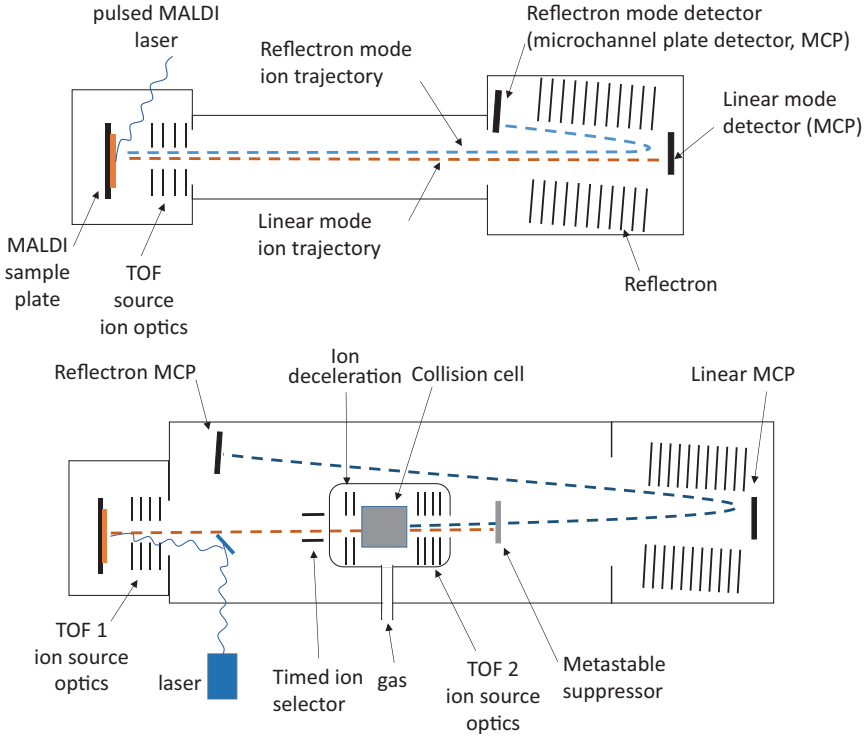


Fig. 2.3 Simplified general diagrams of **a** MALDI-TOF-MS and **b** MALDI-TOF/TOF-MS (based on the Sciex 4800/5800™ systems)

applied) where ions with small *m*/*z*'s travel faster than those with large *m*/*z*'s, and the different times to travel a predefined distance *(d)* forms the basis for their mass separation. The simplified relationship between TOF (t_{TOF}) and *m*/*z* is given by:

$$t_{TOF} = \frac{d}{\sqrt{2U}}\sqrt{\frac{m}{z}}.$$

In principle, the TOF-MS does not have an upper mass limit; however, in practice they are limited by the efficiency of the multichannel plate (MCP) detector (vide infra, Fig. 2.3) in converting low kinetic energy ions (i.e., large *m*/*z*'s) into a detectable electrical current, and the ability of the ionization source to produce ions of large *m*/*z*. Operationally, the relationship between *m*/*z* and t_{TOF} is established by calibration with a set of standard compounds of known *m*/*z* values for their $[M+H]^+$ ions. This calibration is dependent on matrix type and laser intensity (each affects the initial ion velocity during desorption) and the sample position within the MALDI plate (affecting the distance traveled, *d,* and the effective accelerating voltage, *U,* experienced by the desorbed ion). A simplified general diagram of a MALDI-TOF-MS instrument is illustrated in Fig. 2.3a.

Instruments based on this design (linear, but not necessarily reflectron) form part of most, if not all, of the commercially available MALDI-TOF-MS microorganism identification systems that are based on matching a mass spectrum to a mass spectral library of microorganisms (i.e., profile-based MALDI-MS). When operated in the reflectron mode to increase the mass resolution of the measurement (practical up to ~ *m*/*z* 5000), current state-of-the-art TOF mass analyzers (with a properly designed MALDI ion source) specify mass accuracies in the 1 ppm or ± 0.001 (at *m*/*z* 1000). However, no peptide/protein sequence information can be derived from this mass measurement alone and an additional level of selectivity, tandem MS or MS/MS, is required to obtain this information.

Another type of MS available with the MALDI ion source and based on the TOF mass analyzer is the tandem TOF or TOF/TOF-MS. This configuration has two TOF mass analyzers configured in series, separated by a collision cell. This configuration has the capability of obtaining mass spectral protein profiles as well as sequence information of peptides (up to ~4000 Da), and for one manufacturer mid-sized proteins (5–15 kDa). A simplified schematic of a MALDI-TOF/TOF-MS instrument is illustrated in Fig. 2.3b (the following discussion is based on the Sciex 4800™ system (Yergey 2002). An excellent discussion of the inner workings of the Bruker MALDI-TOF/TOF-MS system can be found in Suckau et al. (2003)). Ions formed in the MALDI ion source are accelerated toward the first TOF mass analyzer (TOF1), where ions are separated according to their *m*/*z*'s. In the MS mode, ions are allowed to travel uninterrupted to either the linear or reflectron detector. In the tandem MS (MS/MS) mode, ions of a single *m*/*z* value are selected and allowed to enter the collision cell. This *m*/*z* selection is performed via a timed ion selector, with a series of voltages applied at a unique time on the path of the ion beam so as to deflect all ions except those of the desired *m*/*z* (or TOF). The selected ion is introduced into the collision cell filled with a neutral gas like argon or nitrogen, and upon collision, fragment ions are formed via collision-induced dissociation (CID). Precursor ions can also undergo fragmentation during the MALDI process, via laser-induced dissociation (LID), (Suckau et al. 2003) or after the MALDI process, via post-source decay (PSD), (Neubert et al. 2004; Fagerquist 2013) where metastable ions leave the MALDI ion source and fragment during their voyage through TOF1. In all these fragmentation events, CID, LID, or PSD, the generated fragment ions will have roughly the same velocity as the precursor ion, and thus they cannot be discriminated by their *m*/*z*'s within the TOF1 mass analyzer. The TOF/TOF instrument achieves mass separation of these fragment ions (and obtains useful sequence information) by re-acceleration of this ion packet into the second TOF mass analyzer (TOF2). This second acceleration event becomes the starting point for recording the fragment ion mass spectrum (Yergey et al. 2002). For small proteins (15 kDa), this type of instrument can be used for top-down proteomic measurements, where the fragmentation of a single protein signal from a mixture can yield sequence information about the precursor ion and has been used to discriminate proteins varying by a single amino acid in their sequence (Fagerquist et al. 2010). Finally, when coupled with offline LC and fraction collection directly onto the MALDI plate, this type of instrumentation allows for bottom-up proteomic measurements (Marcus et al. 2007; Benkali et al. 2008; Bodnar et al. 2003).

ESI and ESI-MS Instrumentation

The development of ESI as an ionization source for MS by Yamashita and Fenn (1984) allowed the formation of gas-phase ions of biomolecules in liquid samples, thus enabling the analysis of intact proteins in solutions and of samples separated by LC. A detailed discussion of ESI is beyond the scope of this chapter as many excellent reviews and books have been written on the subject(Bruins et al. 1998; Cech and Enke 2001; Cole 2008). In general, the ESI process in positive ion mode for most biomolecules (proteins and peptides) starts in an acidified solution, that is, by the formation of ions via protonation of basic groups. This is typically accomplished by the addition of a volatile organic acid like acetic acid or formic acid (1 % or 0.1 %, respectively) in a 50 % organic-aqueous solvent (methanol or acetonitrile). The solution is then driven into a metal capillary (~ 50–100 μm inner diameter) connected to a power supply at 3–4 kV (DC voltage). As the liquid emerges at the open end of the capillary, the large electric field causes charge separation of the preformed ions in solution. In the case of biomolecules, an ionized peptide (positive charge) is separated from either a formate ion ($HCOO^-$) or the acetate ion (CH_3COO^-). The use of trifluoroacetic acid (TFA) to acidify solutions for ESI analysis is discouraged as the CF_3COO^- ion forms a strong ion pair with the positively charged biomolecule, making charge separation difficult and thus lowering the ionization efficiency of the ESI process. This accumulation of positively charged ions at the open end of the capillary causes the deformation of the liquid meniscus into what is termed a Taylor cone. Eventually, the electrostatic repulsive forces between the positive charges accumulated in the meniscus exceed the surface tension of the liquid leading to the formation of a fine jet of liquid, which breaks into fine droplets, each containing an excess of positively charged molecules. According to the ion evaporation model (IEM), (Nguyen and Fenn 2007) these droplets undergo a cascade of evaporative and Coulomb fission (charge repulsion) cycles until droplets of about 10 nm in diameter are formed. At this droplet size the effective electric field at the surface is large enough to push one or more solvated ions into the gas phase. A second ionization model, the charge residue model (CRM), describes the generation of an ion when all the solvent is evaporated from the droplet. Although there are many studies showing the prevalence of one model over the other in ESI, the consensus is that large ionized molecules (1000 u) are generated by a process closely described by the CRM. On the other hand, smaller and solvated ions can be emitted from nano-droplets by a process better described by the IEM (Wilm 2011).

In general, for positive ion mode ESI, ionization efficiency is dictated not only by the basicity of the molecule, but also by its hydrophobicity, which determines its concentration at the surface of the droplet (i.e., surface activity). As a result, not all biomolecules present in the sample are ionized with the same efficiency. That is, basic and hydrophobic molecules (with a high surface activity) tend to ionize more efficiently than basic and highly polar molecules. For example, a peptide with a high content of hydrophobic amino acids (phenylalanine, tryptophan) will experience a higher ESI ionization efficiency than a peptide of the same charge but with amino

acids with polar side chains (serine, aspartic acid). Therefore, in the ESI-MS analysis of a complex mixture (e.g., a protein mixture derived from bacteria cell lysate), the observed mass spectrum is neither a quantitative nor qualitative reflection of the composition of the sample. Molecules in the sample with high hydrophobicity and basicity will ionize more efficiently than molecules with lower hydrophobicity and/or basicity, even though the latter may be present at higher concentration. Another characteristic of ESI is the fact that proteins are ionized at multiple sites yielding charge state distributions of ions with multiple protons: $[M+H]^+$, $[M+2H]^{2+}$, …, $[M+nH]^{n+}$. In general, ionization suppression effects in ESI are more pronounced than in MALDI, and coupled with the possibility of observing overlapping charge state distribution from different proteins, ESI-MS is not as straightforward as MALDI-MS for the analysis of complex protein mixtures and is the main reason that ESI-MS is usually coupled with online LC separation.

However, LC-(ESI)-MS offers several key advantageous features for the analysis of closely related bacteria. First, the analysis of large proteins, above 20 kDa, is possible by LC-ESI-MS, increasing the dynamic range of biomarkers available for detection (Everley et al. 2008). In addition, the CID process is more efficient when performed on ions with large charge states (Schaaff et al. 2000) (i.e., large z values), resulting in information-rich fragmentation mass spectra that can identify the precursor peptide (bottom-up proteomic) or protein (top-down proteomic) by its unique amino acid sequence (McFarland et al. 2014). This process, however, requires time-consuming sample preparation and far more complex instrumentation and data analysis than the whole-cell bacteria-MALDI-MS approach (see Fig. 2.1).

Quadrupole-Based Mass Analyzers

One of the earliest MS instruments to be interfaced with ESI is quadrupole-based mass analyzer (Fenn et al. 1989). Early work on the use of quadrupole-based MS instruments for the analysis of microorganisms focused on the detection of mostly lipid biomarkers like phospholipids, triglycerides, and free fatty acids (Anhalt and Fenselau 1975; Meuzelaar and Kistemaker 1973; Huff et al. 1986; Goodacre et al. 1998; Boon et al. 1981; Guckert et al. 1986; DeLuca et al. 1992). In addition to using targeted extraction/derivatization protocols, these early investigations also incorporated rapid thermal desorption and/or pyrolysis methods (with EI) to directly analyze intact bacteria in a manner of minutes (DeLuca et al. 1990). More recently, the triple quadrupole MS (QQQ; or QqQ, where q signifies the collision cell, Q2; vide infra) (Fig. 2.4a) in conjunction with ESI has been used for highly specific detection of microorganisms via targeted bottom-up proteomic approaches (Karlsson et al. 2012; Picotti and Aebersold 2012).

The quadrupole mass analyzer is truly a scanning instrument, in that only ions of a particular m/z can be transmitted through the device (i.e., have a stable trajectory) at a particular time, and thus it is often referred to as a mass filter. The quadrupole mass analyzer consists of a set of four metal rods, ideally each having a parabolic

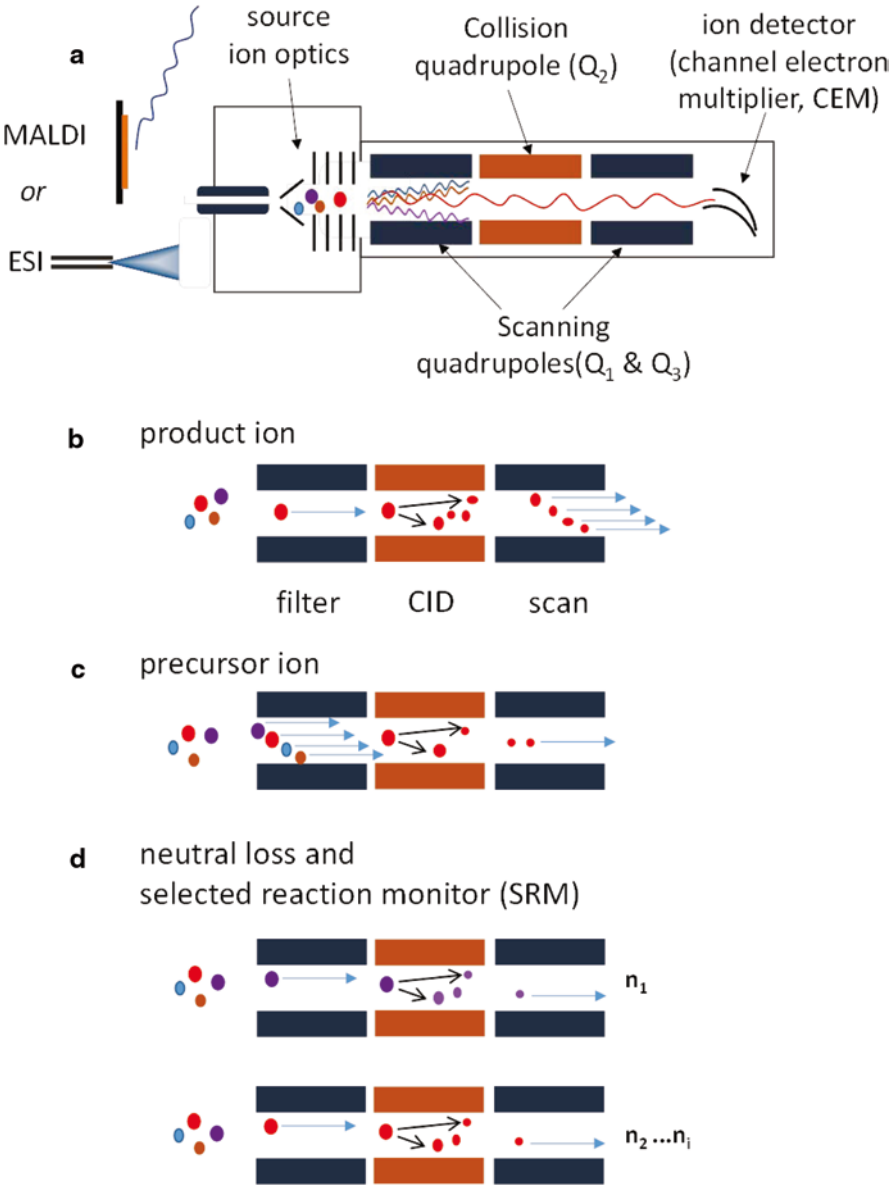


Fig. 2.4 Triple quadrupole MS, QQQ, where Q1 and Q3 indicate scanning quadrupoles and Q2 indicates the collision cell consisting of either an RF-only quadrupole, hexapole, or octapole. In some hybrid instruments Q3 is replaced by either a TOF or a linear quadrupole ion trap. (See text for more details and references for further reading)

surface shape, connected to a DC and radio-frequency (RF) power supplies. By varying the DC and RF voltages applied to opposite rods in the quadrupole mass analyzer (while maintaining the ratio of their magnitudes constant), the stability of ions with different *m*/*z* through the device is sequentially varied, effectively scanning a user-selected and predefined *m*/*z* range. Because modern electronics are able to control voltages and frequencies with high accuracy and precision, these instruments are well suited for quantitative measurements. In addition, the ability of these instruments to perform tandem-MS (MS/MS) measurements makes them ideal for the design of highly specific methodology for the detection of a wide range of chemical and biological species.

The QQQ-MS can be operated in four main scan modes: (i) full scan, (ii) product ion scan, (iii) precursor ion scan, and (iv) SRM (Fig. 2.4a–d). In the full-scan mode, usually Q1 is scanned, while both Q2 and Q3 are operated in RF mode only, essentially acting as ion guides (Fig. 2.4a). In the product ion mode (Fig. 2.4b), Q1 is set to pass only ions of a single *m*/*z,* filtering out all other ions formed in the ion source, and the collision cell, Q2, is operated in RF mode and filled with argon gas ($0.5–2 \times 10^{-3}$ Torr or ~0.2 Pa). Preselected ions emerging from Q1 are accelerated into Q2 where they undergo inelastic collisions with the argon gas, inducing molecular ion dissociation that yields both neutral fragments (not detected) and fragment ions. These fragment ions are then sorted according to their *m*/*z*'s by the scanning Q3. The precursor ion scan mode (Fig. 2.4c) is useful in situations where there is a need to determine the source (precursor) of a particular fragment ion or to survey members of a particular class of compounds that have a common fragment ion (e.g., glycerophospholipids produce a common fragment ion at *m*/*z* 184 regardless of the mass of the precursor ion) (Murphy et al. 2001). To accomplish this measurement, Q1 is scanned and a narrow *m*/*z* window of ions is sequentially introduced into Q2 and fragmented, while Q3 is set fixed at the particular *m*/*z* value of the (common) fragment ion. The mass spectrum is then plotted with the intensity of the fragment ion versus the mass scale of Q1 (not Q3 since it is fixed at a single *m*/*z*). The SRM mode is a highly specific mode of operation of the QQQ instrument where the specific precursor–fragment ion relationship is measured (Fig. 2.4d). For example, a particular peptide known to be a specific biomarker for a disease or microorganism can be detected by setting Q1 to its precursor *m*/*z* value and Q3 to a unique fragment ion of this precursor. The precursor ion, upon fragmentation in Q2 will produce the specific fragment ion that will be transmitted through Q3 and detected. The specificity of the assay is directly related to the specificity of the precursor-product transition (i.e., of the fragmentation reaction), and thus many validation measurements must be performed prior to SRM measurements (Picotti and Aebersold 2012; Lange et al. 2008). In practice, several of these precursor-product reactions can be measured sequentially, and thus the term multiple reaction monitoring (MRM) is also used.

In some hybrid triple quadrupole-based MS systems, the last quadrupole (Q3) is replaced with a TOF analyzer, with an ion path set at a 90° angle from the ions exiting the second quadrupole (or collision cell), and thus it is often termed an

orthogonal TOF. This quadrupole-TOF setup, or Q-TOF (or Qq-TOF, where q signifies the collision cell) increases the resolution and mass accuracy of the product mass spectrum and is also used for bottom-up proteomic measurements (Martinez et al. 2010; Mott et al. 2010; Alvarez et al. 2013). In another hybrid configuration, Q3 is replaced by a linear quadrupole ion trap mass analyzer, which allows ion accumulation for increased sensitivity and MS^n capability (Londry and Hager 2003).

Orbitrap Mass Analyzer

The orbitrap mass analyzer is an ion trap device that provides high accuracy and resolution mass measurement without the need of a magnetic field, (Hu et al. 2005; Zubarev and Makarov 2013) and thus it is more accessible in terms of lab requirements, and initial and operating costs. Some consider it to be the "gold standard" mass spectrometer for proteomic-based measurements (Mitchell 2010). The orbitrap mass analyzer is usually found in a hybrid configuration interfaced with a linear ion trap mass analyzer and transfer octopoles and C-trap (Fig. 2.5) (Senko et al. 2013).

In its core operation mode, ions injected into the orbitrap are trapped in an electrostatic field and oscillate along the central electrode (*z*-axis) with a periodic back

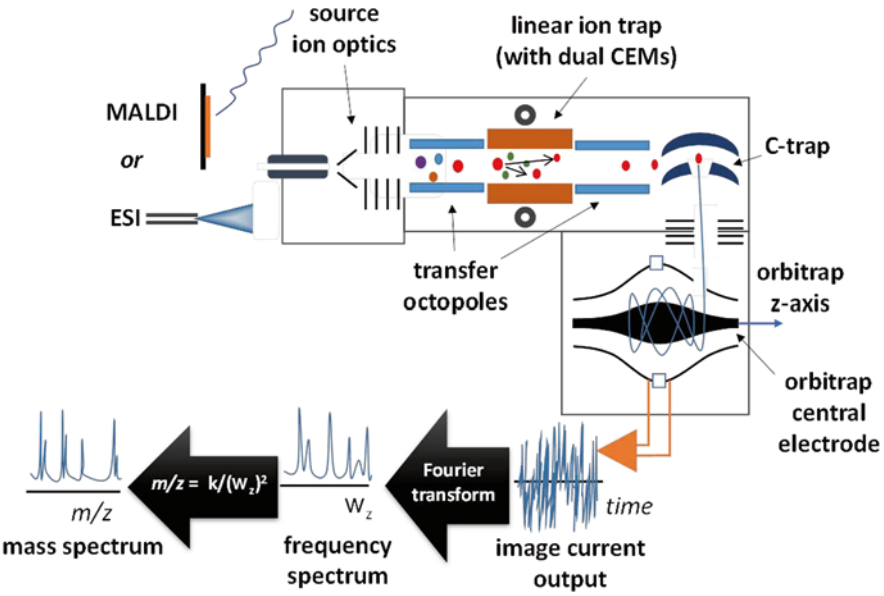


Fig. 2.5 Diagram of a hybrid linear ion trap and orbitrap mass spectrometer

and forth motion. The frequency of this axial oscillation (w_z) is inversely proportional to the square root of the m/z of the trapped ions.

$$\omega_z = \sqrt{k\frac{q}{m}},$$

where the electric charge q is equal to multiples (z) of the electron charge (e) and k is a parameter describing the field (Makarov 2000). The ion oscillations are recorded in the time domain by detecting the transient image current on the electrodes. The respective ion frequencies are derived from the transient by a fast FT and the corresponding m/z values are determined using the above equation.

Orbitrap-based MS systems are predominantly used for bottom-up and middle-down proteomic measurements, (Cannon et al. 2010) where the peptides are fragmented in the linear ion trap and fragments mass analyzed in the orbitrap. In addition, orbitrap-MS systems are used for accurate mass determination of intact proteins (McFarland et al. 2014). These measurements are particularly useful for the detection and differentiation of closely related microorganisms since this methodology allows the detection of a wide range of proteins in the sample, beyond the detection of ribosomal proteins.

MS-Based Methods and Instrumentation for the Differentiation of Closely Related Bacteria: Strain Level and Antibiotic Resistant

The analysis of microorganisms by MALDI-MS has been successful since the range of biomarkers detected afford the required level of selectivity to differentiate samples among a wide range of microorganisms, bacteria and fungi included, thus allowing nontargeted analyses. Extensive work published in the literature demonstrates the ability to obtain phylogenetic classification via MALDI-MS, equivalent to that obtained by 16S rRNA (Seng et al. 2009; Hsieh et al. 2008; Boehme et al. 2013; Li et al. 2014; Deng et al. 2014). It is natural that the use of MALDI-MS has been extended to the identification of bacteria at the strain level and for the differentiation of antibiotic-resistant strains. In this section, the MS-based techniques used in achieving these goals will be highlighted with selected examples from the literature, with an emphasis on MS-hardware and technique (i.e., sample preparation). The reader is also referred to more comprehensive literature reviews on the detection of microorganisms at the strain level and/or antibiotic-resistant strains (Croxatto et al. 2012; Sandrin et al. 2013; Hrabak et al. 2013).

Differentiation of bacteria beyond the species level with MS is challenging since the detected number of unique or characteristic biomarkers decreases as the similarity between microorganisms increases. As a result, the measurement requires a higher level of selectivity, mass spectral profile reproducibility (relative peak intensities), and mass accuracy. For MS-based measurements this may imply strategies involving additional sample preparation steps (e.g., protein extraction, digestion),

the inclusion of a separation step (fractionation, LC), extending the mass range of the analysis (detection of higher MW biomarkers), increasing the selectivity of the MS measurement (MS/MS), or including enzyme substrates or stable isotope reagents (Fig. 2.1).

Profile-Based Techniques for Strain-Level Differentiation

These measurements are attractive since they require the simplest form of instrumentation available, a MALDI-TOF-MS operated in the linear mode ($m/z \sim 2000$–20,000; Fig. 2.3a). Ideally, for profile-based differentiation, a high degree of reproducibility is desired, along with high signal-to-noise ratio (S/N), to consistently produce protein profiles with distinct and unique features (specific) to each strain. The approach used to prepare the sample in MS has enormous consequences on the quality (S/N) and reproducibility of the resulting mass spectral profiles (signal relative intensities), and so this is an obvious experimental step to optimize. This is both true for ESI and MALDI; however, MALDI is more appropriate for samples containing complex mixtures and more robust to small differences in solvent composition and procedural steps than ESI. For the differentiation of microorganisms at the strain level, sample preparation techniques can increase both the reproducibility of the signal(s) and the range of biomarkers detected. In fact, it is generally agreed that the incorporation of a protein extraction step increases the rate of identification of bacteria at the species level, and in some cases at the strain level, (Croxatto et al. 2012; Clark et al. 2013; Lartigue 2013) especially for Gram-positive cocci (Alatoom et al. 2011). This is most likely due to the removal of other cellular components and salts that can cause signal suppression, thus improving the overall S/N of the mass spectrum.

In general, the use of profile-based MS techniques to differentiate microorganisms at the strain level has met with limited success as the presence of unique strain-specific biomarker(s) can be inconsistent and/or microorganism-dependent. It is usually the case for MALDI-MS profiling that the number, nature, and quality of the reference mass spectra improves the reliability of the identification at the species level (Calderaro et al. 2013). For example, Shao and coworkers (Zhu et al. 2013) improved the identification of several strains of *Haemophilus influenza* and *H. haemolyticus* after curating their mass spectral database. The original reference mass spectral database failed to identify any of the *H. haemolyticus* strains at the species level, but was able to do so after the database was updated with reference mass spectra. In addition, cluster analysis of the obtained mass spectra profiles (and using the standard protein extraction protocol) yielded a dendogram clustering showing clear differentiation of *H. haemolyticus* from *H. influenza*. In addition, *H. influenza* was further differentiated by geographical origin, that is, the Chinese strains were differentiated from those of foreign origin.

However, in most cases, different strains of the same species are correctly identified only at the species level by MALDI-MS profiling (Kolecka et al. 2013;

Kierzkowska et al. 2013). For example, in the comparison of two commercially available microorganism identification systems based on the MALDI-MS instrument and using the standardize sample preparation protocol (direct transfer method, vide infra), 54 streptococcal strains were correctly identified only at the species level (Karpanoja et al. 2014). Similarly, in the MALDI-MS analysis of 24 clinical isolates of the fungus *Trichophyton rubrum*, (Pereira et al. 2014) all strains were identified at the species level only, even after efforts to optimize the sample preparation step using different matrices, formic acid extraction, and/or sonication. Finally, in an attempt to identify bacterial strains related to normal and sensitive skin disorders with MALDI-MS profiling, (Hillion et al. 2013) no correlation was found between phylum, genus or bacterial species and the sensitive skin phenotype, even though all bacteria were correctly identified at the species level. It is worth pointing out that these examples illustrate the importance of reproducibility in the generation of replicate mass spectral profiles of bacteria, as small differences in profiles can yield information about strain differentiation. However, this approach can be limited by the overall reproducibility or variance of the resulting mass spectra (both from samples and in the database), and thus this strategy can benefit from a reduction in the total variance of the analysis (vide infra).

Although limited, several strategies have successfully differentiated microorganisms at the strain level using profile-based MALDI-MS and they include: (1) optimization of the sample preparation step and (2) optimization of growth conditions. Other approaches incorporating bioinformatics (Demirev et al. 2001; Tamura et al. 2013) data analyses will not be discussed in this section.

Optimization of the Sample Preparation Step Prior to MALDI-MS This area of research has received a lot of attention from investigators using MALDI-MS due to the pronounced effects that sample preparation has on the resulting mass spectral profiles, the low cost of implementing these changes (mostly reagents and solvents), and the relatively ease of customization depending on the sample type and/or application. However, it is generally agreed that acidic conditions followed by addition of an organic solvent is sufficient to access most of the ribosomal proteins detected in profile-based MALDI-MS measurements. Other approaches have been proposed to either increase the S/N of the profile and/or extend the MW range of detected proteins and include on-probe sample treatment with ethanol, (Madonna et al. 2000) use of additives (crown ethers or thymol), (Liu et al. 2007; Holland et al. 2014) and the implementation of a heating step (Horneffer et al. 2004; Prieto 2006). Two standard sample preparation protocols are currently used in conjunction with most commercially available MALDI-MS bacteria identification systems, and can be broadly classified either as direct transfer or protein extraction. In the direct transfer protocol, a bacterial colony is smeared directly onto the MALDI plate and overlaid with matrix solution, with bacteria inactivation and protein extraction being performed on the MALDI plate as the matrix solution evaporates (< 1 min). Improved identification rates in the direct transfer method were achieved for Gram-positive rods by incorporating a 70 % formic acid pretreatment (on-probe) prior to the addition of the matrix (Werner et al. 2012; Schulthess et al. 2014). In the protein

extraction protocol or ethanol-formic acid procedure, (Freiwald and Sauer 2009) an isolated bacterial colony is first washed with deionized/distilled water, followed by a 75% ethanol/water wash. This step is intended to remove any media contamination and inactivate the bacteria (without spore formation). The resulting pellet is re-suspended in equal volumes of 70% formic acid and acetonitrile (sequentially added, up to 20 μL final volume) mixed and centrifuged. An aliquot (~1 μL) of the supernatant is deposited onto the MALDI plate, dried and overlaid with α-cyano-4-hydroxycinnamic acid (CHCA) matrix dissolved in a 50% acetonitrile/water and 2.5% TFA. For pathogenic bacteria with spore formation, (Lasch et al. 2008) the TFA protein extraction procedure is recommended and involves an aggressive inactivation step with 80% TFA followed by 50% acetonitrile/water.

The use of protein extraction, along a well-curated mass spectral database, can provide accurate identification of many strains at the species level, and in some cases, differentiation at the strain level. However, in most of these situations the differentiation is based on the pattern of several mass spectral peaks and not on a unique biomarker ion. For example, work conducted by Calderaro and coworkers (2014) on the differentiation of *Leptospira* species at the serovar level studied a panel of 20 *Leptospira* reference strains representative of six species. The analysis was performed to supplement their microorganism MALDI-mass spectral database and samples were prepared by the ethanol-formic acid method described above. Using standard chemometrics tools included in commercially available MALDI-MS microorganism identification systems, the authors identified 20 distinct mass spectral signals, their combined pattern being responsible for the differentiation of 12 serovars of *Leptospira interrogans*. For the *L. borgpetersenii* species, discrimination at the strain level of three serovars was based on the unique pattern of five signals in their mass spectra.

Another example highlighting the limited success of the sample preparation step for strain differentiation is found in the work conducted by Huber et al. (Zeller-Peronnet et al. 2013) where 24 strains belonging to the species *Leuconostoc mesenteroides* and *L. pseudomesenteroides* were analyzed by MALDI-MS profiling. Discrimination of the protein profiles by principal component analysis (PCA) generated three distinct clusters, but only half of the microorganisms studied were reliably discriminated at the strain level. The protein profiles in this study were generated from samples prepared by initially subjecting the bacteria suspension to lysozyme digestion (37 °C, 30 min), followed by a standard protein extraction protocol. It was determined that subjecting Gram-positive bacteria to enzyme digestion with lysozyme provides additional lysing of the thick peptidoglycan layer of the cell wall, (Giebel et al. 2008) thus making intracellular proteins more accessible to the MALDI matrix. The authors found that the lysozyme treatment improved only the reproducibility of the profiles (about a 10% improvement in the correlation coefficient), but they did not observe signals at higher *m*/*z*'s reported in other studies (Giebel et al. 2008; Vargha et al. 2006). This discrepancy could be attributed to various experimental factors, including the analysis of different bacteria by each group. However, these results highlight potential issues of irreproducibility in inter-

laboratory studies and limitations of approaches that introduce biological reagents (enzymes) which are prone to biological activity losses and are affected by storage and experimental conditions (e.g., pH, ionic strength, time).

Finally, two examples from the literature illustrate the dependency of strain-level differentiation by MALDI-MS profiles on the type of microorganism being analyzed. In a study using MALDI-MS profiling to differentiate seafood-borne pathogens, (Boehme et al. 2013) authors found mixed levels of success for species-level and strain-level identification. Even though the goal of this investigation was to compare MALDI-MS with 16S rRNA sequencing for their ability to identify food-borne pathogens, the authors achieved species-level identification for all 120 bacterial strains tested with MALDI-MS, and in the case of *Bacillus subtilis* subsp. Spizizenii, subspecies-level classification was possible. Equally, in the analysis of beer-spoiling *Lactobacillus brevis* strains, Behr et al. (Kern et al. 2014) compiled 17 strains of *L. brevis* varying in their environmental source (e.g., brewery vs. sour-dough) and their ability to grow (and spoil) different beers (Lager, Pilsner, etc.). Samples were grown in standardized media and prepared by the ethanol-formic acid protein extraction protocol prior to MALDI-MS analysis. A set of highly reproducible signals allowed the successful assignment of 90 % of the mass spectra collected to the correct strain. Misclassifications were attributed to either highly similar mass spectral patterns or mass spectral patterns with low number of peaks; however, this set of microorganisms strains was always classified correctly for their ability to either strongly or weakly spoil beer.

Optimization of Growth Conditions In this strategy, differentiation of microorganisms at the strain level by MALDI-MS profiles is achieved with the aid of judiciously chosen set of growth conditions. In one such study, the effect of growth conditions on the ability of MALDI-MS profiling to differentiate acetic acid bacteria (AAB) at the strain level was investigated (Wieme et al. 2014). Investigators found that growth medium effects on the mass spectral profile do not affect differentiation at the species level, but rather enhance the level of differentiation at the strain level. For example, eight strains of *Gluconobacter oxydans* were grown in acetic acid medium (AAM), yeast–peptone–mannitol (YPM) agar, and glucose–yeast (GY) agar and their MALDI-mass spectra compared for shared and strain-specific peaks. The results showed that only 7 % of the peaks were consistently present in all mass spectra, regardless of the growth medium used. In addition, it was observed that the number of strain-specific peaks varied from 3–4 with different growth medium, although none were observed when bacteria were grown in GY agar. This approach presents an effective, yet relatively simple and economic way to differentiate a specific set of bacteria at the strain level, for example, antibiotic-resistant strains or enterohemorrhagic serotypes of *Escherichia coli* (vide infra). However, its universal applicability is limited since the effects of growth media on the mass spectral pattern cannot be predicted and thus this approach would require considerable testing and development prior to its implementation for each target microorganism.

Increasing the Reproducibility of the MALDI-MS Measurement

The existence of so-called hot (or sweet) signal spots within a MALDI sample has long been recognized and is mostly due to the inhomogeneous distribution of the analyte within this matrix/sample preparation, (Horneffer et al. 2001) especially when the sample is prepared by a manual/pipette dried-droplet method (i.e., sample deposited and dried first followed by matrix) (Dai et al. 1996). Accordingly, this variance can be mitigated through an increase in the number of measurements, either by averaging a large number of laser shots and/or increasing the number of MALDI sample preparations being analyzed (Szájli et al. 2008). These strategies are already being incorporated in studies aimed at the analysis of microorganisms with MALDI-MS, as most protocols collect about 20–40 replicates for each bacterial sample being analyzed, and average 200–400 laser shots/spectrum. As a result, these measurements most likely have reached the limit of reproducibility that can be achieved with standard manual sample preparation methods.

The heterogeneous nature of the MALDI matrix as well as the uneven analyte distribution within it have been identified as one of the major sources of variance in signal strength between spots of a single MALDI sample preparation. The effect of this uneven distribution on the variance of MALDI mass spectral profiles of bacteria was quantified by using a spray-based method to homogeneously deposit *E. coli* samples (suspended in a CHCA matrix solution) onto the MALDI plate (Fig. 2.6) (Toh-Boyo et al. 2012).

This approach resulted in bacteria being evenly distributed across the deposited sample *(Caution: the spray method is not suitable for clinical samples!)*. Subsequent MALDI-MS analyses of these homogenous sample/matrix preparations yielded highly reproducible mass spectra, regardless of the spatial coordinates of the laser shot on the sample. When compared to the manual/pipette dried-droplet

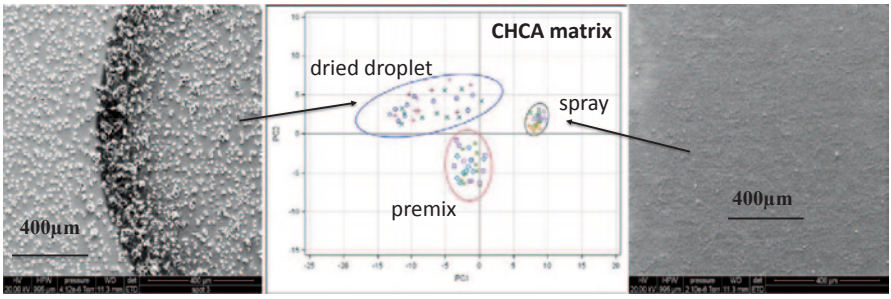


Fig. 2.6 PCA comparing the effect of matrix morphology on the reproducibility of bacteria MALDI-mass spectral profiles. All analyses were performed with *E. coli* (K-12) deposited by a spray-based method (uniform matrix deposition) and two manual pipette methods: dried droplet (simulating the direct transfer method) and premix (where a suspension of bacteria in matrix solution is deposited onto the MALDI plate). Ellipses represent the 95 % prediction space of the PCA clusters of replicate mass spectra for each deposition method (30 mass spectra/cluster). (See text for further details; Adapted from Toh-Boyo et al. 2012, copyright American Chemical Society)

method (equivalent to the direct transfer method), the spray method resulted in a 90% reduction of the total variance of the measurement. Most surprising was the resulting 60% reduction in the variance by the premix deposition method, where bacteria are suspended in the matrix solution prior to deposition, when compared to the manual dried-droplet method.

In principle, this increase in reproducibility would allow for the detection of small features and/or differences between the mass spectra of closely related microorganisms. Assuming the direct transfer method has a variance similar to that of the dried-droplet method used in this study (Fig. 2.6), it can be inferred that a significant reduction in the variance can be realized by simply premixing the intact bacteria with the matrix solution prior to its deposition onto the MALDI plate, rather than performing sequential depositions of bacteria followed by matrix. This improvement in the reproducibility would also be expected to be observed in the protein extraction protocols, adding only a single dilution step in the overall procedure (1:1 sample/matrix). However, to date, all standard protocols (Freiwald and Sauer 2009) used in profile-based MALDI-MS analyses implement a dried-droplet approach to deposit the sample onto the MALDI plate.

Increasing the Selectivity in Protein Biomarkers Detection

The profile-based MALDI-MS approach to identify and differentiate microorganisms, although powerful in the analysis at the species level, has met with mixed success at strain-level differentiation. This limitation stems in part from the fact that under the current experimental conditions ribosomal proteins are serendipitously detected, with the bulk observed in the mass range of 2000 to ~18,000 Da. For example, in *E. coli* there are 55 ribosomal proteins of known sequence that vary in MW from 4,400 to 61,200 Da (Wittmann 1982; Stelzl et al. 2001). Of these, about 80% (~44 proteins) have a MW within the detected mass range in the MALDI-MS measurement. However, this represents a small fraction of the available proteome in bacteria and thus severely limits the selectivity of the method. Case in point: it is known that *E. coli* has 4288 protein-coding genes or open-reading frames (ORF) (Blattner et al. 1997; Han and Lee 2006). Experimentally, however, two-dimensional (2D) gels and nongel methods have roughly identified about 1600 proteins in *E. coli,* (Corbin et al. 2003) with a MW range of 10^3–10^5 Da and *pI* values ranging from 4 to 12 (Han and Lee 2006). Therefore, it can be estimated that profile-based MALDI-MS measurements roughly detect only a small fraction of the possible protein pool in bacteria, or about 2.8% (44/1600) of the total detectable proteins in *E. coli*.

It is thus reasonable to argue that an increase in the selectivity of the method in order to detect a wider range of the microbial proteome will lead to the differentiation of closely related microorganism, as proteins responsible for unique genotypic traits would be detected. This point can be best illustrated with work by Murugaiyan and coworkers (2013), where they identified the protein expression levels between

pathogenic and nonpathogenic strains of the alga *Prototheca zopfii* (using 2D-gels and peptide mass fingerprinting, PMF, by MALDI-MS). Their results indicated that proteins responsible for differentiation of genotype 1 vs. 2 (nonpathogenic vs. pathogenic) were proteins related to energy, carbohydrate metabolism, and signal transduction (interestingly, ribosomal proteins remained unchanged between these two genotypes). One of the proteins upregulated in the pathogenic strain was a cyclic nucleotide-binding domain protein with an average MW of 51 kDa, a protein known to be associated in bacterial adaptation to a changing environment and well above the mass range usually analyzed in MALDI-MS profiling of microorganisms. Because the limited dynamic range in the proteins detected by MALDI-MS is due primarily to the ionization step (i.e., MALDI), merely extending the mass range of the TOF-MS will not overcome this limitation. Furthermore, approaches incorporating modified sample preparation protocols using additives and combinations of organic solvents have had limited and irreproducible success (vide supra).

In order to extend the MW dynamic range of detected proteins by MS (either ESI or MALDI), a separation step is often incorporated prior to detection. In the case of top-down (LeDuc et al. 2004; Zhou and Ning 2012) or bottom-up Zhang et al. (2013a) proteomics approaches, the incorporation of separation step is implemented in order to handle the highly complex protein and/or peptide mixture. In bottom-up proteomics-based measurement, the protein sample is digested into smaller peptides and the protein identity (and its biological source) inferred from the analysis of these peptides. In top-down proteomics-based measurements, the intact protein identity is derived directly from its analysis. In both cases, instrumentation capable of MS/MS measurements is required, although not all MS/MS instruments can perform both top-down and bottom-up analyses, and depending on the ionization mode and ion dissociation mode, the upper mass range limit is different (vide supra). Common to these proteomics measurements is the implementation of a separation step, mainly LC, prior to MS and MS/MS measurements.

Other top-down (Demirev et al. 2005; Fagerquist et al. 2010; Fagerquist et al. 2009; Fagerquist 2013) and bottom-up (Yao et al. 2002a, b) proteomic approaches *without* a separation step utilize MALDI in conjunction with a tandem MS or MS/MS—e.g., TOF/TOF-MS system in Fig. 2.3b—to increase the specificity of the analysis. These approaches do not extend the number of protein biomarkers detected, but rather they can analyze individual protein biomarker signals, and in principle a mixture of bacteria, provided representative signals from all species are generated during the MALDI process. This approach implements the same rapid protein extraction protocol as in the profile-based techniques.

Top-Down Proteomic Approaches

Strategies implementing LC-MS and LC-MS/MS can extend the number of proteins accessible in the microbial proteome for the purpose of strain-level differentiation. In top-down proteomics, the microbial sample is usually processed in

order to extract and isolate a highly enriched protein fraction. Top-down proteomic approaches can implement both MALDI and ESI, as both of these ionization modes can be interfaced with MS/MS instrumentation. Furthermore, both MALDI and ESI can be coupled with LC separation for additional selectivity in the analysis, albeit for MALDI the LC separation step is performed in an offline mode (Marcus et al. 2007; Bodnar et al. 2003; Basile et al. 2005; Maltman et al. 2011). The following discussion will focus on coupling LC with ESI-MS for the analysis of closely related microorganisms.

As stated earlier, ESI (and MALDI) is limited in its ability to ionize every component present in a complex mixture, as in crude bacterial protein extracts, and thus approaches aiming at increasing the dynamic range of proteins detected must incorporate a separation step (e.g., 2D gel, LC). However, strategies incorporating gel-based separation steps, although possessing large peak capacities, are time-consuming and are not amenable for high-throughput analyses of microbial samples. On the other hand, LC-based methods are suitable for high sample throughput, while providing the required selectivity to detect a wide range of proteins present in the sample. Because the analysis is usually carried out in a broad spectrum mode, that is, targeting all possible proteins, an aggressive lysis step is usually implemented. In addition, since both MALDI and ESI are sensitive to high concentrations of ionic species in solution (ESI in particular), physical methods are usually preferred (e.g., pressure, sonication, beads) that preclude the addition of high concentrations of lysis agents. In some instances the released DNA in solution is eliminated by the addition of DNase I. The final protein fraction must be devoid of any solid or suspended matter as it may obstruct/clog valves, syringes and frits present in the LC system. All these steps add to the overall sample preparation time, often requiring several hours of manual labor. Several compilations dealing with sample preparation methods for the analysis of proteins with LC-MS techniques can be found elsewhere, (Aguilar 2004; Shah and Gharbia 2010) and they will not be discussed in detail in this section.

Unlike MALDI-MS, the analysis of intact proteins by LC-(ESI)-MS requires additional processing steps of the mass spectral data in order to extract protein MW information. This is due to the multiple-charged nature of the mass spectral signals of proteins under ESI conditions. This process is best illustrated in the analysis of *Salmonella* spp. protein extraction by LC-ESI-MS. Figure 2.7 shows the initial LC-MS chromatogram, which is deconvoluted, in 30 s time windows, throughout the entire chromatographic period (e.g., 80 min in this example).

That is, the related m/z values from an eluting protonated protein (i.e., $[M+nH]^{n+}$) are converted to a single mass value for the neutral intact protein (i.e., M) (McFarland et al. 2014; Williams et al. 2002). The resulting mass profiles for each 30 s window are combined into a single profile showing peak intensity and mass of the protein (retention time information is also preserved during this data analysis, but not plotted). This intact protein expression profile is easy to interpret, since it represents all of the proteins detected in the sample.

Several advantages result from this approach. First, because LC is incorporated into the ESI-MS analysis, suppression effects are minimized leading to the

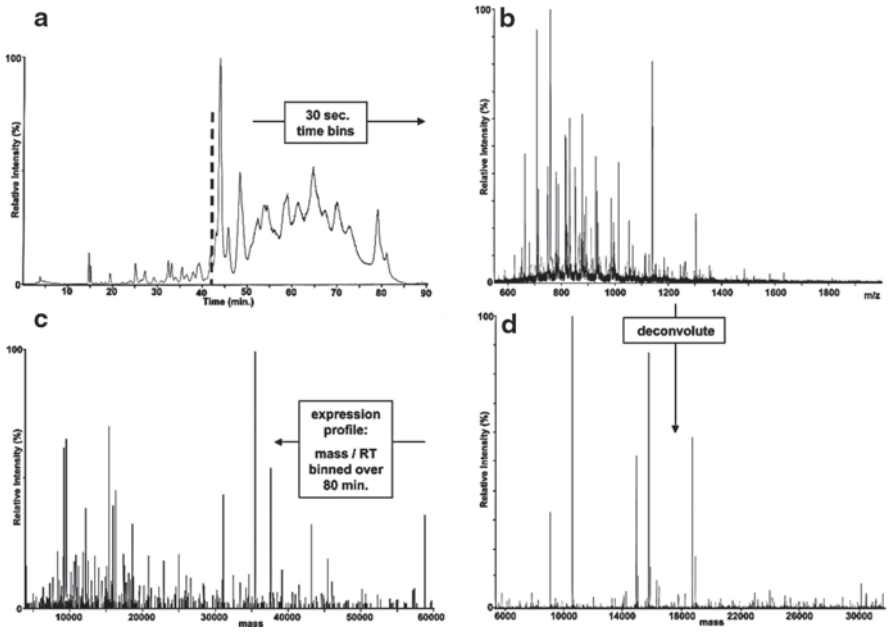


Fig. 2.7 Generation of protein mass profiles for *Salmonella* spp. using LC-ESI-MS. Multiple-charged signals from proteins within a 30 s time window are deconvoluted into single mass values (mass of neutral protein). The process is repeated for the entire chromatographic period (~80 min) and combined into a single mass profile. (Adapted from McFarland et al. 2014, copyright American Chemical Society)

detection of a larger number of proteins that otherwise would not be detected with MALDI-MS alone. Figure 2.8 illustrates this advantage where a comparison is made between the MALDI-mass profile and the LC-(ESI)-MS protein expression profile of *Shigella sonnei,* the latter showing enhanced detection of several proteins above 15 kDa (Everley et al. 2008). This approach was later used to successfully differentiate several strains and isolates of pathogenic and nonpathogenic *E. coli* (Mott et al. 2010).

A second advantage of the LC-ESI-MS approach is the increased mass accuracy of the measurement, making possible the detection of small mass differences between proteins. These mass differences can be equivalent to single-nucleotide polymorphisms (SNP) mutations or post-translational modifications. Recent work by McFarland and coworkers (2014) best illustrates the implementation of LC-(ESI)-MS (intact protein mass) and MS/MS top-down analyses to bacteria differentiation at the strain level. Proteins extracted from bacterial samples of *Salmonella typhimurium* (strain LT2) and *S. heidelberg* (strain A39) were first separated by reversed-phase (RP) LC and the eluent analyzed directly by ESI-MS using Q-TOF MS system (operated in the full-scan mode or MS). Following the data processing

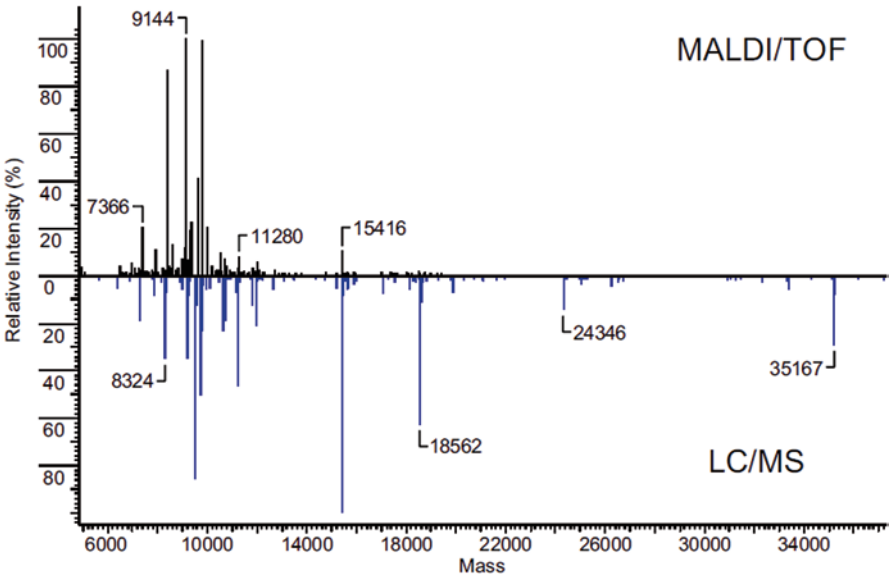


Fig. 2.8 Comparison of MALDI-MS profile (protein extraction; sinapinic acid matrix) with the LC-MS mass profile for *Shigella sonnei*. The *x*-axis represents the mass of the neutral proteins, M, and not the *m*/*z* of the $[M+nH]^{n+}$ ions in the ESI- or MALDI-mass spectra. (Adapted from Everley et al. 2008, with permission from Springer)

shown in Fig. 2.7, the resulting deconvoluted mass spectra were then displayed as intact protein expression profiles, in a mirrored configuration for easy comparison (Fig. 2.9). Although the similarity of these two serovars in terms of their protein expression profiles (mass maps) make them almost indistinguishable, close inspection reveals that several proteins showed detectable mass shifts (highlighted in Fig. 2.9) between the samples. These mass shifts most likely represent protein products of SNP containing genes that differentiate these two strains. The identification of proteins in Fig. 2.9 was accomplished in a second analysis by LC-MS/MS of the intact proteins with an orbitrap mass analyzer (Fig. 2.5).

A third advantage of this approach, when combined with top-down and bottom-up proteomic measurements, is the ability of "reverse engineering" unique segments within a protein sequence into polymerase chain reaction (PCR) primers that have specificity toward a desired phenotypic trait (Williams et al. 2005). This is especially useful for organisms with unsequenced genomes. Overall, the extended dynamic range of detected and identified intact proteins with LC-MS and MS/MS allows for an increase in the analysis selectivity among different microbial strains. However, development of these techniques for clinical analyses would have to include automated sample preparation and analysis in order to achieve a high sample throughput. In general, its implementation is expected to provide a flexible platform for enhanced discrimination of closely related microorganisms, including antibiotic-resistant strains, which will be presented next.

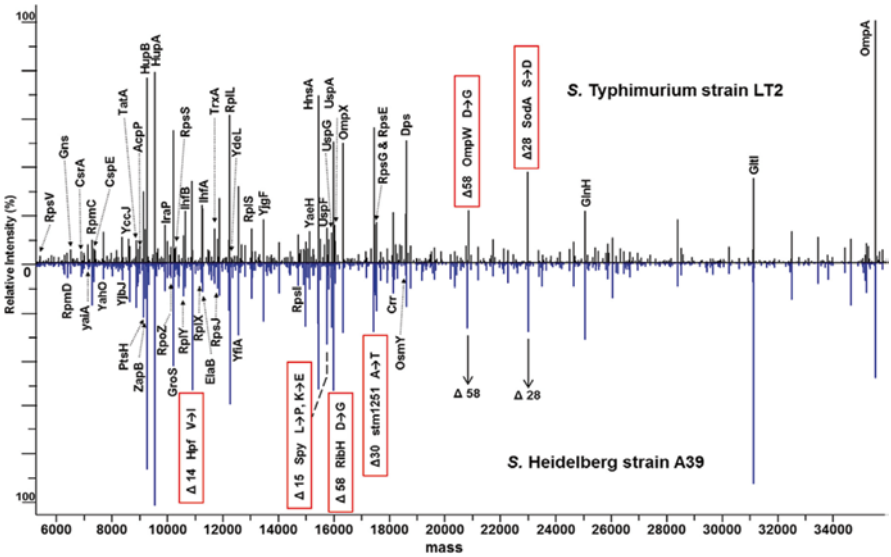


Fig. 2.9 Mass profiles obtained by LC-MS analysis of protein extracts of *S. typhimurium* strain LT2 and *S. heidelbergh* strain A39. The *x*-axis represents the mass of the neutral proteins, M, and not the *m*/*z* of the $[M+nH]^{n+}$ ions in the ESI-mass spectra. (Adapted from McFarland et al. 2014, copyright American Chemical Society)

Differentiation and Detection of Antibiotic-Resistant Bacteria

Clinically relevant is the detection of antibiotic-resistant strains, as accurate information about the infectious microorganism can result in improved patient outcomes and a concomitant reduction in health care costs. Knowledge of the resistance mechanism is key in order to develop accurate and specific MS-based methods for the detection of antibiotic (or antimicrobial)-resistant microorganisms. Bacteria can acquire antibiotic resistance via these five mechanisms: (i) mutation of the target site, (ii) enzymatic modification of the antibiotic, (iii) active efflux of antibiotics from the cell, (iv) restricted permeability of antibiotics to the active site, and (v) acquisition of an alternative metabolic pathway(s) insensitive to the antibiotic (Mc Dermott et al. 2003; Vranakis et al. 2014). Accordingly, several MS-based techniques have been developed to detect antibiotic resistance in bacteria derived by some or all of these mechanisms. These detection strategies include: (1) profile-based MALDI-MS of intact cells or protein extracts, (2) detection of the enzymatic modification of the antibiotic (a.k.a., enzyme activity methods), (3) direct detection of the enzyme degrading the antibiotic, and (4) profile-based MALDI-MS of intact cells grown in stable isotope/selective media. A brief discussion of each technique will follow with representative examples from the published literature.

Profile-Based MALDI-MS of Intact Cells or Protein Extracts In this approach commercial MALDI-MS systems are used to differentiate antibiotic-resistant strains

from their susceptible counterparts. The profiles generated from these two samples are usually indistinguishable in terms of unique biomarkers present (in the mass range of 2000 to 20,000 Da). That is, no unique biomarker(s) is responsible for antibiotic resistance differentiation within this mass range analyzed. This is understandable since most of the enzymes responsible for antibiotic resistance (carbapenemase, (Queenan and Bush 2007) vide infra) have MWs in the 28–31 kDa range. However, several studies have been successful at this differentiation by a careful control of the experimental conditions (Jackson et al. 2005; Goldstein et al. 2013) either using the direct transfer method (Goldstein et al. 2013; Majcherczyk et al. (2006) or the standard ethanol-wash and formic acid/acetonitrile protein extraction protocol, (Wybo et al. 2011; Griffin et al. (2012) while performing the MS analysis in the 2000–20,000 Da mass range. The method relies on the detection of subtle differences between mass spectral profiles of susceptible and resistant strains of the microorganisms, which must be present in the mass spectral database. For example, in the MALDI-MS differentiation of *Bacteroides fragilis* susceptible and resistant strains to the antibiotic meropenem, no unique signal was associated with resistance, but rather the two groups were discriminated based on the profiles of approximately 10 signals (Wybo et al. 2011). As a result, the profile-based MALDI-MS approach has shown very limited success as a universal method for antibiotic resistance detection and must be validated on a case-by-case basis. However, because this methodology is already in place in most clinical laboratories, it is expected that implementation of protein extraction protocols in conjunction with other procedures that reduce the variance of the measurement (vide supra) could lead to more confident differentiations between these closely related strains.

Enzyme Activity Via Antibiotic Degradation Product Detection This approach is based on monitoring the products of an enzyme-catalyzed reaction by MS (Gerber et al. 1999; Bothner et al. 2000; Gerber et al. 2001; Basile et al. 2002; Chennamaneni et al. 2014). The advantage of using MS over commonly used optical methods (e.g., UV-vis absorption or fluorescence) is that specific substrates labeled with chromophore molecules are not required, that is, substrates can be used in their native state, and products are differentiated from the substrate molecule simply by a characteristic mass change. In addition, given the resolution of most MS systems and a judicious choice of substrate molecules (MW's), multiple substrate–enzyme reactions can be monitored simultaneously with a single MS-based enzyme activity approach (Basile et al. 2002).

The application of MS-based enzyme activity to detect antibiotic-resistant bacteria was first reported by Hrabák et al. where carbapenemase activity was detected using MALDI-MS in viable intact cells of Enterobacteriaceae and *Pseudomonas* spp. (Hrabak et al. 2011) and later expanded to *Acinetobacter baumannii* (Hrabák et al. 2012). Carbapenemase activity detection in bacteria with MALDI-MS requires the incubation of viable bacteria (in a suitable buffer system, e.g., 20 mM Tris–HCl, 0.01 % sodium dodecyl sulfate (SDS), pH 7.0) with the substrate molecule, in this case, the antibiotic molecule (e.g., meropenem). This suspension is incubated for ~2 h at 35 °C, in which period meropenem molecules are enzymati-

cally hydrolyzed at the β-lactam moiety followed by decarboxylation. The bacteria suspension is then centrifuged and the supernatant is analyzed by MALDI-MS. The MALDI-MS analysis is performed in the mass range of 160–600 Da in order to detect the low MW products. The matrix 2,5-dihydroxybenzoic acid (DHB) was used as it has a very low chemical background in this low mass range. For example, for meropenem, the intact molecule (unmodified) is detected at *m*/*z* 384.16 (calculated monoisotopic mass of the $[M+H]^+$ ion), while the product (hydrolyzed and decarboxylated) is detected at *m*/*z* 358.18 (calculated monoisotopic mass of the $[M+H]^+$ ion; the sodium adduct ion, $[M+Na]^+$, is also detected at *m*/*z* 380.16). Mass spectra of susceptible strains of bacteria will show signals corresponding to the unreacted, intact antibiotic molecule, while resistant strains will show both the intact antibiotic and hydrolysis/decarboxylation product. Because the antibiotic molecules undergo spontaneous, but slow, hydrolysis (a.k.a., auto-hydrolysis), a background signal or measurement must be made in order to provide quantitative measure of the enzyme activity. This approach was successful in detecting antibiotic resistance from different carbapenemases (NDM-1, KPC-2, KPC-3, VIM-1, OXA-48, and OXA-162) in several microbial species (Hrabák et al. 2012).

Other laboratories have successfully implemented this technique to detect carbapenemase activity in several species of *Pseudomonas* and Enterobacteriaceae using the antibiotic (i.e., substrate) ertapenem (Burckhardt and Zimmermann 2011) and imipenem (Kempf et al. 2012; Alvarez-Buylla et al. 2013). A detailed characterization of this enzyme assay was conducted with *E. coli* cell lysate/extract (expressing chromosomally encoded AmpC β-lactamase) and using both a MALDI-TOF-MS and a MALDI-QQQ-MS in the MRM mode (see Fig. 2.4) to detect β-lactamase enzyme activity with the substrate penicillin G (PenG) (Hooff et al. 2012). The implementation of an SRM detection mode that allowed for accurate kinetic degradation studies was able to detect enzyme activity within 5–15 min of incubation time as well as establishing statistical parameters in terms of inter- and intraday reproducibility. A quantitative measurement of carbapenemase activity was also performed by incorporating a ^{18}O-labeled internal standard antibiotic molecule in conjunction with ESI-QQQ detection (where Q3 was a linear quadrupole ion trap, rather than a quadrupole mass filter), (Wang et al. 2013) increasing the accuracy, specificity and reproducibility of the measurement while at the same time reducing false positives due to auto-hydrolysis of the drug (substrate).

These analyses showed successful detection of carbapenem antibiotic resistance utilizing available instrumentation and without any special reagents (other than the antibiotic themselves). However, this analysis does require a separate sample preparation step, and instrumentation parameters, and thus may add to the overall sample preparation time and hinder high-throughput measurements. The ability to perform SRM measurements with these approaches can in principle provide a high-throughput platform for the detection of bacteria resistance toward multiple β-lactam drugs in a single measurement.

Direct Detection of Enzymes Responsible for Antibiotic Degradation/Resistance In this strategy antibiotic resistance in bacteria is established by the direct detection

of the carbapenem-hydrolyzing β-lactamase(s). Two studies will be described that demonstrate this strategy via a direct MALDI-MS measurement and a bottom-up proteomic approach.

Combined, MALDI with TOF-MS systems have a practical upper mass limit detection of ~60–160 kDa, depending on sample complexity and preparation, and so it is feasible to use MALDI-MS systems for the detection of proteins other than the ribosomal proteins used in profile-based identifications. Most carbapenemases have MW's that range between 28 and 31 kDa, and thus their detection by commercially available MALDI-MS is feasible. However, as mentioned earlier in this chapter, the prevalence of ribosomal proteins in the MALDI-mass spectrum profile of bacteria is a direct consequence of their abundance and ease of ionization. That is, detection of other proteins present in the sample may be suppressed by the presence of these ribosomal proteins. As a result, MALDI matrix and sample reparation conditions need to be optimized in order to detect higher MW proteins in these complex samples. Indeed, proof-of-principle work by Camara and Hays (2007) demonstrated that by optimizing the sample preparation and MALDI matrix it was possible to detect a 29 kDa β-lactamase in *E. coli*. The method used a sample preparation protocol that included a 0.1% TFA cell wash, followed by a protein extraction in formic acid/isopropyl alcohol/water (17:33:50 by vol.) and used sinapinic acid as the MALDI matrix (Wang et al. 1998). Although limited in scope and bacterial species analyzed, this study demonstrated the detection of higher MW proteins using a modified protein extraction protocol and without the use of LC prior to the MS analysis step. However, this approach may require optimization of the MALDI sample preparation step for each type of sample (e.g., Gram-type, genus, etc.), thus limiting its applicability as a universal detection protocol for known and unknown microbial samples.

The detection of the β-lactamase enzyme in bacteria was also accomplished via a bottom-up proteomic approach (Fig. 2.1) in order to provide increase specificity to the assay via protein identification. Two examples from the literature will be used to illustrate this approach. In the first report, Hu et al. detected β-lactam resistance in *Acinetobacter baumannii* (Chang et al. 2013) by implementing a bottom-up proteomic approach using a microwave heating-assisted trypsin digestion of the protein extract followed by RP LC-MS/MS. The authors were able to identify a unique tryptic peptide in all the β-lactam-resistant clinical isolates of *A. baumannii* tested. In addition, because the analysis was performed in a data-dependent mode, a protein distribution profile was also obtained that can be used to further classify the sample. In a second study involving bottom-up proteomics, Hensbergen and coworkers (Fleurbaaij et al. 2014) employed capillary electrophoresis (CE)-MS/MS (using a Qq-TOF-MS system) to detect antibiotic-resistant Gram-negative bacteria. A total 14 tryptic peptides unique to antibiotic-resistant bacteria were identified in this study, all derived from the OXA-48 and KPC carbapenemases.

Because these bottom-up proteomic approaches are conducted in a data-dependent mode, it is foreseeable that other β-lactamases can be identified. In addition, an increase in the detection and identification confidence level is accomplished by the detection of multiple peptides per protein. The discovery of unique peptide bio-

markers corresponding to β-lactamases could be used to develop targeted analyses using an MRM mode in a QQQ instrument for increase in specificity and rapid data analysis. However, given the stochastic nature of proteomic-based approaches, (Zhang et al. 2013b) in particular those involving a separation step and MS/MS detection, avoiding or decreasing the number of false-negative outcomes may require the use of internal standards, detection of multiple biomarkers within a measurement, and tests using selective growth media. This shortcoming will also require the analysis of several biological replicates, which unfortunately increases the analysis time mostly due to the added individual chromatographic steps.

Profile-Based MALDI-MS with Stable-Isotope/Selective Growth Media In this approach MALDI-MS profiles are obtained for bacteria grown in selective growth media containing antibiotic and with nutrients enriched with heavy (stable) isotopes of ^{13}C, or both ^{13}C and ^{15}N. Operationally, antibiotic-resistant bacteria would be able to grow in the presence of antibiotic, and during this process incorporate nutrients with the heavy isotopes, which are used in the biosynthesis of heavy homologs of protein biomarkers. By comparing these results with those from the same analysis performed in control media (i.e., natural isotope abundance) and without the antibiotic, mass shifts between these protein signals can be used to establish antibiotic resistance. Two approaches have been demonstrated using different growth media and data analysis/algorithm. In the first published study, bacteria were grown in 98% ^{13}C isotope-enriched media (and control media with natural isotope abundance) (Demirev et al. 2013). In a second approach, (Sparbier et al. 2013; Jung et al. 2014) samples were grown in media containing "heavy" lysine (^{13}C and ^{15}N enriched lysine) and control or "light" media (containing naturally occurring lysine). In both of these schemes, it is essential to measure a control mass spectrum to establish mass shifts of the proteins biosynthesized in the presence of heavy isotopes. The main advantage of these analyses is that they provide a universal method to detect *any form of antibiotic resistance* in bacteria.

Conclusions

The use of MS-based techniques for the differentiation of closely related microorganisms requires the close interplay of biochemical knowledge of the sample and the capabilities and requirements of MS hardware. The main information derived from MS analysis, m/z of the ions produced during the ionization process, provides a unique and specific set of biomarkers for the differentiation of microorganisms. Because of the wide acceptance of MALDI-TOF-MS systems by the microbiology community, it is reasonable to channel efforts at developing new analytical methods for the detection of closely related microorganism based on this platform, as is the case for enzyme activity (Hrabak et al. 2011) and stable-isotope media (Demirev et al. 2013) tests outlined above. In addition, the development of methods involving new sample preparation protocols and using MALDI-TOF-MS platforms is worth

pursuing, given the occurrence of this instrument in many laboratories. For example, efforts should continue to explore methodology intended to inactivate and/or extract proteins (Machen et al. 2013) that can enhance the ability to differentiate microorganisms at the strain level. Finally, approaches measuring metabolite (and their levels) should also be revisited and/or developed as new evidence points to the effect of antimicrobial drug resistance on the metabolomic phenotype of bacteria, yielding to the discovery of some unique metabolites (Derewacz et al. 2013). Their analysis can be performed with a wide range of accessible MS instrumentation, including MALDI-TOF-MS and bench-top ESI-MS systems.

References

Aguilar M-I. HPLC of peptides and proteins: methods and protocols. Totowa: Humana Press; 2004.

Alatoom AA, Cunningham SA, Ihde SM, Mandrekar J, Patel R. Comparison of direct colony method versus extraction method for identification of gram-positive cocci by use of bruker biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2011;49(8):2868–73.

Alvarez MTB, Shah DJ, Thulin CD, Graves SW. Tissue proteomics of the low-molecular weight proteome using an integrated cLC-ESI-QTOFMS approach. Proteomics. 2013;13(9):1400–11.

Alvarez-Buylla A, Picazo JJ, Culebras E. Optimized method for *Acinetobacter* species carbapenemase detection and identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2013 51:1589–92.

Anhalt JP, Fenselau C. Identification of bacteria using mass spectrometry. Anal Chem. 1975:47(2):219–25.

Basile F. Rapid sample preparation for microorganism analysis by mass spectrometry. In Fenselau C, Demirev P, Editors. Rapid characterization of microorganisms by mass spectrometry. ACS symposium series; American Chemical Society: Washington, DC. vol. 1065. 2011. pp 5–34.

Basile F, Ferrer I, Furlong ET, Voorhees KJ. Simultaneous multiple substrate tag detection with ESI-Ion trap MS for in vivo bacterial enzyme activity profiling. Anal Chem. 2002;74(16):4290–3.

Basile F, Kassalainen GE, Ratanathanawongs Williams SK. Interface for direct and continuous sample–matrix deposition onto a MALDI probe for polymer analysis by thermal field flow fractionation and off-line MALDI-MS. Anal Chem. 2005;77(9):3008–12.

Benkali K, Marquet P, Rérolle JP, Meur YL, Gastinel LN. A new strategy for faster urinary biomarkers identification by Nano-LC-MALDI-TOF/TOF mass spectrometry. BMC Genomics. 2008;9(1):541.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. The complete genome sequence of *Escherichia coli* K-12. Science. 1997;277(5331):1453–62.

Bodnar WM, Blackburn RK, Krise JM, Moseley MA. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. J Am Soc Mass Spectrom. 2003;14(9):971–9.

Boehme K, Fernandez-No IC, Pazos M, Gallardo JM, Barros-Velazquez J, Canas B, Calo-Mata P. Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing versus MALDI-TOF MS Fingerprinting. Electrophoresis. 2013;34:877–87.

Boon JJ, Boer WRD, Kruyssen FJ, Wouters JTM. Pyrolysis mass spectrometry of whole cells, cell walls and isolated cell wall polymers of *Bacillus subtilis* var. Niger WM. J Gen Microbiol. 1981;122(1):119–27.

Bothner B, Chavez R, Wei J, Strupp C, Phung Q, Schneemann A, Siuzdak G. Monitoring enzyme catalysis with mass spectrometry. J Biol Chem. 2000;275(18):13455–9.

Boyd RK, Basic C, Bethem RA. Measurement, dimensions and units. In trace quantitative analysis by mass spectrometry. New Jersey: Wiley; 2008. pp. 6–8.

Bruins AP. Mechanistic aspects of electrospray ionization. J Chromatogr A. 1998;794(1):345–57.

Burckhardt I, Zimmermann S. Using matrix-assisted laser desorption ionization-time of flight mass spectrometry to detect carbapenem resistance within 1 to 2.5 hours. J Clin Microbiol. 2011;49:3321–4.

Calderaro A, Piccolo G, Montecchini S, Buttrini M, Gorrini C, Rossi S, Arcangeletti MC, De Conto F, Medici MC, Chezzi C. MALDI-TOF ms analysis of human and animal *Brachyspira* species and benefits of database extension. J. Proteomics. 2013;78:273–80.

Calderaro A, Piccolo G, Gorrini C, Montecchini S, Buttrini M, Rossi S, Piergianni M, De CF, Arcangeletti MC, Chezzi C, Medici MC. *Leptospira* species and serovars identified by MALDI-TOF mass spectrometry after database implementation. BMC Res Notes. 2014;7:330.

Camara JE, Hays FA. Discrimination between wild-type and ampicillin-resistant *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Bioanal Chem. 2007;389:1633–8.

Cannon J, Lohnes K, Wynne C, Wang Y, Edwards N, Fenselau C. High-throughput middle-down analysis using an orbitrap. J Proteome Res. 2010;9(8):3886–90.

Cech NB, Enke CG. Practical implications of some recent studies in electrospray ionization fundamentals. Mass Spectrom Rev. 2001;20(6):362–87.

Chang C-J, Lin J-H, Chang K-C, Lai M-J, Rohini R, Hu A. Diagnosis of B-Lactam resistance in *Acinetobacter baumannii* using shotgun proteomics and LC-Nano-Electrospray ionization ion trap mass spectrometry. Anal Chem. 2013;85:2802–8.

Chennamaneni NK, Kumar AB, Barcenas M, Spáčil Z, Scott CR, Tureček F, Gelb MH. Improved reagents for newborn screening of mucopolysaccharidosis Types I, II, and VI by tandem mass spectrometry. Anal Chem. 2014;86(9):4508–14.

Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. 2013;26(3):547–603.

Cohen SL, Chait BT. Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. Anal Chem. 1996;68(1):31–7.

Cole RB. Electrospray mass spectrometry: fundamentals, instrumentation, and applications. Oxford: Wiley-Blackwell; 2008.

Corbin RW, Paliy O, Yang F, Shabanowitz J, Platt M, Lyons CE, Root K, McAuliffe J, Jordan MI, Kustu S, Soupene E, Hunt DF. Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. Proc Natl Acad Sci. 2003;100(16):9232–7.

Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012;36(2):380–407.

Dai Y, Whittal RM, Li L. Confocal fluorescence microscopic imaging for investigating the analyte distribution in MALDI matrices. Anal Chem. 1996;68(15):2494–500.

DeLuca S, Sarver EW, Harrington P de B, Voorhees KJ. Direct analysis of bacterial fatty acids by curie-point pyrolysis tandem mass spectrometry. Anal Chem. 1990;62(14):1465–72.

DeLuca SJ, Sarver EW, Voorhees KJ. Direct analysis of bacterial glycerides by curie-point pyrolysis-mass spectrometry. J Anal Appl Pyrolysis 1992;23(1):1–14.

Demirev PA, Lin JS, Pineda FJ, Fenselau C. Bioinformatics and mass spectrometry for microorganism identification: proteome-wide post-translational modifications and database search algorithms for characterization of Intact H. Pylori. Anal Chem. 2001;73(19):4566–73.

Demirev PA, Feldman AB, Kowalski P, Lin JS. Top-down proteomics for rapid identification of intact microorganisms. Anal Chem. 2005;77(22):7455–61.

Demirev PA, Hagan NS, Antoine MD, Lin JS, Feldman AB. Establishing drug resistance in microorganisms by mass spectrometry. J Am Soc Mass Spectrom. 2013;24(8):1194–201.

Deng J, Fu L, Wang R, Yu N, Ding X, Jiang L, Fang Y, Jiang C, Lin L, Wang Y, Che X. Comparison of MALDI-TOF MS, gene sequencing and the Vitek 2 for identification of seventy-three clinical isolates of enteropathogens. J Thorac Dis. 2014;6(5):539–44.

Derewacz DK, Goodwin CR, McNees CR, McLean JA, Bachmann BO. Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism. Proc Natl Acad Sci USA. 2013;110:2336–41 (S 2336/1–S 2336/188).

Everley RA, Mott TM, Wyatt SA, Toney DM, Croley TR. Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species. J Am Soc Mass Spectrom. 2008;19(11):1621–28.

Fagerquist CK. Top-down proteomic identification of bacterial protein biomarkers and toxins using MALDI-TOF-TOF-MS/MS and post-source decay. Rev Anal Chem. 2013;32(2):127–33.

Fagerquist CK, Garbus BR, Williams KE, Bates AH, Boyle S, Harden LA. Web-based software for rapid top-down proteomic identification of protein biomarkers, with implications for bacterial identification. Appl Environ Microbiol. 2009;75(13):4341–53.

Fagerquist CK, Garbus BR, Miller WG, Williams KE, Yee E, Bates AH, Boyle S, Harden LA, Cooley MB, Mandrell RE. Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight–time-of-flight mass spectrometry and top-down proteomics. Anal Chem. 2010;82(7):2717–25.

Fang M. Tandem mass spectrometry-based proteomics, protein characterisation and biomarker discovery in microorganisms. In Shah, H., Gharbia, S., Editors. Mass Spectrometry for Microbial Proteomics;, Wiley: Hoboken; 2010. pp. 75–6.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science 1989;246(4926):64–71.

Fleurbaaij F, Heemskerk AAM, Russcher A, Klychnikov OI, Deelder AM, Mayboroda OA, Kuijper EJ, van Leeuwen HC, Hensbergen PJ. Capillary-electrophoresis mass spectrometry for the detection of carbapenemases in (Multi-)drug-resistant gram-negative bacteria. Anal Chem. 2014;86(18):9154–61.

Freiwald A, Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry. Nat Protoc. 2009;4(5):732–42.

Gerber SA, Scott CR, Turecek F, Gelb MH. Analysis of rates of multiple enzymes in cell lysates by electrospray ionization mass spectrometry. J Am Chem Soc. 1999;121(5):1102–3.

Gerber SA, Scott CR, Tureček F, Gelb MH. Direct profiling of multiple enzyme activities in human cell lysates by affinity chromatography/electrospray ionization mass spectrometry: application to clinical enzymology. Anal Chem. 2001;73(8):1651–57.

Giebel RA, Fredenberg W, Sandrin TR. Characterization of environmental isolates of *Enterococcus* spp. by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Water Res. 2008;42(4–5):931–40.

Goldstein JE, Zhang L, Borror CM, Rago JV, Sandrin, TR. Culture conditions and sample preparation methods affect spectrum quality and reproducibility during profiling of *Staphylococcus aureus* with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Lett Appl Microbiol. 2013;57(2):144–50.

Goodacre R, Timmins éadaoin M, Burton R, Kaderbhai N, Woodward AM, Kell DB, Rooney PJ. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. Microbiology. 1998;144(5):1157–70.

Griffin PM, Price GR, Schooneveldt JM, Schlebusch S, Tilse MH, Urbanski T, Hamilton B, Venter D. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. J Clin Microbiol. 2012;50:2918–31.

Gross JH. Mass spectrometry a textbook. 2nd ed. Berlin: Springer; 2011.

Guckert JB, Hood MA, White DC. Phospholipid ester-linked fatty acid profile changes during nutrient deprivation of *Vibrio cholerae*: increases in the trans/cis ratio and proportions of cyclopropyl fatty acids. Appl Environ Microbiol. 1986;52(4):794–801.

Han M-J, Lee SY. The *Escherichia coli* proteome: past, present, and future prospects. Microbiol Mol Biol Rev. 2006;70(2):362–439.

Hillion M, Mijouin L, Jaouen T, Barreau M, Meunier P, Lefeuvre L, Lati E, Chevalier S, Feuilloley MGJ. Comparative study of normal and sensitive skin aerobic bacterial populations. Microbiol Open 2013;2:953–61.

Holland RD, Wilkes JG, Rafii F, Sutherland JB, Persons CC, Voorhees KJ, Lay JO. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1996;10(10):1227–32.

Holland RD, Duffy CR, Rafii F, Sutherland JB, Heinze TM, Holder CL, Voorhees KJ, Lay Jr JO. Identification of bacterial proteins observed in MALDI TOF mass spectra from whole cells. Anal Chem. 1999;71(15):3226–30.

Holland RD, Wilkes JG, Cooper WM, Alusta P, Williams A, Pearce B, Beaudoin M, Buzatu D. Thymol treatment of bacteria prior to matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis aids in identifying certain bacteria at the subspecies level. Rapid Commun Mass Spectrom. 2014;28(23):2617–26.

Hooff GP, van Kampen JJA, Meesters RJW, van Belkum, A, Goessens, WHF, Luider TM. Characterization of B-Lactamase enzyme activity in bacterial lysates using MALDI-mass spectrometry. J Proteome Res. 2012;11(1):79–84.

Horneffer V, Forsmann A, Strupat, K, Hillenkamp, F; Kubitscheck, U. Localization of analyte molecules in MALDI preparations by confocal laser scanning microscopy. Anal Chem. 2001;73(5):1016–22.

Horneffer V, Haverkamp J, Janssen H-G, Notz R. MALDI-TOF-MS analysis of bacterial spores: wet heat-treatment as a new releasing technique for biomarkers and the influence of different experimental parameters and microbiological handling. J Am Soc Mass Spectrom. 2004;15(10):1444–54.

Hrabak J, Walkova R, Studentova V, Chudackova E, Bergerova T. Carbapenemase activity detection by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 2011;49:3222–7.

Hrabák J, Študentová V, Walková R, Žemličková H, Jakubů V, Chudáčková E, Gniadkowski M, Pfeifer Y, Perry JD, Wilkinson K, Bergerová, T. Detection of NDM-1, VIM-1, KPC, OXA-48, and OXA-162 carbapenemases by Matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50(7):2441–3.

Hrabak J, Chudackova E, Walkova R. Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. Clin Microbiol Rev. 2013;26:103–14.

Hsieh S-Y, Tseng C-L, Lee Y-S, Kuo A-J, Sun C-F, Lin Y-H, Chen J-K. Highly efficient classification and identification of human pathogenic bacteria by MALDI-TOF MS. Mol Cell Proteomics 2008;7(2):448–56.

Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The orbitrap: a new mass spectrometer. J Mass Spectrom. 2005;40(4):430–43.

Huff SM, Matsen JM, Windig, W, Meuzelaar HLC. Pyrolysis mass spectrometry of bacteria from infected human urine. I. Influence of culturing and antibiotics. Biol Mass Spectrom. 1986;13(6):277–86.

Jackson KA, Edwards-Jones V, Sutton CW, Fox AJ. Optimisation of intact cell MALDI method for fingerprinting of methicillin-resistant *Staphylococcus aureus*. J Microbiol Methods 2005;62(3):273–84.

Jaskolla TW, Karas M. Compelling evidence for lucky survivor and gas phase protonation: the unified MALDI analyte protonation mechanism. J Am Soc Mass Spectrom. 2011;22(6):976–88.

Jung JS, Eberl T, Sparbier K, Lange C, Kostrzewa M, Schubert S, Wieser A. Rapid detection of antibiotic resistance based on mass spectrometry and stable isotopes. Eur J Clin Microbiol Infect Dis. 2014;33:949–55.

Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem. 1988;60(20):2299–301.

Karlsson C, Malmström L, Aebersold R, Malmström J. Proteome-wide selected reaction monitoring assays for the human pathogen streptococcus pyogenes. Nat Commun. 2012;3:1301.

Karpanoja P, Harju I, Rantakokko-Jalava K, Haanpera M, Sarkkinen H. Evaluation of two matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of viridans group streptococci. Eur J Clin Microbiol Infect Dis. 2014;33:779–88.

Kempf M, Bakour S, Flaudrops C, Berrazeg M, Brunel J-M, Drissi M, Mesli E, Touati A, Rolain J-M. Rapid detection of carbapenem resistance in acinetobacter baumannii using matrix-assisted laser desorption ionization-time of flight mass spectrometry. PLoS ONE 2012;7:e31676.

Kern CC, Vogel RF, Behr J. Differentiation of *Lactobacillus brevis* strains using matrix-assisted-laser-desorption-ionization-time-of-flight mass spectrometry with respect to their beer spoilage potential. Food Microbiol. 2014;40:18–24.

Kierzkowska M, Majewska A, Kuthan RT, Sawicka-Grzelak A, Mlynarczyk G. A Comparison of Api 20A vs. MALDI-TOF MS for routine identification of clinically significant anaerobic bacterial strains to the species level. J Microbiol Methods 2013;92:209–212.

Kolecka A, Khayhan K, Groenewald M, Theelen B, Arabatzis M, Velegraki A, Kostrzewa M, Mares M, Taj-Aldeen SJ; Boekhout T. Identification of medically relevant species of arthroconidial yeasts by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2013;51:2491–500.

Laiko VV, Baldwin MA, Burlingame AL. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem. 2000;72(4):652–7.

Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. 2008;4:222.

Lartigue M-F. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry for bacterial strain characterization. Infect Genet Evol. 2013;13:230–5.

Lasch P, Nattermann H, Erhard M, Stämmler M, Grunow R, Bannert N, Appel B, Naumann D. MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. Anal Chem. 2008;80(6):2026–2034.

LeDuc RD, Taylor GK, Kim Y-B, Januszyk TE, Bynum LH, Sola JV, Garavelli JS, Kelleher NL. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. Nucleic Acids Res. 2004;32(suppl 2):W340–W345.

Li Y, Gu B, Liu G, Xia W, Fan K, Mei Y, Huang P, Pan S. MALDI-TOF MS versus VITEK 2 ANC card for identification of anaerobic bacteria. J Thorac Dis. 2014;6:517–23.

Liu H, Du Z, Wang J, Yang R. Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2007;73(6):1899–907.

Londry FA, Hager JW. Mass selective axial ion ejection from a linear quadrupole ion trap. J Am Soc Mass Spectrom. 2003;14(10):1130–47.

Machen A, Kobayashi M, Connelly MR, Wang YF (Wayne). Comparison of heat inactivation and cell disruption protocols for identification of mycobacteria from solid culture media by use of vitek matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2013;51(12):4226–9.

Madonna AJ, Basile F, Ferrer I, Meetani MA, Rees JC, Voorhees KJ. On-probe sample pretreatment for the detection of proteins above 15 KDa from whole cell bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2000;14(23):2220–9.

Madonna AJ, Voorhees KJ, Taranenko NI, Laiko VV, Doroshenko VM. Detection of cyclic lipopeptide biomarkers from bacillus species using atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem. 2003;75(7):1628–37.

Majcherczyk PA, McKenna T, Moreillon P, Vaudaux P. The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*. FEMS Microbiol Lett. 2006;255:233–9.

Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem. 2000;72(6):1156–62.

Maltman DJ, Brand S, Belau E, Paape R, Suckau D, Przyborski SA. Top-down Label-Free LC-MALDI analysis of the peptidome during neural progenitor cell differentiation reveals com-

plexity in cytoskeletal protein dynamics and identifies progenitor cell markers. Proteomics 2011;11(20):3992–4006.
Marcus K, Schäfer H, Klaus S, Bunse C, Swart R, Meyer HE. A new fast method for nanoLC–MALDI-TOF/TOF–MS analysis using monolithic columns for peptide preconcentration and separation in proteomic studies. J Proteome Res. 2007;6(2):636–43.
Martinez AR, Abranches J, Kajfasz JK, Lemos JA. Characterization of the *Streptococcus sobrinus* acid-stress response by interspecies microarrays and proteomics. Mol Oral Microbiol. 2010;25(5):331–42.
Mc Dermott PF, Walker RD, White DG. Antimicrobials: modes of action and mechanisms of resistance. Int J Toxicol. 2003;22(2):135–43.
McFarland MA, Andrzejewski D, Musser SM, Callahan JH. Platform for identification of *Salmonella* serovar differentiating bacterial proteins by top-down mass spectrometry: *S. Typhimurium* vs *S. Heidelberg*. Anal Chem. 2014;86(14):6879–86.
McLafferty FW, Tureek F. Interpretation of mass spectra. 4th ed. University Science Books. Calif: Mill Valley; 1993.
Meuzelaar HLC, Kistemaker PG. Technique for fast and reproducible fingerprinting of bacteria by pyrolysis mass spectrometry. Anal Chem. 1973;45(3):587–90.
Mitchell P. Proteomics retrenches. Nat Biotechnol. 2010;28(7):665–70.
Mott TM, Everley RA, Wyatt SA, Toney DM, Croley TR. Comparison of MALDI-TOF/MS and LC-QTOF/MS methods for the identification of enteric bacteria. Int J Mass Spectrom. 2010;291(1–2):24–32.
Murphy RC, Fiedler J, Hevko J. Analysis of nonvolatile lipids by mass spectrometry. Chem Rev. 2001;101(2):479–526.
Murugaiyan J, Weise C, von BM, Roesler U. Two-dimensional proteome reference map of *Prototheca zopfii* revealed reduced metabolism and enhanced signal transduction as adaptation to an infectious life style. Proteomics 2013;13:2664–9.
Neubert H, Halket JM, Ocaña MF, Patel RKP. MALDI post-source decay and LIFT-TOF/TOF investigation of A-Cyano-4-Hydroxycinnamic acid cluster interferences. J Am Soc Mass Spectrom. 2004;15(3):336–43.
Nguyen S, Fenn JB. Gas-Phase ions of solute species from charged droplets of solutions. Proc Natl Acad Sci. 2007;104(4):1111–7.
Pereira L, Dias N, Santos C, Lima N. The use of MALDI-TOF ICMS as an alternative tool for *Trichophyton rubrum* identification and typing. Enferm Infecc Microbiol Clin. 2014:32:11–7.
Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods. 2012;9(6):555–66.
Price P. Standard definitions of terms relating to mass spectrometry. J Am Soc Mass Spectrom. 1991;2(4):336–48.
Prieto MC. Erratum to: MALDI-TOF-MS analysis of bacterial spores: wet heat-treatment as a new releasing technique for biomarkers and the influence of different experimental parameters and microbiological handling. J Am Soc Mass Spectrom. 2006;17(9):1323–3.
Queenan AM, Bush K. Carbapenemases: the versatile -lactamases. Clin Microbiol Rev. 2007;20(3):440–58.
Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal Chem. 2001;73(4):746–50.
Sandrin TR, Goldstein JE; Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom Rev. 2013;32(3):188–217.
Schaaff TG, Cargile BJ, Stephenson JL, McLuckey SA. Ion trap collisional activation of the $(M+2H)^{2+}$–$(M+17H)^{17+}$Ions of human hemoglobin B-Chain. Anal Chem. 2000;72(5):899–907.
Schulthess B, Bloemberg GV, Zbinden R, Bottger EC, Hombach M. Evaluation of the Bruker MALDI biotyper for identification of gram-positive rods: development of a diagnostic algorithm for the clinical laboratory. J Clin Microbiol. 2014;52(4):1089–97.
Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49(4):543–51.

Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, Mullen C, Earley L, Hardman M, Blethrow JD, Bui H, Specht A, Lange O, Denisov E, Makarov A, Horning S, Zabrouskov V. Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. Anal Chem. 2013;85(24):11710–4.

Sparbier K, Lange C, Jung J, Wieser A, Schubert S, Kostrzewa M. MALDI biotyper-based rapid resistance detection by stable-isotope labeling. J Clin Microbiol. 2013;51(11):3741–8.

Stelzl U, Connell S, Nierhaus KH, Wittmann-Liebold B. Ribosomal proteins: role in ribosomal functions. eLS; John Wiley & Sons, Ltd., 2001. pp.1–3.

Suarez S, Ferroni A, Lotz A, Jolley KA, Guerin P, Leto J, Dauphin B, Jamet A, Maiden MCJ, Nassif X, Armengaud J. Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. J Microbiol Methods. 2013;94:390–6.

Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal Bioanal Chem. 2003;376(7):952–65.

Szájli E, Fehér T, Medzihradszky KF. Investigating the quantitative nature of MALDI-TOF MS. Mol Cell Proteomics. 2008;7(12):2410–18.

Tamura H, Hotta Y, Sato H. Novel accurate bacterial discrimination by MALDI-Time-of-Flight MS based on ribosomal proteins coding in S10-Spc-Alpha operon at strain level S10–GERMS. J Am Soc Mass Spectrom. 2013;24(8):1185–93.

Toh-Boyo GM, Wulff SS, Basile F. Comparison of sample preparation methods and evaluation of intra- and intersample reproducibility in bacteria MALDI-MS profiling. Anal Chem. 2012;84(22):9971–80.

Vargha M, Takáts Z, Konopka A, Nakatsu CH. Optimization of MALDI-TOF MS for strain level differentiation of arthrobacter isolates. J Microbiol Methods 2006;66(3):399–409.

Vranakis I, Goniotakis I, Psaroulaki A, Sandalakis V, Tselentis Y, Gevaert K; Tsiotis G. Proteome studies of bacterial antibiotic resistance mechanisms. J. Proteomics 2014;97:88–99.

Wang M, Shen Y, Turko IV, Nelson DC, Li S. Determining carbapenemase activity with 18O labeling and targeted mass spectrometry. Anal Chem Wash DC US. 2013;85:11014–11019.

Wang Z, Russon L, Li L, Roser DC, Long SR. Investigation of spectral reproducibility in direct analysis of bacteria proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1998;12(8):456–64.

Werner G, Fleige C, Fessler AT, Timke M, Kostrzewa M, Zischka M, Peters T, Kaspar H, Schwarz S. Improved identification including MALDI-TOF mass spectrometry analysis of Group D streptococci from bovine mastitis and subsequent molecular characterization of corresponding *Enterococcus faecalis* and *Enterococcus faecium* isolates. Vet Microbiol. 2012;160:162–9.

Wieme AD, Spitaels F, Aerts M, De Bruyne K, Van Landschoot A, Vandamme P. Effects of growth medium on matrix-assisted laser desorption-ionization time of flight mass spectra: a case study of acetic acid bacteria. Appl Environ Microbiol. 2014;80:1528–38, 12

Williams TL, Leopold P, Musser S. Automated postprocessing of electrospray LC/MS data for profiling protein expression in bacteria. Anal Chem. 2002;74(22):5807–13.

Williams TL, Monday SR, Edelson-Mammel S, Buchanan R, Musser SM. A top-down proteomics approach for differentiating thermal resistant strains of *Enterobacter sakazakii*. Proteomics 2005;5(16):4161–9.

Wilm M. Principles of electrospray ionization. Mol Cell Proteomics 2011;10(7):M111.009407–M111.009407.

Wittmann HG. Components of bacterial ribosomes. Annu Rev Biochem. 1982;51(1):155–83.

Wybo I, Bel AD, Soetens O, Echahidi F, Vandoorslaer K, Cauwenbergh MV, Piérard D. Differentiation of cfiA-Negative and cfiA-positive *Bacteroides fragilis* isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2011;49(5):1961–4.

Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. J. Phys. Chem. 1984; 88(20): 4451–9

Yao Z-P, Afonso C, Fenselau C. Rapid microorganism identification with on-slide proteolytic digestion followed by matrix-assisted laser desorption/ionization tandem mass spectrometry and database searching. Rapid Commun Mass Spectrom. 2002a;16(20):1953–6.

Yao Z-P, Demirev PA, Fenselau C. Mass spectrometry-based proteolytic mapping for rapid virus identification. Anal Chem. 2002b;74(11):2529–34.

Yergey AL, Coorssen JR, Backlund PS, Blank PS, Humphrey GA, Zimmerberg J, Campbell JM, Vestal ML. De novo sequencing of peptides using MALDI/TOF-TOF. J Am Soc Mass Spectrom. 2002;13(7):784–91.

Zeller-Peronnet V, Brockmann E, Pavlovic M, Timke M, Busch, U, Huber I. Potential and limitations of MALDI-TOF MS for discrimination within the species *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides*. J Fuer Verbraucherschutz Leb. 2013;8:205–14.

Zenobi R, Knochenmuss R. Ion Formation in MALDI mass spectrometry. Mass Spectrom Rev. 1998;17(5):337–66.

Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR. Protein analysis by shotgun/bottom-up proteomics. Chem Rev. 2013a;113(4):2343–94.

Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR. Protein analysis by shotgun/bottom-up proteomics. Chem Rev. 2013b;113(4):2343–94.

Zhou H, Ning Z, E. Starr A, Abu-Farha M, Figeys D. Advancements in top-down proteomics. Anal Chem. 2012;84(2):720–34.

Zhu B, Xiao D, Zhang H, Zhang Y, Gao Y, Xu L, Lv J, Wang Y, Zhang J, Shao Z. MALDI-TOF MS distinctly differentiates nontypable *Haemophilus influenzae* from *Haemophilus haemolyticus*. PLoS ONE. 2013;8:e56139.

Zubarev RA, Makarov A. Orbitrap mass spectrometry. Anal Chem. 2013;85(11):5288–96.

Chapter 3
Sample Preparation Methods for the Rapid MS Analysis of Microorganisms

Shobha Devi, Anren Hu and Yen-Peng Ho

Sample preparation is the most critical steps in microbial analysis to generate accurate and informative data. Well-designed methods are required to enable unambiguous and sensitive identification of microbial cells/biomarkers from complex sample mixtures. However, there is no standard protocol for sample preparation because the microbial samples are complex. Moreover, sample preparation strategies depend on the type, source, abundance, and physical properties of samples. Traditional procedures have been used to collect, isolate, and identify pathogens from different samples. Furthermore, biochemical, serological, and molecular biology methods have been employed for the definitive identification of microbial isolates. These established methods are often time-consuming and labor-intensive. Mass spectrometry (MS) has become one of the main tools to accomplish the rapid identification and quantification of microbial cells/biomarkers. To simplify the complexity of the samples and improve the detection of low-abundance microbial cells/biomarkers, sample preparation methods that can selectively enrich target analytes and simultaneously eliminate interferences are greatly desired prior to MS analysis. In response to this challenge, numerous sample treatment techniques have been developed. In this chapter, we describe the sample preparation methods for the identification of microorganisms by MS, including techniques of enrichment, cell lysis, and separation of microbial cells/biomarkers. Two types of procedures for the microbial sample preparation can be used for the MS identification. In one sample preparation procedure, intact cells in the form of suspension are analyzed after they are enriched or cultivated. In the other procedure, extraction, enrichment, and/or separation of biomarkers from the cells are carried out prior to MS analysis. The workflow of these approaches is outlined in Fig. 3.1.

Y.-P. Ho (✉) · S. Devi
Department of Chemistry, National Dong Hwa University, 97401 Hualien, Taiwan
e-mail: ypho@mail.ndhu.edu.tw

A. Hu
Institute of Medical Biotechnology, Tsu-Chi University, Hualien, Taiwan

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_3

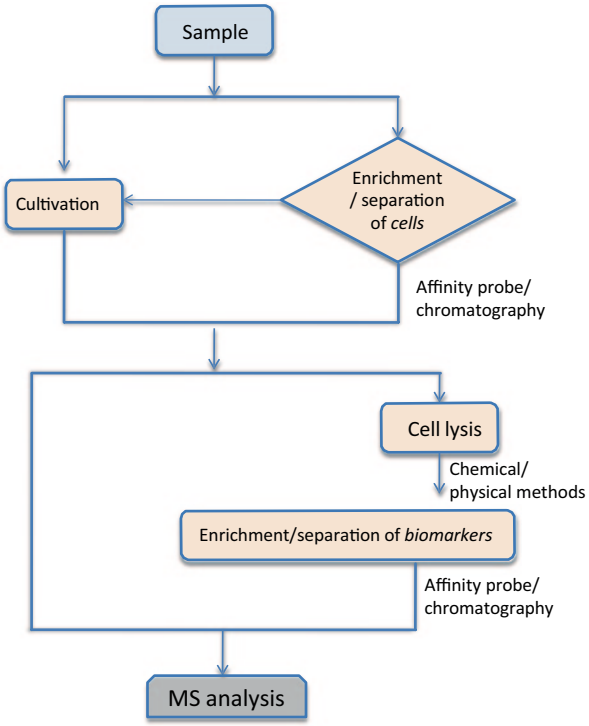


Fig. 3.1 Workflow of sample preparation approaches for microbial analysis. *MS* mass spectrometry

Enrichment Methods

Various strategies have been developed for the enrichment of microbial samples. The enrichment methods involving non-covalent, covalent interactions, and immunoassays are discussed in this section. In general, affinity enrichment steps include binding of targets, washing, and elution. A general scheme is provided in Fig. 3.2. Microbial cells/markers are isolated and concentrated after the incubation of the sample solution with affinity probes. The enriched cells are lysed or directly mixed with a matrix-assisted laser desorption/ionization (MALDI) matrix solution and subjected to mass spectrometry (MS) analysis. Moreover, biomarkers obtained from the enriched cells may be concentrated and separated prior to MS analysis

Enrichment of Microbial Cells/Biomarkers Involving Non-covalent Interactions

A simple concentration step prior to MS analysis may improve the detection and identification of microorganisms from complex biological mixtures. Affinity meth-

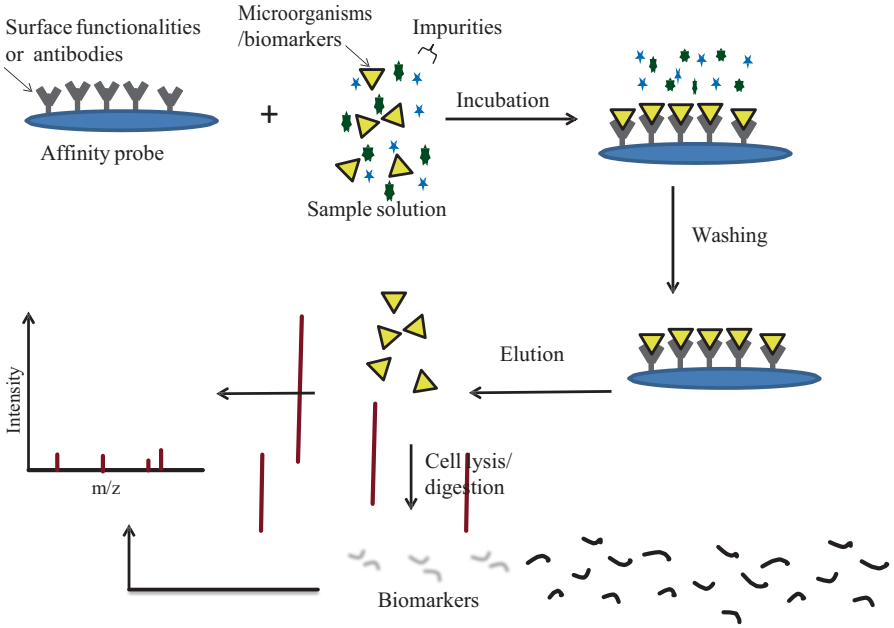


Fig. 3.2 General scheme for the enrichment of microbial cells or biomarkers with affinity probes.

ods have been developed for effective cleaning and enriching microorganisms from complex matrices.

Non-covalent binding between the cells/biomarkers and the functionalities of the affinity surface may arise from columbic (electrostatic) interactions, hydrogen bonding, or hydrophobic interactions.

Cells

Various affinity probes have been used to concentrate and purify the bacteria of interest. One such probe for the selective enrichment of bacteria using lectins has been previously reported (Bundy and Fenselau 1999). Lectins are glycoproteins that have selective affinity for carbohydrates. These molecules have the capacity to interact with bacteria through non-covalent interactions involving hydrogen bonding, hydrophobic, and van der Waals' interactions with various carbohydrates (lipopolysaccharides and peptidoglycans) located on the cell surface. The selectivity of lectins for a particular carbohydrate can be used as a probe for selective isolation of bacterial species. For example, the lectin, concanavaline A (Con A), is reactive with various Gram-negative bacteria due to its binding to lipopolysaccharides on the cell surface. Further improvement of this technique was achieved using wheat germ agglutinin (WGA)-lectin bound to an affinity membrane. WGA-lectin probes have been employed for the enrichment of *Escherichia coli* and *Salmonella typhimurium* from milk, urine, and processed chicken samples (Bundy and Fenselau 2001).

Surface-activated glass slides with immobilized lectins have been reported to selectively capture bacteria *(E. coli)* and bacterial spores *(Bacillus cereus* and *Bacillus subtilis)* (Afonso and Fenselau 2003).

Numerous magnetic nanoparticles (MNPs) of various sizes, shapes, and compositions have been used as affinity probes to selectively concentrate trace amounts of bacteria from complex biological and food samples. Nanoparticle (NP)-based magnetic separation has been reported to perform exceptionally well for pre-concentration, isolation, and enrichment of microorganisms in comparison to other common separation techniques. Microbial cells often bind with NPs through electrostatic or hydrophobic interactions. Generally, the procedure involves addition of functionalized NPs into cell suspensions and incubation of the suspensions for the efficient attachment of NPs to the microorganisms. Then the NP-microbial cell conjugates can be isolated by magnetic separation and subjected to MS analysis.

Various carbohydrates have been recognized as receptors for the attachment of pathogens (Sharon 2006). Silica-coated MNPs (Fe_3O_4@SiO_2) modified with D-mannose have been employed to concentrate *E. coli* strain ORN178, which possesses mannose-specific receptor sites (El-Boubbou et al. 2007). Pigeon ovalbumin (POA), a phosphoprotein containing high level of galactose units, also serves as an affinity probe to enrich bacterial cells through disaccharide–protein interactions. POA-immobilized Fe_3O_4@Al_2O_3 MNPs have been used for the selective enrichment of *E. coli* and *Pseudomonas aeruginosa* from urine samples at a concentration as low as 4×10^4 cells/mL, corresponding to 10^2 cells deposited on the matrix-assisted laser desorption/ionization (MALDI) plate (Liu et al. 2008, 2009).

Vancomycin-modified Fe-Pt MNPs have been reported for selective enrichment of vancomycin-resistant enterococci and other Gram-positive bacteria *(Staphylococcus aureus* and *Staphylococcus epidermidis)*. The selective interaction between MNPs and bacterial cell walls (expressing D-Ala-D-Ala as the terminal peptides) is through multiple hydrogen bonding (Gu et al. 2003).

Vancomycin-immobilized MNPs were employed for the selective enrichment of Gram-positive bacteria from urine samples. The optimal detectable concentration was 7.4×10^4 and 7.8×10^4 CFU/mL for *S. aureus* and *Staphylococcus saprophyticus*, respectively (Lin et al. 2005). Vancomycin-modified NPs have also been utilized to concentrate Gram-positive bacteria from tap water prior to MALDI-MS. The optimal detection limit was found to be 5×10^2 CFU/mL (Li et al. 2010). The architecture and orientation of vancomycin on the surface of silica-encapsulated Fe_3O_4-NPs and the overall surface coverage have been found critical in mediating fast and effective interaction between the NPs and the pathogen cell wall. Only one orientation/architecture in a series of modified NPs led to the efficient and reproducible capture of several important pathogenic bacteria (Kell et al. 2008).

Anion-exchange MNPs also served as affinity probes to separate/concentrate bacterial cells. The positively charged NPs may interact with bacteria (generally

carrying negative charges). Commercially available anion-exchange MNPs or microparticles have been used to enrich various bacterial species from tap water and reservoir water. The detection limit was 1×10^3 CFU/mL and the analytical time was around 2 h (Guo et al. 2009; Li et al. 2009). Cationic ionic liquid-modified magnetic nanoparticles (CILMS) have been used for the enrichment of pathogenic bacteria from blood samples (Bhaisare et al. 2014).

Cultivation of captured bacteria may further improve the detection limit. Pathogens in milk/pudding/coffee have been captured with magnetized zirconium hydroxide, directly cultured without colony isolation, and then analyzed using MALDI and LC-MS/MS. The limit of detection for *Enterococcus faecalis* spiked into milk was down to a level as low as 32 CFU/mL (Chen et al. 2012). The synthesis of both positively and negatively charged NPs with various functionalities including Fe_3O_4, SiO_2, TiO_2, ZrO_2, polyethyleneimine (PEI), and polyacrylic acid (PAA) to concentrate target Gram-negative and Gram-positive bacteria has been reported. The capture efficiency of all MNPs was examined. The affinity of Fe_3O_4@PEI for each of the bacterial species was attributed to its polymer structure having more positive charges on the surface. The capture mechanism of selective binding of negatively charged TiO_2 NPs with bacteria (also negatively charged) may include not only electrostatic attraction but also covalent and hydrogen bonding (Reddy et al. 2014).

The size effect of magnetic particles (nano- vs. micro-sized particles) on the capture efficiency of microorganisms may require further studies. Nano-sized particles have greater surface area to volume ratio and better dispersion properties. Micro-sized particles have higher surface coverage of functionalities. Both types of particles have been successfully applied to cell enrichment.

Biomarkers

Affinity chromatography is a powerful tool for the concentration of microbial proteins and has potential applications for diagnosis and even for therapy. Lectin affinity chromatography was used for capturing glycosylated proteins, which are potential virulence factors, from many parasites including apicomplexan (Fauquenoy et al. 2008). Surface proteins related to pathogenesis of *Mycoplasma hyopneumoniae* have been labeled with biotin and affinity captured for MS characterization (Reolon et al. 2014). Isotope-coded glycosylation-site-specific tagging has been employed to enrich low abundant *N*-glycosylated proteins from a complex bacterial extract (Kaji et al. 2003).

Immobilized MNPs can also be used for the selective enrichment of biomarkers. 12-Hydroxy octadecanoic acid-modified $BaTiO_3$ NPs have been used to extract hydrophobic compounds, including phospholipids and membrane proteins from *E. coli* (Kailasa and Wu 2013). Oleic acid capped $Mg(OH)_2$ NPs have been used as extracting and concentrating probes for hydrophobic proteins from *E. coli* and *B. subtilis* prior to MALDI-MS analysis (Kailasa and Wu 2012).

Surface-enhanced laser desorption/ionization (SELDI-MS) technology plays an important role in rapid identification of biomarkers from various microorganisms. SELDI is an affinity-based method in which the protein mixture is adsorbed onto various chemically or biochemically modified surfaces such as anion exchange (AX), cation exchange (CX), and hydrophobic or immobilized metal ion affinity chromatography (IMAC) surfaces. Some proteins in the sample bind to the surface while others are removed by washing. After washing, the adsorbed proteins are mixed with matrix and allowed to crystallize. Binding of the proteins to the surface itself acts as a separation step and the proteins can be easily analyzed. Pathogenic Gram-negative bacteria *Francisella tularensis* has been analyzed employing AX, CX, and IMAC (loaded with copper ions) chip surfaces. These approaches enable the discrimination among the different species and the subspecies of Francisella (Lundquist et al. 2005; Seibold et al. 2007). The *Campylobacter* species have been analyzed using SELDI-MS with hydrophobic and IMAC protein-chip array surfaces (Kiehntopf et al. 2011).

SELDI-MS based on protein-chip arrays have been utilized for the selective and rapid identification of *S. aureus* from other non *S. aureus* species (Yang et al. 2009). Furthermore, the protein chips have also been employed for the rapid identification of antibiotic resistance of *E. coli* (Dubska et al. 2011).

Enrichment of Cells/Biomarkers Involving Covalent Interactions

Cells

Titania-coated magnetic iron oxide (Fe_3O_4@TiO_2) NPs have been employed to enrich five Gram-negative bacteria including *E. coli* O157:H7, *uropathogenic E. coli, Shigella sonnei, P. aeruginosa*, and *Klebsiella pneumonia* and three Gram-positive bacteria *(Listeria monocytogenes, S. saprophyticus*, and *S. aureus)* (Chen et al. 2008). The Fe_3O_4@TiO_2 NPs exhibited much higher capture capacities towards Gram-negative bacteria. The outer membrane of Gram-negative bacteria contains lipopolysaccharide, as a major component, whereas most Gram-positive bacteria lack lipopolysaccharides. Therefore, the metal oxide-coated MNPs interacted weakly with the Gram-positive bacteria but strongly with the Gram-negative bacteria through metal coordination bonding with the phosphate groups on the lipopolysaccharides. A titanium-based bacterial chip has been utilized to capture pathogenic bacteria. This reusable bacterial chip can be directly used as a MALDI target plate for the rapid and sensitive bacterial analysis in MALDI-MS (Gopal et al. 2013).

Biotinylation of cell surface proteins can be used for cell enrichment. This approach has been employed by covalent modification of the membrane proteins with a cleavable reactive ester derivative of biotin (sulfo-*N*-hydroxysuccinimide-SS-biotin). Then, the biotinylated cells were enriched with streptavidin-coated resin (Scheurer et al. 2005).

Biomarkers

Enrichment of biomarkers through covalent interactions may benefit from the stability of the bonding at different salt concentrations, pH values, and temperatures. IMAC can be used to concentrate biomolecules based on the metal ion chelation with the biomolecules. IMAC is highly useful for the enrichment of phosphopeptides from complex mixtures using various metal ions such as Ga(III) and Fe(III). For instance, this method has been employed to enrich phosphopeptides from *Saccharomyces cerevisiae* (Gruhler et al. 2005).

Covalent chromatography has been applied to the enrichment of thiol-containing proteins from *E. coli* lysate. Initially, the free thiol groups were derivatized with the 2,2-dipyridyl disulfide reagent. Then, the peptides were passed through a thiopropyl sepharose resin column, which facilitated the exchange of the disulfide-modified peptides with the thiol groups of the column (Wang and Regnier 2001). Covalent chromatography and IMAC (loaded with copper) have been used to selectively capture cysteine and histidine containing peptides respectively from *E. coli* cell lysate (Wang et al. 2002). Disulfide containing membrane proteins have been enriched from *Pseudomonas putida* using a similar approach (An et al. 2011).

The magnetic microspheres have been employed for specific enrichment of *F. tularensis* glycoproteins. The capture was based on the irreversible reaction of the hydrazide groups of microspheres with the aldehydes generated by the oxidation of the cis-diol groups of glycoproteins. Biomarkers containing various functional groups (carboxyl, amino, sulfhydryl, etc.) can also be conjugated with these microspheres through covalent bonding (Horak et al. 2012).

Organomercurial agarose beads have been developed to enrich cysteine containing peptides from yeast cell lysates. A significant increase of proteins that were suitable for identifying yeast proteins has been observed in MS analysis (Raftery 2008).

Biomarkers may be easily tagged through the modification of the reactive functional groups such as amino, carboxylic, hydroxyl and thiol groups. The most popularly used chemical affinity tag for the enrichment of biomarkers is the biotin tag, which can be attached to various biomarkers through covalent bonding. Enrichment of biotinylated proteins can be done by both avidin and streptavidin immobilized affinity chromatography. Various research groups have employed biotinylation to enrich different target proteins. Biotinylation has been employed to enrich and detect membrane proteins from *Deinococcus radiodurans* (Goshe et al. 2003) and *P. aeruginosa* (Blonder et al. 2004).

Immuno-Affinity Enrichment of Microbial Cells/Biomarkers

Although the interaction involved in immuno-affinity methods is basically non-covalent in nature, it appears to be unique enough to warrant separate discussion. Immunological methods have been extensively used for the studies of pathogenic microorganisms since the late 1950s. Traditional immunoassays are time-consuming

and in some cases the antibody cross-reactivity with food and biological matrices may lead to false-positive results and/or a worse detection limit. Immuno-affinity MS is a more specific affinity approach capable of selectively targeting and characterizing protein biomarkers. In MS-based immunoassays, proteins are affinity retrieved from biological samples via surface-immobilized antibodies, and are then detected via MS analysis.

Cells

Micro-sized magnetic beads coated with antibodies (against a chosen microbial species) can bind to target microbial cells in complex biological samples. The magnetized beads are easily dispersed in solution because of their small sizes (2–5 μm in diameter) and can be retrieved using a magnetic field. Sample debris and nontarget organisms and molecules are removed by washing.

This method enriches specific microbial species, while the magnetic separation involving non-covalent interaction described in Section 1.1 is mostly nonspecific. Magnetic beads coated with antibodies that were specific to the antigen of *Salmonella choleraesuis* have been applied for the isolation/concentration of *S. choleraesuis* from river water, human urine, and chicken blood (Madonna et al. 2001).

Immunoglobulin (IgG) functionalized magnetic Au-NPs and Pt-NPs have been utilized to selectively enrich targeted bacteria from sample solutions (Ho et al. 2004; Ahmad and Wu 2013).The captured bacteria with the beads or particles were directly applied onto the MALDI target plates prior to MS analysis.

Captured bacteria can be lysed after the magnetic separation. The released proteins are digested and the peptides are analyzed using LC-MS. An antibody against *Bacillus anthracis* spores were immobilized on IgG magnetic beads for the immunocapture of intact spores prior to their detection using multiple reaction monitoring (MRM)-MS (Chenau et al. 2011). Immuno-LC-MS/MS has also been used for the selective isolation and detection of *Yersinia pestis*. Optimizing the immunocapture conditions for the enrichment of intact bacterial cells from complex matrices improved the detection limit to 2×10^4 CFU/mL in milk or tap water and in soil (Chenau et al. 2014).

Biomarkers

Staphylococcal enterotoxin B (SEB) is a potent bacterial protein toxin responsible for food poisoning, as well as a potential biological warfare agent. An affinity probe has been prepared by immobilizing anti-SEB antibody on the surface of paratoluene-sulfonyl-functionalized magnetic beads. Immobilization and affinity capture procedures were optimized to maximize the density of anti-SEB IgG on the surface of magnetic beads and the amount of captured SEB. MALDI-MS detection of the enriched SEB from different matrices, such as cultivation media of *S. aureus* strains and raw milk samples (Schlosser et al. 2007).

Enrichment of microbial toxins including ricin, SEB, and botulinum neurotoxins (BoNT) has been performed using multiplex-immuno-affinity purification. Specific monoclonal antibodies for each of the four toxins were selected from a pool of antibodies, the selected antibodies allowed for the specific and simultaneous capture of toxins. This assay enabled unambiguous identification of toxins in complex food matrices with a detection limit of 500 fmol. Additionally, it allowed for the rapid differentiation of closely related BoNT sero- and subtypes ((Kull et al. 2010).

Cell Lysis

Microbial cells need to be lysed to release intracellular components prior to MS analysis. Various cell lysis strategies are available depending on the sample types and target biomarkers.

Gram-negative and Gram-positive bacteria differ from each other due to the difference in composition of the cell walls. The complexity of the microbial sample, limited availability of the sample, and need for rapid identification have prompted the development of cell lysis methodologies which can be coupled to MS techniques. Various cell lysis techniques are available for the release of intracellular constituents, including chemical (acids, detergents), enzymatic (lysozyme), physical (bead milling, ultra-sonication, French press, freeze-thawing, osmotic shock, corona plasma discharge), and the combined methods.

Chemical Methods

Strong organic acids, bases, alcohols, and detergents/surfactants have been employed in microbial sample preparations, due to their high efficiency in disrupting membranes and solubilizing proteins. Different cell types require different buffer formulations. Conditions such as pH, salt concentration, and temperature are considered to be important parameters in the sample preparation protocols.

Trifluoroacetic acid (0.1% TFA) can effectively extract proteins from both Gram-positive/negative bacteria (Nilsson 1999). Comparison of various solvents involving water, 40% ethanol, 0.1% TFA, and various solvent mixtures indicated that 0.1% TFA could yield the most MALDI-MS peaks (Ruelle et al. 2004a; Liu et al. 2007). Extraction of proteins from *E. coli* has been achieved using a range of solvents with varying polarity. The different combinations of extraction solvents, involving 10:45:45 formic acid (FA)/methanol/water, 17:33:50 FA/isopropylalcohol/water, 17:33:50 FA/methanol/water, 33:67 acetonitrile (ACN)/water, and 0.1:99.9 TFA/water, have also been examined. Among these, the second and the third solvent combinations provided the highest number of signals from *E. coli* extract (Domin et al. 1999).

Bacillus anthracis is the etiological agent of anthrax in humans/animals (Demirev and Fenselau 2008; Lasch et al. 2009). Numerous low molecular-weight proteins can be readily extracted from the spores of *B. anthracis* and related species. Many of these proteins have been identified as small acid-soluble spore proteins (SASPs), due to their basic nature. They can be selectively solubilized in acids and easily protonated to provide strong signals when ionized by MALDI or electrospray ionization (ESI). SASPs have been proposed as candidate biomarkers capable of discriminating between various Bacillus spores. Solvents that have been reported to extract SASPs include 10 % TFA (Castanha et al. 2007), 50 % acetic acid (Fenselau et al. 2007), 30 % ACN/40 % FA (Dickinson et al. 2004), 5 %ACN/TFA (70:30, v/v) (Dybwad et al. 2013), and 1N HCl (Hathout et al. 2003).

Various types of detergents/surfactants have been reported for microbial cell lysis to extract proteins. Detergents break the structure of cell membranes by disrupting the lipid–lipid, protein–lipid, and protein–protein interactions. Detection of proteins with high molecular weight up to 140 kDa has been achieved by mixing nonionic surfactant (1.0 mM *N*-octyl-B-D-galactopyranoside) with bacterial cells on the MALDI target (Meetani and Voorhees 2005). The release of proteins from *E. coli* has been achieved by a lysis buffer containing 2 % Triton, which facilitated the identification of many proteins (Chong et al. 1997). The most commonly used detergent is sodium dodecyl sulfate (SDS), which is very efficient in cellular lysis and protein solubilization. However, it poses a major impediment for LC-MS/MS experiments. In particular, the presence of SDS hinders the resolving power of reverse phase LC. The introduction of SDS into a mass spectrometer during electrospray may lead to ion suppression or contamination inside the ion source. Therefore, the detergents used for lysis must be removed using suitable cleanup methods. Detergent cleanup methods including trichloroacetic acid precipitation, chloroform/methanol/water extraction, a commercial detergent removal spin column method, and filter-aided sample preparation (FASP) have been investigated. These methods have been compared by lysing the bacterial samples in a sample buffer containing 2 % SDS. The FASP method outperformed the other three SDS cleanup methods. An additional fractionation step enhanced the number of proteins identified from bacterial cell lysates by 8–25 % (Sharma et al. 2012). Melittin, a lytic peptide, shows lytic activities against bacteria and mammalian cells. The leucine-substituted analogs of melittin exhibited selective lysis of *E. coli* and *Bacillus megaterium* (Pandey et al. 2010).

Guanidine HCl has been reported to be very efficient in cell lysis with minimal amount of bacterial sample. Comparative study on effective small-scale cell lysis using guanidine HCl and trifluoroethanol (TFE) indicated that guanidine HCl gave better cell lysis results for complex microbial samples. Furthermore, comparison of lysis techniques including sonication and modified guanidine lysis showed that a higher number of proteins were identified using guanidine HCl (Thompson et al. 2008).

Lysozyme is commonly used for cell lysis to enhance the release of proteins from microbial cells. The results of the lysozyme treatment on the *Enterococcus* species indicated that the most reproducible MALDI-MS profile was achieved by treating

the bacteria with lysozyme for 30 min. It was noted that the isolates exposed to lysozyme have more intense peaks when compared to the untreated isolates. The lysozyme-treated cells yielded fewer peaks in the low-mass range (900–1400 Da), but the whole-mass range was increased up to 10 kDa (Giebel et al. 2008). An increase in the number of peaks for *S. aureus, Streptococcus haemolyticus*, and *Streptococcus pyogenes* (Gram-positive bacteria) has been achieved by the treatment of lysozyme combined with sonication.

Gram-positive bacteria contain a thick peptidoglycan layer in the cell wall which may impede the release of analyte molecules from the intact cells during MALDI analysis. When these Gram-positive bacteria were exposed to lysozyme, digestion of peptidoglycan might occur and yield the additional higher-mass signals in the spectra (Smole et al. 2002). In contrast, a decrease in the MALDI-MS quality after the lysozyme treatment has also been observed for both Gram-positive (Vargha et al. 2006) and Gram-negative bacteria (Williams et al. 2003). Decrease in signal quality might be attributed to the introduction of additional components to the analyzed mixture.

Physical Methods

Physical methods are also shown to be efficient in disrupting microbial cells. Various physical methods such as micro-beads disruption, ultra-sonication, freeze/thaw cycle, thermolysis, and corona plasma discharge have been employed for efficient cell lysis. A simple and inexpensive method for the disruption of small volumes of bacteria and yeast is by suspending glass beads within the sample and vortexing the sample repeatedly. The beads disrupt cell membranes through shear forces, grinding between the beads, and collision with the beads. *Campylobacter* or *Streptococcus* cells have been successfully disrupted with zirconia/silica beads (Fagerquist et al. 2005; Teramoto et al. 2007).

Thermolysis also induces cell lysis. For example, *F. tularensis* cell lysis has been achieved by heating the sample at 65 °C for 2 h (Lundquist et al. 2005) prior to SELDI-MS analysis. Effective release of high-mass biomarkers from *Bacillus* spores has been achieved using wet heat treatment.

Wet heat treatment has been performed for 3–30 min by two techniques using either a screw-cap tube submerged in a glycerol bath at 120 °C or an eppendorf tube submerged in a water bath at 100 °C. Both techniques were successful in releasing high-mass biomarkers (Horneffer et al. 2004). Osmotic lysis and French press have been utilized for *Shewanella oneidensis* cell lysis (Brown et al. 2010). Reproducible patterns of spectral markers and increased sensitivity have been achieved for *Cryptosporidium parvum* employing freeze-thaw cell lysis procedure (Magnuson et al. 2000).

Ultrasonic vibrations at a frequency of 25 kHz is commonly used for cell disruption. The duration of ultrasound needed depends on the cell type, sample size, and cell concentration.

Ultra-sonication has been successfully applied for the lysis of *E. coli* to detect high-mass proteins with higher signal-to-noise ratio (Easterling et al. 1998).

Nonthermal plasma or corona treatment of microorganisms at room temperature and ambient pressure produced unique biomarkers when analyzed by mass spectrometers (Birmingham et al. 1999). Atmospheric pressure nonthermal plasmas rapidly lysed bacterial spores *(B. subtilis)* after exposing the spores to ionized gas for a few minutes (Birmingham 2006). Barrier discharges have also been utilized as cell lysis devices (Pineda et al. 2000). Corona plasma discharge showed potential to benefit the MALDI-MS profiling performance with rapid sample treatment. For example, detection of signals of *B. cereus* spores required only 3 s of treatment with corona plasma discharge (Victor Ryzhov and Catherine 2000). The treatment of intact cells deposited on the MALDI target with corona plasma discharge of 15 s was found to be useful for the signal enhancement. This effect was more notable in the analysis of Gram-positive bacteria because of their higher cell wall stability. MS analysis of plasma-treated bacterial spores revealed new biomarkers in the mass spectra, which were undetectable in the spectra of unprocessed samples (Birmingham et al. 1999).

Separation of Cells and Biomarkers

Intact Cells

Capillary electrophoresis (CE) is a separation technique based on the differential migration rates of charged species in an applied electric field. Microorganisms carry charged or chargeable groups (amino acids and carbohydrates) on their outer surface; therefore, CE can be used to separate intact bacterial, viral, and fungal cells. CE allows for the rapid and efficient separation of microorganisms with least consumption of sample and reagents. CE analysis of *Tobacco mosaic* virus and *Lactobacillus casei* using a fused-silica capillary coated with methyl cellulose has been carried out (Hjerten et al. 1987). Off-line coupling of CE to MALDI-MS may be used for microbial identification. Separation and designation of two distinct subcomponents of cow pea mosaic virus were done by capillary zone electrophoresis and MALDI-MS analysis of the capsid proteins (Liang and Schneider 2009). CE coupled off-line with desorption electrospray ionization (DESI) MS has been employed for the separation/identification of *E. coli* and *S. cerevisiae* (Petr et al. 2009).

Field flow fractionation (FFF) can also be used for microbial cell separation. In the FFF technique, a field (may be gravitational, centrifugal, thermal-gradient, electrical, magnetic, etc.) is applied perpendicular to the fluid flow, causing particles to migrate with different velocities. Fields of sedimentation, diffusion, and electrical diffusion are manipulated to optimize the separations of microbes. Separation of *Pseudomonas putida* and *E. coli* has been achieved by hyperlayer FFF. Fractions of the whole cells were collected after the separation at different time intervals, dif-

ferent sizes, and possibly different growth stages of bacteria. The bacterial analysis by FFF/MALDI-MS has been completed in 1 h (Lee et al. 2003). Hyperlayer flow FFF coupled online with inductively coupled plasma MS has been employed for the separation/analysis of *S. oneidensis* (Jackson et al. 2005).

Biomarkers

CE is also useful for the separation of biomolecules. The selective proteotypic-peptide analyses of protein digests obtained from various bacterial cell (Gram positive/Gram negative) extracts were successfully performed using CE-MS. Minor bacterial species present in the complex mixture at even 1% relative abundance could be identified with high confidence (Hu et al. 2007). CE time of flight (TOF)-MS has also been found to be applicable in separation and identification of metabolites including amino acids, amine, nucleotides, sugars, lipids, and other substances from various microbial cells (Garcia et al. 2008). Tryptic peptides from whole cell lysates have been separated and analyzed by CE-ESI-MS/MS for the sensitive and specific identification of β-lactamases in multidrug-resistant Gram-negative bacterial species *(K. pneumoniae, E. coli*, and *Enterococcus cloacae)*. Analysis of clinical isolates identified the presence of β-lactamase peptides (OXA-48 and KPC) in all of the carbapenemase positive samples (Fleurbaaij et al. 2014).

LC is the most widely used method for the separation of large variety of microbial biomarkers. LC-MS has been employed for the separation and detection of protein biomarkers from *E. coli* (Ho and Hsu 2002). A multidimensional LC-MS method has been used for the rapid determination of bacterial proteins to identify *B. anthracis* strains (Krishnamurthy et al. 2007). Reproducible intact protein markers identified using the LC-MS approach were used to correctly identify unknown pathogens at the species (Everley et al. 2008) and strain level (Everley et al. 2009). Multidimensional LC separation coupled with MS/MS has been employed to obtain the whole-cell proteome, providing insight into the pathogenesis of *Streptococcus pneumonia* (Sun et al. 2011). In order to identify biomarkers for *Ruegeria lacuscaerulensis*, tryptic digests of low molecular weight proteins have been separated and identified by the shotgun nano-LC-MS/MS approach (Christie-Oleza et al. 2013). *Acinetobacter baumannii* DU202, a clinical isolate, exhibits resistance to many antibiotics, including imipenem, tetracycline, ampicillin, and chloramphenicol. The proteomes of the cytoplasm, cell wall, and membrane of *A. baumannii* DU202 have been analyzed by LC-MS/MS. Combining the proteomic analysis with genome sequence data provided the comprehensive picture of antibiotic resistance in this strain (Lee et al. 2014). Identification of β-lactam resistance in *A. baumannii* has been achieved by shotgun proteomics and nano-LC-MS. Various antibiotic-resistant proteins, including AmpC, β-lactamase, and carO, have been successfully identified in clinical-resistant strains of *A. baumannii* (Chang et al. 2013).

Gas chromatography-based separation methods for the identification of bacterial metabolites are well established. The maximum molecular weight of compounds

that can be analyzed when coupled to MS is < 1000 Da. GC-MS and direct-infusion MS methods have been employed to produce specific and discriminant metabolite profiles from different yeast mutants (*S. cerevisiae* strains) (Mas et al. 2007). The analysis of volatile compounds generated by bacteria may be a possible alternative method for the identification of pathogenic bacteria. The rapid detection of *L. monocytogenes* from milk has been achieved by extracting, separating, and detecting the volatile organic compounds (VOCs) by headspace-solid phase micro-extraction coupled to GC-MS. The limit of detection was found to be $1–1.5 \times 10^2$ CFU/mL of cells in milk (Tait et al. 2014). High-resolution pyrolysis gas chromatography/MS selected ion monitoring technology has also been employed to detect *L. monocytogenes* from food products (beef and milk) (Li et al. 2014). The identification of VOCs produced by microorganisms also assisted in determining the bacterial infections. Differentiation between a methicillin-resistant and a methicillin-sensitive isolates of *S. aureus* was possible due to the significant differences between the produced VOCs compounds including 1,1,2,2-tetrachloroethane, 2-heptanone, and 1,4-dichlorobenzene (Boots et al. 2014). Needle trap GC-MS has been employed for the enrichment and analysis of VOCs in *E. coli* and *P. aeruginosa* (Zscheppank et al. 2014).

Sample Preparation for MALDI-MS

Currently, most of the rapid microbial analyses are based on MALDI techniques. Selecting an optimal matrix is a crucial step in developing all of the sample preparation protocols for the MALDI-MS. The MALDI matrices that are frequently used for microbial analysis are α-cyano-4-hydroxycinnamic acid (CHCA), ferulic acid (FA), and sinapinic acid (SA). It has been demonstrated that the use of a different matrix for the same kind of sample led to a significant change in the MALDI mass spectrum. For example, the MALDI mass spectra of *E. coli* obtained with CHCA and a mixture of SA/4-methoxycinnamic acid showed significantly different signals (Demirev et al. 1999). CHCA is one of the most common matrix for bacterial identification.

Analysis of bacterial cells using CHCA yielded better signal-to-noise ratio and sensitivity and more number of intense signals in the lower mass range (Ryzhov and Fenselau 2001; Shaw et al. 2004; Ruelle et al. 2004b). Presence of more signals in the lower mass range may be due to the formation of doubly protonated ions (Pineda et al. 2003). SA is a matrix of choice for the analysis of high-mass proteins (Ryzhov and Fenselau 2001; Moura et al. 2008). Increased signal-to-noise ratio and intensity of signals and improved spectral quality due to the decrease in peak broadening were also found with SA (Ochoa and Harrington 2005). In comparison to SA and CHCA matrices, FA has been reported to be better for the detection of high-mass biomarkers from Gram-positive/Gram-negative bacteria (Madonna et al. 2000). This phenomenon may be attributed to the higher tolerance to salts and contaminants with FA. 2,5-Dihydroxybenzoic acid (DHB) is a matrix suitable

for the analysis of carbohydrates and small peptides. MALDI profiling with this matrix produced a lower degree of peak broadening (Ryzhov and Fenselau 2001), increased reproducibility (Jones et al. 2003), and higher MALDI-MS/MS fragment ions (Demirev et al. 2005).

2-(4-Hydroxyphenylazo)-benzoic acid was employed as a matrix for MALDI-MS analysis of *Rhizobium* cells, which yielded a maximum number of signals in the 1–10 kDa range (Mandal et al. 2007). Ionic liquid matrices (ILMs) based on SA and DHB in conjugation with various bases including aniline, dimethyl aniline, diethylaniline, dicyclohexylamine, and pyridine have been reported for the analysis of intact bacteria in MALDI-MS (Abdelhamid et al. 2013).

The choice of matrix solvents in sample preparation protocols for the MALDI-MS is also an essential step. Solvent volatility was frequently modified in order to achieve fast or slow crystallization. Large biomolecules require slow crystallization to have more time to incorporate into the matrix crystal. The homogeneity of the matrix layer containing small crystals can be achieved by fast solvent evaporation. Bacterial cell lysate or extract was often mixed with the suitable composition of matrix solvent to enhance the MS mass range and signal intensities. For example, an *E. coli* suspension mixed with the CHCA matrix in 0.1% TFA/ACN (60:40) generated peaks only in the low-mass region, whereas more peaks in the high-mass range (over 10 kDa) were observed when the matrix was dissolved in ACN/isopropanol/0.1% TFA (49:49:2) (Ruelle et al. 2004b). Acetonitrile was found in almost all the matrix solvent compositions because of its excellent solubilization capability both for matrices and peptides/proteins. In addition, several co-matrices have been employed in sample preparation protocols to modulate the sample crystallization properties or to suppress the ionization of interferences. For example, addition of crown ether (at 0.01 M concentration) to the 5-chloro-2-mercaptobenzothiazole matrix has been reported to produce higher sensitivity and better spectral reproducibility in the analysis of Gram-positive bacteria (Evason et al. 2000).

References

Abdelhamid HN, Gopal J, Wu HF. Synthesis and application of ionic liquid matrices (ILMs) for effective pathogenic bacteria analysis in matrix assisted laser desorption/ionization (MALDI-MS). Analytica Chimica Acta. 2013;767:104–11. doi:10.1016/j.aca.2012.12.054.

Afonso C, Fenselau C. Use of bioactive glass slides for matrix-assisted laser desorption/ionization analysis: application to microorganisms. Anal Chem. 2003;75(3):694–7.

Ahmad F, Wu H-F. Rapid and sensitive detection of bacteria via platinum-labeled antibodies and on-particle ionization and enrichment prior to MALDI-TOF mass spectrometry. Microchim Acta. 2013;180(5–6):485–92. doi:10.1007/s00604-013-0951-5.

An B, Lee S, Lee E, Lee J, Wi S, Jung H, Park W, Lee S, Chung B. Functional switching of a novel prokaryotic 2-Cys peroxiredoxin (PpPrx) under oxidative stress. Cell Stress Chaperones. 2011;16(3):317–28. doi:10.1007/s12192-010-0243-5.

Bhaisare ML, Abdelhamid HN, Wu B-S, Wu H-F. Rapid and direct MALDI-MS identification of pathogenic bacteria from blood using ionic liquid-modified magnetic nanoparticles (Fe3O4@SiO2). J Mater Chem B. 2014;2(29):4671–83. doi:10.1039/C4TB00528G.

Birmingham JG. Plasma lysis for identification of bacterial spores using ambient-pressure nonthermal discharges. Plasma Sci IEEE Trans. 2006;34(4):1270–4. doi:10.1109/TPS.2006.877740.

Birmingham J, Demirev P, Ho Y-P, Thomas J, Bryden W, Fenselau C. Corona plasma discharge for rapid analysis of microorganisms by mass spectrometry. Rapid Commun Mass Spectrom. 1999;13(7):604–6. doi:10.1002/(SICI)1097-0231(19990415)13:7<604::AID-RCM529>3.0.CO;2-M.

Blonder J, Goshe MB, Xiao W, Camp DG, Wingerd M, Davis RW, Smith RD. Global analysis of the membrane subproteome of *Pseudomonas aeruginosa* using liquid chromatography-tandem mass spectrometry. J Proteome Res. 2004;3(3):434–44. doi:10.1021/pr034074w.

Boots AW, Smolinska A, van Berkel JJ, Fijten RR, Stobberingh EE, Boumans ML, Moonen EJ, Wouters EF, Dallinga JW, Van Schooten FJ. Identification of microorganisms based on headspace analysis of volatile organic compounds by gas chromatography-mass spectrometry. J Breath Res. 2014;8(2):027106. doi:10.1088/1752-7155/8/2/027106.

Brown RN, Romine MF, Schepmoes AA, Smith RD, Lipton MS. Mapping the subcellular proteome of Shewanella oneidensis MR-1 using sarkosyl-based fractionation and LC–MS/MS protein identification. J Proteome Res. 2010;9(9):4454–63. doi:10.1021/pr100215h.

Bundy J, Fenselau C. Lectin-based affinity capture for MALDI-MS analysis of bacteria. Anal Chem. 1999;71(7):1460–3. doi:10.1021/ac981119h.

Bundy JL, Fenselau C. Lectin and carbohydrate affinity capture surfaces for mass spectrometric analysis of microorganisms. Anal Chem. 2001;73(4):751–7. doi:10.1021/ac0011639.

Castanha ER, Vestal M, Hattan S, Fox A, Fox KF, Dickinson D. *Bacillus cereus* strains fall into two clusters (one closely and one more distantly related) to *Bacillus anthracis* according to amino acid substitutions in small acid-soluble proteins as determined by tandem mass spectrometry. Mol Cell Probes. 2007;21(3):190–201. doi:10.1016/j.mcp.2006.11.002.

Chang CJ, Lin JH, Chang KC, Lai MJ, Rohini R, Hu A. Diagnosis of beta-lactam resistance in *Acinetobacter baumannii* using shotgun proteomics and LC-nano-electrospray ionization ion trap mass spectrometry. Anal Chem. 2013;85(5):2802–8. doi:10.1021/ac303326a.

Chen W-J, Tsai P-J, Chen Y-C. Functional nanoparticle-based proteomic strategies for characterization of pathogenic bacteria. Anal Chem. 2008;80(24):9612–21. doi:10.1021/ac802042x.

Chen CT, Reddy PM, Ma YR, Ho YP. Mass spectrometric identification of pathogens in foods using a zirconium hydroxide immobilization approach. Int J Mass spectrom. 2012;312:45–52. doi:10.1016/j.ijms.2011.05.014.

Chenau J, Fenaille F, Ezan E, Morel N, Lamourette P, Goossens PL, Becher F. Sensitive detection of *Bacillus anthracis* spores by immunocapture and liquid chromatography-tandem mass spectrometry. Anal Chem. 2011;83(22):8675–82. doi:10.1021/ac2020992.

Chenau J, Fenaille F, Simon S, Filali S, Volland H, Junot C, Carniel E, Becher F. Detection of *Yersinia pestis* in environmental and food samples by intact cell immunocapture and liquid chromatography-tandem mass spectrometry. Anal Chem. 2014;86(12):6144–52. doi:10.1021/ac501371r.

Chong BE, Wall DB, Lubman DM, Flynn SJ. Rapid profiling of *E. coli* proteins up to 500 kDa from whole cell lysates using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom RCM. 1997;11(17):1900–8. doi:10.1002/(sici)1097-0231(199711)11:17<1900::aid-rcm95>3.0.co;2-k.

Christie-Oleza JA, Piña-Villalonga JM, Guerin P, Miotello G, Bosch R, Nogales B, Armengaud J. Shotgun nanoLC-MS/MS proteogenomics to document MALDI-TOF biomarkers for screening new members of the *Ruegeria* genus. Environ Microbiol. 2013;15(1):133–47. doi:10.1111/j.1462-2920.2012.02812.x.

Demirev PA, Fenselau C. Mass spectrometry in biodefense. J Mass Spectrom. 2008;43(11):1441–57. doi:10.1002/jms.1474.

Demirev PA, Ho Y-P, Ryzhov V, Fenselau C. Microorganism identification by mass spectrometry and protein database searches. Anal Chem. 1999;71(14):2732–8. doi:10.1021/ac990165u.

Demirev PA, Feldman AB, Kowalski P, Lin JS. Top-down proteomics for rapid identification of intact microorganisms. Anal Chem. 2005;77(22):7455–61. doi:10.1021/ac051419g.

Dickinson DN, Duc MT L, Haskins WE, Gornushkin I, Winefordner JD, Powell DH, Venkateswaran K. Species differentiation of a diverse suite of *Bacillus* spores by mass spectrometry-based protein profiling. Appl Environ Microbiol. 2004;70(1):475–82.

Domin MA, Welham KJ, Ashton DS. The effect of solvent and matrix combinations on the analysis of bacteria by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun Mass Spectrom RCM. 1999;13(4):222–6. doi:10.1002/(sici)1097-0231(19990228)13:4<222::aid-rcm440>3.0.co;2-y.

Dubska L, Pilatova K, Dolejska M, Bortlicek Z, Frostova T, Literak I, Valik D. Surface-enhanced laser desorption ionization/time-of-flight (SELDI-TOF) mass spectrometry (MS) as a phenotypic method for rapid identification of antibiotic resistance. Anaerobe. 2011;17(6):444–7. doi:10.1016/j.anaerobe.2011.05.008.

Dybwad M, van der Laaken AL, Blatny JM, 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. 2013;79(17):5372–83. doi:10.1128/aem.01724-13.

Easterling ML, Colangelo CM, Scott RA, Amster IJ. Monitoring protein expression in whole bacterial cells with MALDI time-of-flight mass spectrometry. Anal Chem. 1998;70(13):2704–9. doi:10.1021/ac971344j.

El-Boubbou K, Gruden C, Huang X. Magnetic glyco-nanoparticles: a unique tool for rapid pathogen detection, decontamination, and strain differentiation. J Am Chem Soc. 2007;129(44):13392–3. doi:10.1021/ja076086e.

Evason DJ, Claydon MA, Gordon DB. Effects of ion mode and matrix additives in the identification of bacteria by intact cell mass spectrometry. Rapid Commun Mass Spectrom RCM. 2000;14(8):669–72. doi:10.1002/(sici)1097-0231(20000430)14:8<669::aid-rcm932>3.0.co;2-7.

Everley RA, Mott TM, Wyatt SA, Toney DM, Croley TR. Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species. J Am Soc Mass Spectrom. 2008;19(11):1621–8. doi:10.1016/j.jasms.2008.07.003.

Everley RA, Mott TM, Toney DM, Croley TR. Characterization of *Clostridium* species utilizing liquid chromatography/mass spectrometry of intact proteins. J Microbiol Methods. 2009;77(2):152–8. doi:10.1016/j.mimet.2009.01.013.

Fagerquist CK, Miller WG, Harden LA, Bates AH, Vensel WH, Wang G, Mandrell RE. Genomic and proteomic identification of a DNA-binding protein used in the "fingerprinting" of *Campylobacter* species and strains by MALDI-TOF-MS protein biomarker analysis. Anal Chem. 2005;77(15):4897–907. doi:10.1021/ac040193z.

Fauquenoy S, Morelle W, Hovasse A, Bednarczyk A, Slomianny C, Schaeffer C, Van Dorsselaer A, Tomavo S. Proteomics and glycomics analyses of N-glycosylated structures involved in *Toxoplasma gondii*-host cell interactions. Mol Cell Proteomics. 2008;7(5):891–910. doi:10.1074/mcp.M700391-MCP200.

Fenselau C, Russell S, Swatkoski S, Edwards N. Proteomic strategies for rapid characterization of micro-organisms. Eur J Mass Spectrom. 2007;13(1):35–9.

Fleurbaaij F, Heemskerk AA, Russcher A, Klychnikov OI, Deelder AM, Mayboroda OA, Kuijper EJ, van Leeuwen HC, Hensbergen PJ. Capillary-electrophoresis mass spectrometry for the detection of carbapenemases in (multi-)drug-resistant Gram-negative bacteria. Anal Chem. 2014;86(18):9154–61. doi:10.1021/ac502049p.

Garcia DE, Baidoo EE, Benke PI, Pingitore F, Tang YJ, Villa S, Keasling JD. Separation and mass spectrometry in microbial metabolomics. Curr Opin Microbiol. 2008;11(3):233–9. doi:10.1016/j.mib.2008.04.002.

Giebel RA, Fredenberg W, Sandrin TR. Characterization of environmental isolates of *Enterococcus* spp. by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Water Res. 2008;42(4–5):931–40. doi:10.1016/j.watres.2007.09.005.

Gopal J, Hasan N, Wu HF. Fabrication of titanium based MALDI bacterial chips for rapid, sensitive and direct analysis of pathogenic bacteria. Biosens Bioelectron. 2013;39(1):57–63. doi:10.1016/j.bios.2012.06.036.

Goshe MB, Blonder J, Smith RD. Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. J Proteome Res. 2003;2(2):153–61. doi:10.1021/pr0255607.

Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics. 2005;4(3):310–27. doi:10.1074/mcp.M400219-MCP200.

Gu H, Ho P-L, Tsang KWT, Wang L, Xu B. Using biofunctional magnetic nanoparticles to capture vancomycin-resistant *Enterococci* and other Gram-positive bacteria at ultralow concentration. J Am Chem Soc. 2003;125(51):15702–3. doi:10.1021/ja0359310.

Guo Z, Liu Y, Li S, Yang Z. Interaction of bacteria and ion-exchange particles and its potential in separation for matrix-assisted laser desorption/ionization mass spectrometric identification of bacteria in water. Rapid Commun Mass Spectrom. 2009;23(24):3983–93. doi:10.1002/rcm.4338.

Haihong Liu ZD, Jin W, Yang R. Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2007;73 (6):1899–907.

Hathout Y, Setlow B, Cabrera-Martinez RM, Fenselau C, Setlow P. Small, acid-soluble proteins as biomarkers in mass spectrometry analysis of Bacillus spores. Appl Environ Microbiol. 2003;69(2):1100–7.

Hjerten S, Elenbring K, Kilar F, Liao JL, Chen AJ, Siebert CJ, Zhu MD. Carrier-free zone electrophoresis, displacement electrophoresis and isoelectric focusing in a high-performance electrophoresis apparatus. J Chromatogr. 1987;403:47–61.

Ho YP, Hsu PH. Investigating the effects of protein patterns on microorganism identification by high-performance liquid chromatography-mass spectrometry and protein database searches. J Chromatogr A. 2002;976(1–2):103–11.

Ho KC, Tsai PJ, Lin YS, Chen YC. Using biofunctionalized nanoparticles to probe pathogenic bacteria. Anal Chem. 2004;76(24):7162–8. doi:10.1021/ac048688b.

Horak D, Balonova L, Mann BF, Plichta Z, Hernychova L, Novotny MV, Stulik J. Use of magnetic hydrazide-modified polymer microspheres for enrichment of *Francisella tularensis* glycoproteins. Soft Matter. 2012;8(9):2775–86. doi:10.1039/C2SM07036G.

Horneffer V, Haverkamp J, Janssen HG, ter Steeg PF, Notz R. MALDI-TOF-MS analysis of bacterial spores: wet heat-treatment as a new releasing technique for biomarkers and the influence of different experimental parameters and microbiological handling. J Am Soc Mass Spectrom. 2004;15(10):1444–54. doi:10.1016/j.jasms.2004.06.017.

Hu A, Lo AA, Chen CT, Lin KC, Ho YP. Identifying bacterial species using CE-MS and SEQUEST with an empirical scoring function. Electrophoresis. 2007;28(9):1387–92. doi:10.1002/elps.200600637.

Jackson BP, Ranville JF, Neal AL. Application of flow field flow fractionation-ICPMS for the study of uranium binding in bacterial cell suspensions. Anal Chem. 2005;77(5):1393–7. doi:10.1021/ac049278q.

Jones JJ, Stump MJ, Fleming RC, Lay JO Jr, Wilkins CL. Investigation of MALDI-TOF and FT-MS techniques for analysis of *Escherichia coli* whole cells. Anal Chem. 2003;75(6):1340–7.

Kailasa SK, Wu HF. Dispersive liquid-liquid microextraction using functionalized Mg(OH)2 NPs with oleic acid as hydrophobic affinity probes for the analysis of hydrophobic proteins in bacteria by MALDI MS. Analyst. 2012;137(19):4490–6. doi:10.1039/c2an35788g.

Kailasa SK, Wu HF. Surface modified BaTiO3 nanoparticles as the matrix for phospholipids and as extracting probes for LLME of hydrophobic proteins in *Escherichia coli* by MALDI-MS. Talanta. 2013;114:283–90. doi:10.1016/j.talanta.2013.05.032.

Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K-i, Takahashi N, Isobe T. Concanavaline. Nat Biotech. 2003;21(6):667–72.

Kell AJ, Stewart G, Ryan S, Peytavi R, Boissinot M, Huletsky A, Bergeron MG, Simard B. Vancomycin-modified nanoparticles for efficient targeting and preconcentration of Gram-positive and Gram-negative bacteria. ACS Nano. 2008;2(9):1777–88. doi:10.1021/nn700183g.

Kiehntopf M, Melcher F, Hanel I, Eladawy H, Tomaso H. Differentiation of *Campylobacter* species by surface-enhanced laser desorption/ionization-time-of-flight mass spectrometry. Foodborne Pathog Dis. 2011;8(8):875–85. doi:10.1089/fpd.2010.0775.

Krishnamurthy T, Deshpande S, Hewel J, Liu H, Wick CH, Yates Iii JR. Specific identification of *Bacillus anthracis* strains. Int J Mass spectrom. 2007;259(1–3):140–6.

Kull S, Pauly D, Stormann B, Kirchner S, Stammler M, Dorner MB, Lasch P, Naumann D, Dorner BG. Multiplex detection of microbial and plant toxins by immunoaffinity enrichment and matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem. 2010;82(7):2916–24. doi:10.1021/ac902909r.

Lasch P, Beyer W, Nattermann H, Stammler M, Siegbrecht E, Grunow R, Naumann D. Identification of *Bacillus anthracis* by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. Appl Environ Microbiol. 2009;75(22):7229–42. doi:10.1128/aem.00857-09.

Lee H, Williams SKR, Wahl KL, Valentine NB. Analysis of whole bacterial cells by flow field-flow fractionation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Chem. 2003;75(11):2746–52. doi:10.1021/ac020698u.

Lee SY, Yun SH, Lee YG, Choi CW, Leem SH, Park EC, Kim GH, Lee JC, Kim SI. Proteogenomic characterization of antimicrobial resistance in extensively drug-resistant *Acinetobacter baumannii* DU202. J Antimicrob Chemother. 2014;69(6):1483–91. doi:10.1093/jac/dku008.

Li S, Guo Z, Liu Y, Yang Z, Hui HK. Integration of microfiltration and anion-exchange nanoparticles-based magnetic separation with MALDI mass spectrometry for bacterial analysis. Talanta. 2009;80(1):313–20. doi:10.1016/j.talanta.2009.06.069.

Li S, Guo Z, Wu HF, Liu Y, Yang Z, Woo CH. Rapid analysis of Gram-positive bacteria in water via membrane filtration coupled with nanoprobe-based MALDI-MS. Anal Bioanal Chem. 2010;397(6):2465–76. doi:10.1007/s00216-010-3777-6.

Li X, Lv P, Wang L, Guo A, Ma M, Qi X. Application of high resolution pyrolysis gas chromatography/mass spectrometry (HRPGC/MS) for detecting Listeria monocytogenes. J Chromatogr B Anal Technol Biomed Life Sci. 2014;971:107–11. doi:10.1016/j.jchromb.2014.06.032.

Liang S, Schneider RJ. Capillary zone electrophoresis of Cowpea mosaic virus and peak identification. Electrophoresis. 2009;30(9):1572–8. doi:10.1002/elps.200800468.

Lin YS, Tsai PJ, Weng MF, Chen YC. Affinity capture using vancomycin-bound magnetic nanoparticles for the MALDI-MS analysis of bacteria. Anal Chem. 2005;77(6):1753–60. doi:10.1021/ac048990k.

Liu JC, Tsai PJ, Lee YC, Chen YC. Affinity capture of uropathogenic *Escherichia coli* using pigeon ovalbumin-bound $Fe_3O_4@Al_2O_3$ magnetic nanoparticles. Anal Chem. 2008;80(14):5425–32. doi:10.1021/ac800487v.

Liu JC, Chen WJ, Li CW, Mong KK, Tsai PJ, Tsai TL, Lee YC, Chen YC. Identification of *Pseudomonas aeruginosa* using functional magnetic nanoparticle-based affinity capture combined with MALDI MS analysis. Analyst. 2009;134(10):2087–94. doi:10.1039/b908069d.

Lundquist M, Caspersen MB, Wikstrom P, Forsman M. Discrimination of *Francisella tularensis* subspecies using surface enhanced laser desorption ionization mass spectrometry and multivariate data analysis. FEMS Microbiol Lett. 2005;243(1):303–10. doi:10.1016/j.femsle.2004.12.020.

Madonna AJ, Basile F, Ferrer I, Meetani MA, Rees JC, Voorhees KJ. On-probe sample pretreatment for the detection of proteins above 15 KDa from whole cell bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom RCM. 2000;14(23):2220–9. doi:10.1002/1097-0231(20001215)14:23<2220::aid-rcm155>3.0.co;2-4.

Madonna AJ, Basile F, Furlong E, Voorhees KJ. Detection of bacteria from biological mixtures using immunomagnetic separation combined with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2001;15(13):1068–74. doi:10.1002/rcm.344.

Magnuson ML, Owens JH, Kelty CA. Characterization of *Cryptosporidium parvum* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2000;66(11):4720–4.

Mandal SM, Pati BR, Ghosh AK, Das AK. Letter: influence of experimental parameters on identification of whole cell Rhizobium by matrix-assisted laser desorption/ionization time-of-flight

mass spectrometry. Eur J Mass Spectrom (Chichester Eng). 2007;13(2):165–71. doi:10.1255/ejms.842.

Mas S, Villas-Boas SG, Hansen ME, Akesson M, Nielsen J. A comparison of direct infusion MS and GC-MS for metabolic footprinting of yeast mutants. Biotechnol Bioeng. 2007;96(5):1014–22. doi:10.1002/bit.21194.

Meetani MA, Voorhees KJ. MALDI mass spectrometry analysis of high molecular weight proteins from whole bacterial cells: pretreatment of samples with surfactants. J Am Soc Mass Spectrom. 2005;16(9):1422–6. doi:10.1016/j.jasms.2005.04.004.

Moura H, Woolfitt AR, Carvalho MG, Pavlopoulos A, Teixeira LM, Satten GA, Barr JR. MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. Fems Immunol Med Microbiol. 2008;53(3):333–42. doi:10.1111/j.1574-695X.2008.00428.x.

Nilsson CL. Fingerprinting of Helicobacter pylori strains by matrix-assisted laser desorption/ionization mass spectrometric analysis. Rapid Commun Mass Spectrom RCM. 1999;13(11):1067–71. doi:10.1002/(sici)1097-0231(19990615)13:11<1067::aid-rcm612>3.0.co;2-n.

Ochoa ML, Harrington PB. Immunomagnetic isolation of enterohemorrhagic *Escherichia coli* O157:H7 from ground beef and identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and database searches. Anal Chem. 2005;77(16):5258–67. doi:10.1021/ac0502596.

Pandey BK, Ahmad A, Asthana N, Azmi S, Srivastava RM, Srivastava S, Verma R, Vishwakarma AL, Ghosh JK. Cell-selective lysis by novel analogues of melittin against human red blood cells and *Escherichia coli*. BioChemistry. 2010;49(36):7920–9. doi:10.1021/bi100729m.

Petr J, Ryparova O, Ranc V, Hinnerova P, Znaleziona J, Kowalska M, Knob R, Maier V, Frebort I, Lemr K, Sevcik J. Assessment of CE for the identification of microorganisms. Electrophoresis. 2009;30(3):444–9. doi:10.1002/elps.200800232.

Pineda FJ, Lin JS, Fenselau C, Demirev PA. Testing the significance of microorganism identification by mass spectrometry and proteome database search. Anal Chem. 2000;72(16):3739–44.

Pineda FJ, Antoine MD, Demirev PA, Feldman AB, Jackman J, Longenecker M, Lin JS. Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. Anal Chem. 2003;75(15):3817–22. doi:10.1021/ac034069b.

Raftery MJ. Enrichment by organomercurial agarose and identification of cys-containing peptides from yeast cell lysates. Anal Chem. 2008;80(9):3334–41. doi:10.1021/ac702539q.

Reddy PM, Chang K-C, Liu Z-J, Chen C-T, Ho Y-P. Evaluation of functionalized magnetic iron oxide (Fe_3O_4) nanoparticles for capturing gram-positive and gram-negative bacteria. J Biomed Nanotechnol. 2014;10:11.

Reolon LA, Martello CL, Schrank IS, Ferreira HB. Survey of surface proteins from the pathogenic *Mycoplasma hyopneumoniae* strain 7448 using a biotin cell surface labeling approach. Plos One. 2014;9(11):e112596. doi:10.1371/journal.pone.0112596.

Ruelle V, Moualij BE, Zorzi W, Ledent P, Pauw ED. Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2004a;18(18):2013–9. doi:10.1002/rcm.1584.

Ruelle V, Moualij BE, Zorzi W, Ledent P, Pauw ED. Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2004b;18(18):2013–9. doi:10.1002/rcm.1584.

Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal Chem. 2001;73(4):746–50. doi:10.1021/ac0008791.

Scheurer SB, Rybak JN, Roesli C, Brunisholz RA, Potthast F, Schlapbach R, Neri D, Elia G. Identification and relative quantification of membrane proteins by surface biotinylation and two-dimensional peptide mapping. Proteomics. 2005;5(11):2718–28. doi:10.1002/pmic.200401163.

Schlosser G, Kačer P, Kuzma M, Szilágyi Z, Sorrentino A, Manzo C, Pizzano R, Malorni L, Pocsfalvi G. Coupling immunomagnetic separation on magnetic beads with matrix-assisted laser desorption ionization-time of flight mass spectrometry for detection of Staphylococcal Enterotoxin B. Appl Environ Microbiol. 2007;73(21):6945–52. doi:10.1128/aem.01136-07.

Seibold E, Bogumil R, Vorderwulbecke S, Dahouk S A, Buckendahl A, Tomaso H, Splettstoesser W. Optimized application of surface-enhanced laser desorption/ionization time-of-flight MS to differentiate *Francisella tularensis* at the level of subspecies and individual strains. FEMS Immunol Med Microbiol. 2007;49(3):364–73. doi:10.1111/j.1574-695X.2007.00216.x.

Sharma R, Dill BD, Chourey K, Shah M, VerBerkmoes NC, Hettich RL. Coupling a detergent lysis/cleanup methodology with intact protein fractionation for enhanced proteome characterization. J Proteome Res. 2012;11(12):6008–18. doi:10.1021/pr300709k.

Sharon N. Carbohydrates as future anti-adhesion drugs for infectious diseases. Biochim Biophys Acta. 2006;1760(4):527–37. doi:10.1016/j.bbagen.2005.12.008.

Shaw EI, Moura H, Woolfitt AR, Ospina M, Thompson HA, Barr JR. Identification of biomarkers of whole *Coxiella burnetii* phase I by MALDI-TOF mass spectrometry. Anal Chem. 2004;76(14):4017–22. doi:10.1021/ac030364k.

Smole SC, King LA, Leopold PE, Arbeit RD. Sample preparation of Gram-positive bacteria for identification by matrix assisted laser desorption/ionization time-of-flight. J Microbiol Methods. 2002;48(2–3):107–15.

Sun X, Jia H-L, Xiao C-L, Yin X-F, Yang X-Y, Lu J, He X, Li N, Li H, He Q-Y. Bacterial proteome of *Streptococcus pneumoniae* through multidimensional separations coupled with LC-MS/MS. OMICS J Integr Biol. 2011;15(7–8):477–82. doi:10.1089/omi.2010.0113.

Tait E, Perry JD, Stanforth SP, Dean JR (2014) Bacteria detection based on the evolution of enzyme-generated volatile organic compounds: determination of *Listeria monocytogenes* in milk samples. Analytica Chimica Acta. 848:80–7. doi:10.1016/j.aca.2014.07.029.

Teramoto K, Sato H, Sun L, Torimura M, Tao H. A simple intact protein analysis by MALDI-MS for characterization of ribosomal proteins of two genome-sequenced lactic acid bacteria and verification of their amino acid sequences. J Proteome Res. 2007;6(10):3899–907. doi:10.1021/pr070218 l.

Thompson MR, Chourey K, Froelich JM, Erickson BK, VerBerkmoes NC, Hettich RL. Experimental approach for deep proteome measurements from small-scale microbial biomass samples. Anal Chem. 2008;80(24):9517–25. doi:10.1021/ac801707s.

Vargha M, Takats Z, Konopka A, Nakatsu CH. Optimization of MALDI-TOF MS for strain level differentiation of Arthrobacter isolates. J Microbiol Methods. 2006;66(3):399–409. doi:10.1016/j.mimet.2006.01.006.

Victor Ryzhov YH, Catherine F. Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl Environ Microbiol. 2000;66(9):3828–34.

Wang S, Regnier FE. Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography. J Chromatogr A. 2001;924(1–2):345–57.

Wang S, Zhang X, Regnier FE. Quantitative proteomics strategy involving the selection of peptides containing both cysteine and histidine from tryptic digests of cell lysates. J Chromatogr A. 2002;949(1–2):153–62.

Williams TL, Andrzejewski D, Lay JO, Musser SM. Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. J Am Soc Mass Spectrom. 2003;14(4):342–51. doi:10.1016/s1044-0305(03)00065-5.

Yang YC, Yu H, Xiao DW, Liu H, Hu Q, Huang B, Liao WJ, Huang WF. Rapid identification of *Staphylococcus aureus* by surface enhanced laser desorption and ionization time of flight mass spectrometry. J Microbiol Methods. 2009;77(2):202–6. doi:10.1016/j.mimet.2009.02.004.

Zscheppank C, Wiegand HL, Lenzen C, Wingender J, Telgheder U. Investigation of volatile metabolites during growth of *Escherichia coli* and *Pseudomonas aeruginosa* by needle trap-GC-MS. Anal Bioanal Chem. 2014;406(26):6617–28. doi:10.1007/s00216-014-8111-2.

Chapter 4
Advantages Offered by Proteomic Strategies for Rapid Biodetection

Catherine Fenselau

Mass spectrometry has been applied for the analysis of microorganisms in the field and in the clinic because of its speed, sensitivity, broadband capability, specificity, and automation. It is a physicochemical method that provides complementary information orthogonal to biochemical and morphological methods. We claim broadband because everything has a mass and thus a mass spectrum. Other chemical and biological detectors ask, *"Is it there?"* The mass spectrometer asks, *"What is there?"* To achieve specificity the mass spectrometer measures an intrinsic property—mass. Both molecular and fragment masses provide detailed fingerprints, which can also be interpreted. These capabilities allow mass spectrometry to be applied to achieve rapid characterization of microorganisms in biodefense, space exploration, medical diagnostics, food safety, drug safety, environmental and workplace monitoring, antibiotic susceptibility testing, biotechnology process control, and others.

Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) instruments have been widely used for these applications because of their speed, simplicity, ruggedness, and capability to produce spectra from unfractionated cells that reveal protein biomarkers. Such spectra allow bacteria and other targets to be identified by sophisticated library searching or pattern recognition. This approach requires highly controlled instrumental and experimental conditions. Also, it requires that the target has already been defined and its spectrum put into the library.

An alternative, interpretive approach uses proteomic strategies and bioinformatics, referencing the unknown spectrum obtained under any conditions to a database intrinsic to each organism, i.e., derived from each organism's genome. Proteomic approaches, which relate observed protein or peptide biomarkers to sequence databases, offer a number of advantages. They take advantage of the ever-growing genomic and protein sequence repositories under development around the world. They provide identifications that are not dependent on growth conditions, sample preparation, or matrix selection. One database supports all ionization techniques

C. Fenselau (✉)
Department of Chemistry & Biochemistry, University of Maryland, College Park, MD 20742, USA
e-mail: Fenselau@umd.edu

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_4

and mass analyzers. Proteomics and bioinformatics enable the identification of specific biomarkers and can determine the uniqueness of biomarkers identified, thus increasing specificity in the analysis of strains and enabling forensic applications. Currently, proteomics strategies have been applied successfully in the rapid analysis of unfractionated vegetative bacteria, spores, viruses, fungi, and toxins. Various workflows have been demonstrated (Fenselau et al. 2007) and the most rigorous are:

- Identification of proteins and species based on microsequences (partial sequences) of peptides
- Identification of proteins, species, and strains based on protein microsequences
- Identification based on combined protein masses and peptide microsequencing

Genome and protein sequences are most authoritatively available in the UniProt knowledge system, into which most public information has converged in recent years. Unfortunately, some sequences are held in private hands for commercial gain.

This chapter will illustrate the greater flexibility provided by proteomic interpretation of whole-cell spectra and the superior capabilities of proteomic approaches to address three major challenges in biodetection—analysis of components in mixtures, identification of genetically engineered microorganisms, and characterization of bacteria without sequenced genomes. These kinds of samples cannot be reliably analyzed by library searching.

Identification of Components in Mixtures

Most environments contain a complex mixture of background bacteria and spores against which a bioagent or a workplace contaminant must be identified. Many clinical samples also contain mixtures of microorganisms. Thus, mixture analysis may be more the norm than the exception in rapid analysis. Traditional approaches to the characterization of components of mixtures include selection of single colonies on culture plates and selective biocapture of predetermined targets using antibodies or adhesins. Proteomic strategies can be successful without purification if they identify proteins that are characteristic of the different components. This can be accomplished with either top-down and bottom-up workflows, both of which require tandem mass spectrometers. A bottom-up workflow is represented in Fig. 4.1, which can be realized on rugged low-resolution tandem instruments. The application of this workflow (on a MALDI ion trap-TOF tandem instrument) to a synthetic mixture of *Bacillus thuringiensis* subs. Kurstaki and *Bacillus subtilis* 168 spores is illustrated in Fig. 4.2. Tryptic peptides are formed in situ and characterized from protein biomarkers for both workflows. The approach requires additional sample manipulation in situ, including proteolysis to produce peptides for sequence analysis, and has been automated for analysis of spores in field trials on an atmospheric pressure (AP) MALDI ion trap (personal communication from Dr. Vladimir Doroshenko). Advantages of working with peptides (bottom-up) include that smaller

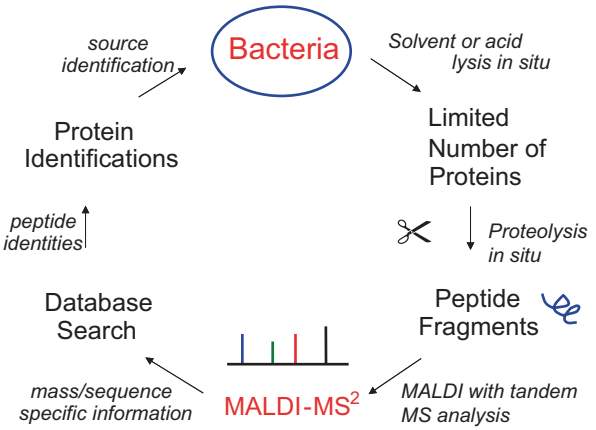


Fig. 4.1 A proteomic strategy for rapid identification of bacteria in mixtures. (Provided by Prof. Bettina Warschied)

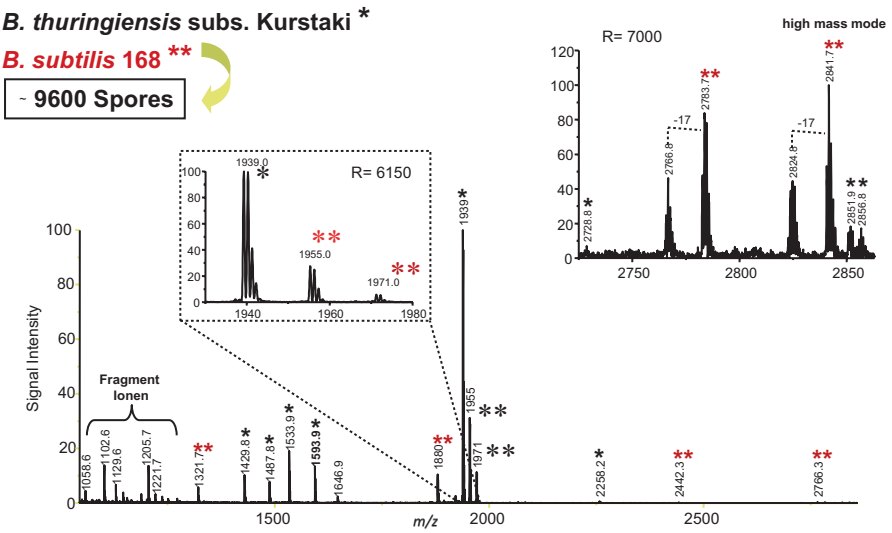


Fig. 4.2 MALDI spectrum of tryptic peptides formed on the sample holder from a mixture of *B. thuringiensis* subs. Kurstaki and *B. subtilis* 168 spores totaling around 9600 spores in a ratio of 10:1. All the annotated peptides were subsequently selected for collisional activation and protein assignment. (Adapted from Warscheid et al. 2003)

ions are measured with greater sensitivity and greater mass accuracy in field portable systems. Top-down analysis can proceed with only lysis of the organism and desalting. In addition, it provides more reliable analysis of mixtures of proteins and cells. However, fragmentation of intact proteins (top-down) using either MALDI or electrospray ionization (ESI) and decharging precursor and fragment ions for

bioinformatics currently requires more complex instruments and is better suited for reference and regulatory laboratories than for implementation in the field (e.g., see Demirev et al. 2005; Fagerquist et al. 2010; McFarland et al. 2014; Wynne et al. 2009).

Rapid Detection of Genetic Engineering in Bacteria

One obvious approach to recognizing a genetically modified organism is to guess what protein is to be expressed and target it. Guessing is challenging however, since the protein might be a toxin, a sensitizer, an antivaccine, or other. A more reliable strategy might be to look for the machinery required for genetic modification. Plasmids are commonly used to introduce guest genes into bacteria, and "off-the-shelf" plasmids are available in the public sector. Plasmids also carry genes that produce proteins required for successful transfection and expression, and genes for a small number of proteins that confer resistance to antibiotics (Russell et al. 2007). Thus, rapid identification of the resistance protein beta-lactamase in *Escherichia coli* would suggest that host cell *E.coli* had been engineered in some way and could trigger caution, and additional analysis. The presence of beta-lactamase correlates with the presence of a plasmid. Resistance to ampicillin, for example, can be tested in a Petri dish in about 24 h (Fig. 4.3), or the presence of beta-lactamase can be directly demonstrated by mass spectrometry in 5–10 min. MALDI-TOF spectra in

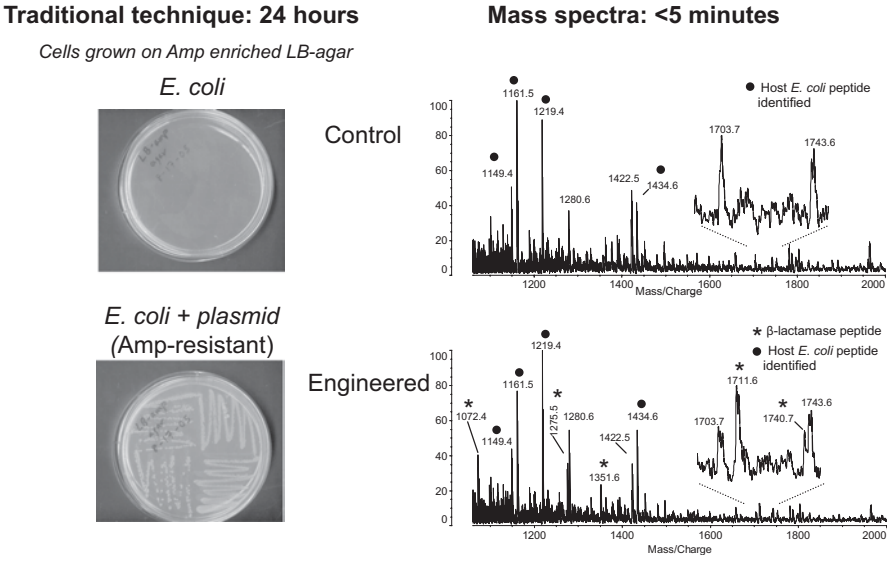
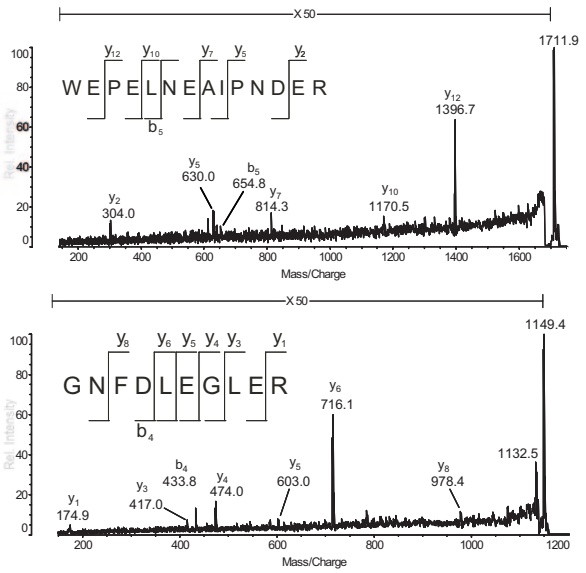


Fig. 4.3 Characterization of beta-lactamase in *E. coli* using *(left)* cell culture and *(right)* MALDI TOF mass spectrometry. The mass spectra show peptides formed by on-probe digestion of *(top) E. coli* and *(bottom) E. coli* transformed with the pPLC28 plasmid. (Adapted from Russell et al. 2007)



peptide from β-lactamase, the antibiotic resistance gene in the plasmid.

4 peptides give protein score of 116 (MASCOT)

peptide from the E.coli protein tryptophanase.

2 peptides give protein score of 73 (MASCOT)

Fig. 4.4 TOF/TOF spectra acquired after on-probe tryptic digestion of *E.coli* ΔH1Δtrp transformed with the pPLc28 plasmid. (Adapted from Russell et al. 2007)

Fig. 4.3 show peptides released by tryptic digestion from *E. coli,* and from *E. coli* transformed with the pPLc28 plasmid. Figure 4.4 presents TOF/TOF spectra of two of these peptides, which are found to be derived from the resistance protein and from a protein associated with the host bacterium *E. coli.*

In many ways, this proteomic strategy is analogous to that applied to mixtures of cells. We want to identify proteins in the sample that have more than one source. The bottom-up approach maximizes sensitivity and mass accuracy for field applications. Top-down proteomics involving fragmentation of intact proteins in more complex instruments identify full-length protein sequences with higher specificity and identify the changes in amino acids that distinguish organisms.

Characterizing Bacteria with Unsequenced Genomes

In the third grand challenge, we point out that the most automatable use of proteomic strategies requires a comprehensive database of protein sequences. The identification of proteins is straightforward, using either bottom-up or top-down proteomics approaches if the genome of the target has been sequenced. It is more challenging when little genomic or protein sequence information is available. One strategy, which can sometimes circumvent this limitation, depends on recognition of proteins with high homology in related species that have been sequenced (Edwards et al. 2011). This makes it possible to provide a phylogenetic context for unsequenced organisms.

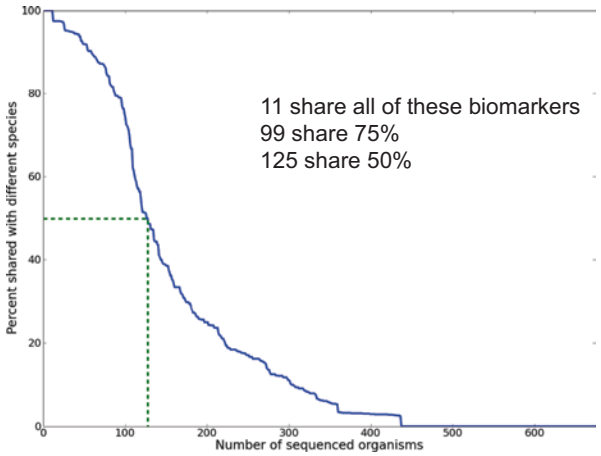


Fig. 4.5 Percentage of 675 sequenced bacteria who share ribosomal proteins (4–16 kDa). (Adapted from Wynne et al. 2009)

Many bacteria carry proteins with similar or identical sequences. This is especially true for ribosomal proteins and is readily observed across the large family of Enterobacteriaceae, as has recently been demonstrated for *Yersinia rohdei* (Wynne et al. 2009) , *Erwinia herbicola,* and *Enterobacter cloacae* (Wynne et al. 2010) . *Y. rohdei* is often used as a simulant for the pathogen *Yersinia pestis.* A lysate was prepared for analysis on a high-resolution tandem mass spectrometer, in this case equipped with an electrospray source (ESI) and an interfaced high-performance liquid chromatography (HPLC). Proteins were identified by searching decharged precursor and product ions against a custom-made library that included all the Yersinia protein sequences publically available at the time and also sequences from six other Enterobacteriaceae. Table 4.1 lists ten proteins that were identified in *Y. rohdei,* and other Enterobacteriaceae species in which identical proteins occur. Based on this table, a phylogenetic tree was constructed, which placed *Y. rohdei* in the same clade as a phylogenetic tree based on 16S-RNA sequences (Wynne et al. 2009) . Finally, to point out the value of looking for shared biomarkers, the graph in Fig. 4.5 evaluates the overlap of ribosomal proteins (4–16 kDa) among 675 organisms with sequenced genomes. The dotted line indicates that 127 of these organisms share at least 50 % of these biomarkers with other species. Of this set of sequenced organisms, 201 share 25 % of their potential ribosomal biomarkers and 238 share none. Although recently sequenced organisms are not represented in this analysis, the graph suggests that we can expect that many, but not all, unsequenced organisms will be characterized by identification of shared protein biomarkers.

This objective is best approached using a top-down strategy, because the mass of the intact protein is an important constraint in the database searches, and can define mutations and modifications. Top-down protein analysis has been successfully carried out on all tandem high-resolution analyzers, and with both MALDI and ESI.

Table 4.1 Proteins and their sources matched by top-down spectra from *Y. rohdei* searched against a custom library. (Adapted from Wynne et al. 2009)

m/z	Charge	Number of matching fragments	Observed mass	Theoretical mass	Protein description	Organism	E value
756.70	8	27	6044.11	6044.82	50s Ribosomal protein L32	*Yersinia enterocolitica, Yersinia pseudotuberculosis, Yersinia pestis, Yersinia pestis* strain Antiqua	2.49E-36
781.10	8	13	6239.55	6240.4	50s Ribosomal protein L33	*Enterobacter* strain 638, *Salmonella typhimurium, Shigella sonnei, Klebsiella pneumoniae*	8.15E-16
802.90	8	22	6413.56	6414.6	50s Ribosomal protein L30	*Yersinia pestis* (strain Antiqua), *Yersinia enterocolitica, Yersinia pestis, Yersinia pseudotuberculosis*	6.00E-22
857.72	8	9	6852.67	6852.95	Carbon storage regulator	*Yersinia pestis* (strain Nepal516), *Yersinia pseudotuberculosis, Yersinia enterocolitica, Yersinia pestis*	5.84E-09
807.80	9	24	7260.92	7261.41	50s Ribosomal protein L29	*Yersinia enterocolitica*	6.79E-24
763.10	11	8	8368.61	8368.77	30s Ribosomal protein S21	*Salmonella typhimurium, Yersinia pestis* (strain Antiqua), *Yersinia pseudotuberculosis, Yersinia pestis, Enterobacter* strain 638, *Shigella sonnei, Yersinia enterocolitica, Klebsiella pneumoniae*	2.82E-05
682.76	13	14	8862.89	8863.32	50s Ribosomal protein L28	*Yersinia enterocolitica*	1.64E-12

Table 4.1 (continued)

m/z	Charge	Number of matching fragments	Observed mass	Theoretical mass	Protein description	Organism	E value
643.22	14	7	8991.92	8992.34	50s Ribosomal protein L27	*Yersinia pseudotuberculosis, Shigella sonnei, Yersinia pestis* (strain Antiqua), *Yersinia pestis* (strain Nepal), *Yersinia pestis*	6.06E-05
1105.83	11	8	12155.73	12156.2	50s Ribosomal protein L22	*Yersinia enterocolitica*	1.78E-06

Going Forward

This chapter illustrates significant advantages offered by proteomic strategies (in contrast to library matching) for the identification and characterization of bacteria and other microorganisms. The point is made in other chapters of this book that proteomics strategies also offer indispensable benefits in both establishing and using protocols to characterize strains and to measure antibiotic resistance. In addition to resistance, mass spectrometry-based proteomics will continue to provide answers to important questions about the biochemistry of microorganisms, biofilms, and microbiomes. Because of the expanding market in this area, manufacturers are expected to continue to improve the speed, sensitivity, reliability, and automation of appropriate commercial instruments.

References

Demirev P, Feldman AB, Kowalski P, Lin JS. Top-down proteomics for rapid identification of intact microogranisms. Anal Chem. 2005;77:7455–61.

Edwards N, Wynne D, Dhabaria A, Fenselau C. Top-down protein analysis and phylogenetic characterization of unsequenced bacteria. In: Fenselau C, Demirev P, editors. Rapid characterization of microorganisms by mass spectrometry. Washington, DC: American Chemical Society; 2011. p. 121.

Fagerquist CK, Garbus BR, Miller WG, Williams KE, Yee E, Bates AH, Boyle S, Harden LA, Cooley MB, Mandrell RE. Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. Anal Chem. 2010;82:2717–25.

Fenselau C, Russell S, Swatkoski S, Edwards N. Proteomic strategies for rapid characterization of micro-organisms. Eur J Mass Spectrom. 2007;13:35–9.

McFarland MA, Andrzejewski D, Musser SM, Callahan JH. Platform for identification of *Salmonella* serovar differentiating bacterial proteins by top-down mass spectrometry: *S. Typhimurium* vs *S. Heidelberg*. Anal Chem. 2014;86:6879–86.

Russell SC, Edwards N, Fenselau C. Detection of plasmid insertion in *Escherichia coli* by MALDI-TOF mass spectrometry. Anal Chem. 2007;79:5399–406.

Warscheid B, Jackson K, Sutton C, Fenselau C. MALDI analysis of *Bacilli* in spore mixtures by applying a quadrupole ion trap time-of-flight tandem mass spectrometer. Anal Chem. 2003;75:5608–17.

Wynne C, Fenselau C, Demirev PA, Edwards N. Top-down identification of protein biomarkers in bacteria with unsequenced genomes. Anal Chem. 2009;81:9633–42.

Wynne C, Edwards N, Fenselau C. Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an orbitrap. Proteomics. 2010;10:3631–43.

Chapter 5
Bottom-Up Proteomics Methods for Strain-Level Typing and Identification of Bacteria

Jacek P. Dworzanski

Introduction

The microbiological methods used for the detection and identification of bacteria were by necessity based on culturing and staining techniques combined with microscopic evaluation of cells. However, over the past few decades the use of molecular methods gained importance in microbiological laboratories and led to tremendous changes in a way of detecting microorganisms, their identification at the species level, and typing of isolates to infer subspecies diversity. Although routine identification methods continue to be based on the determination of the morphology, differential staining, and physiology of a microbial isolate, currently these methods are gradually supplanted by the use of diverse genomic and proteomic-based approaches that include mass spectrometry (MS) techniques, among others.

MS-based methods represent a broad group of highly versatile approaches that use precise mass measurements to infer identity of diverse biomolecules. Although for many decades the scope of investigated molecules was limited by their molecular mass and polarity, developments in soft ionization techniques like electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) substantially broaden the range of investigated species. Nowadays not only proteins and nucleic acids but also multimolecular complexes, and even whole viruses can be mass analyzed by modern MS instruments and used to infer genomic information encoded in nucleotide and amino acid sequences. Therefore, MS-based analysis of nucleic acid amplicons and proteins is increasingly replacing the older, time-consuming, and labor-intensive approaches.

Currently, both "top-down" and "bottom-up" methods are used to analyze microbial proteins by MS. In top-down approach, proteins are analyzed to determine molecular masses of intact proteins and to characterize them by using gas-phase fragmentation techniques. The bottom-up characterization of proteins uses prote-

J. P. Dworzanski (✉)
Leidos, Inc., 1816 Queen Anne Sq, Bel Air, MD 21015, USA
e-mail: jacek.p.dworzanski@gmail.com

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_5

olysis and analysis of released peptides by tandem MS to reveal their amino acid sequences. Bottom-up analysis of a protein mixture is usually called shotgun proteomics, to indicate analogy to shotgun genomic sequencing (Yates 1998).

Molecular criteria used for defining bacterial species have been progressing from the determination of nucleotide content (e.g., mol% G-C), DNA–DNA hybridization (DDH) and melting point analysis, which measure the degree of genetic similarity between two genomes, 16S rRNA gene sequencing, and multilocus sequence analysis (MLSA) of housekeeping genes, up to proteomics-based analysis and sequencing the whole microbial genome.

Since the 1960s, a means for determining relatedness of strains was based on a comparison of genomic similarities measured by DDH between DNA strands. DDH has driven the construction of current bacterial taxonomy and has become a gold standard for the delineation of bacterial species, which were defined as a collection of strains with a DDH value of at least 70% (Wayne et al. 1987). However, these methods are difficult and laborious; therefore, other genomic approaches were developed to replace DDH, including DNA sequencing by hybridization with custom-designed microarrays, or comparison of 16S rRNA sequences used with the assumption that if strains share less than 97% of sequence similarity, they belong to different species (Stackebrandt and Goebel 1994). In fact, sequencing of 16S rRNA combined with searching a database (DB) with millions of entries has become the most commonly used method for identifying and classifying microbial species (Cole et al. 2009; Quast et al. 2013). However, the 16S rRNA gene has limited specificity, for example, only 80% of isolates were recently found to be unambiguously identified at the species level (Chatellier et al. 2014). Therefore, genes with less-conserved sequences from protein-coding loci, for example, DNA gyrase *(gyrB)* or RNA polymerase *(rpoB)* have to be used instead. Unfortunately, different genes may give different patterns of interspecies relationships due to horizontal gene transfer (HGT) or unequal rates of nucleotide substitution. Therefore, sequence analysis of 6−8 housekeeping genes (a multilocus approach) was designed to increase the resolution and to buffer the potential impact of the HGT on the determined relatedness. Despite being successful in phylogenetic discrimination of strains at the subspecies level, it has major drawbacks that arise from a putative bias in gene selection and amplification primer availability.

The universally adopted genomic approaches to strain subtyping still use DNA fingerprinting techniques based on: (i) analysis of restriction nuclease digested whole cell DNA fragments resolved by pulsed-field gel electrophoresis (PFGE), or (ii) polymerase chain reaction (PCR)-amplified segments targeting loci with a variable number of tandem repeats (VNTR), which reveal relatedness at a microevolutionary level by using the technique called multiple locus VNTR analysis (MLVA). These DNA fingerprinting techniques are used for high accuracy isolate characterization, for example, by the Centers for Disease Control and Prevention PulseNet program (http://www.cdc.gov/pulsenet/pathogens/pfge.html) to recognize, investigate, and control outbreaks of food infections. However these procedures are also quite lengthy. For example, a standard operating procedure of PFGE takes up to 5 days and includes the isolation and growth of the culture, cell lysis, digestion

of DNA with restriction nuclease in agarose gel, followed by gel electrophoresis; staining and documentation of a gel. The other fingerprinting procedure used to discriminate between closely related strains is called "optical mapping" or whole-genome mapping, which provides maps of a chromosome based on optical analysis of DNA fragments obtained by digestion with a restriction nuclease. The experimental data obtained by this technique can be correlated directly to DNA sequence information in the public databases so that markers for resistance or virulence can be easily recognized (Miller 2013).

Reliable characterization of microorganisms at subspecies level is increasingly essential in clinical, biotechnological, environmental, and epidemiological studies. Currently, reliable characterization of microorganisms is based on genetic and genomic criteria inferred from complete genome sequences, considered as the reference standard for determining bacterial phylogenies. The most widely adopted tools for comparing and analyzing complete genome sequences are based on in silico calculation of digital DDH-type indices representing conservation of the core genome, the DNA content measured as the proportion of DNA shared by two genomes (Goris et al. 2007), and alignment-free approaches using oligonucleotide frequencies for phylogenomic inferences (Bohlin et al. 2008). The DDH-type indices include average nucleotide identity (ANI) of all orthologous genes shared by two genomes (Konstantinidis and Tiedje 2005a) or its equivalent, average amino acid identity (AAI) of protein-coding genes (Konstantinidis and Tiedje 2005b), calculated using BLAST or BLASTP algorithms; the maximal unique matches index (MUMi) (Deloger et al. 2009); and refinements of these approaches, for instance, by using the rapid alignment tool MUMer (Richter and Rosselló-Móra 2009). More recently, a similar method called the genome BLAST distance phylogeny makes use of DNA rather than genes and uses a set of local alignment tools and a special formula to calculate a genome-to-genome distance (Meier-Kolthoff et al. 2013).

Of these, ANI and AAI indices have been most widely used as possible next-generation gold standards for species delineation because they represent a robust measure of the genetic distance between two sequenced bacterial strains and are strongly correlated with DDH data. In addition, they are also strongly correlated with 16S rRNA gene sequence similarity, the percentage of conserved DNA, the mutation rate of the genome, and offer resolution at the subspecies level (Konstantinidis and Tiedje 2005a, b; Goris et al. 2007). However, the major drawback of this approach is that it is only available for a pair of strains with complete genome sequences.

Importantly, both AAI index and DDH values for bacterial strains can be predicted experimentally by using a proteomics-derived index termed the fraction of shared (tryptic) peptides (FSP, Dworzanski et al. 2010). FSP is calculated from the peptide-centric bottom-up proteomics MS data sets acquired during analysis of an unknown bacterial strain and searched, with a suitable search engine, against DB proteomes predicted from complete genome sequences of reference strains. In this approach, the high-throughput proteome identification of thousands of released peptides reveals amino acid sequence information translated from genomic sequences that may be used not only for predicting strain similarities but also for

identifications of genes that are actually expressed. Consequently, bottom-up proteomics allows high-resolution typing and subspecies level identifications, reflecting both genomic similarities and supplanting traditional typing approaches based on serological (e.g., H-antigen typing) and phenotypic properties, like antibiotic resistance.

Currently, the important pieces of information about an isolated bacterial strain, that is, the species, serovar, subtype, or its antibiotic resistance are generated by separate tests. However, the bottom-up proteomics analysis potentially allows finding this kind of information in just one test comprising liquid chromatography (LC)-MS/MS analysis and data mining with a suite of bioinformatics tools. In this chapter, I will focus on a group of highly versatile bottom-up shotgun-proteomics methods allowing for the identification, classification, and characterization of microorganisms by revealing: (i) strain identity, (ii) serotype, (iii) virulence, (iv) antimicrobial resistance profile, and (v) a subtype reflecting differences in both the gene content and single amino acid variations (SAVs) of expressed proteins.

Cell Harvesting and Protein Extraction

Samples analyzed by MS-based approaches for bacteria identification are initially processed in microbiological laboratory settings; therefore, researchers should follow standard procedures used for sample collection and preconcentration, and these methods will not be discussed here.

Generally, clinical, environmental, or food samples are processed to obtain pure cultures either directly, for example, from blood samples, or by isolating them from other cells and/or food and environmental matrices using diverse enrichment techniques. Such cells are then grown to obtain pure cultures by using diverse selective or enriched liquid and agar-solidified media supporting the growth of a wide range of microorganisms. The microbial cells are then harvested, washed with buffers or distilled water, and processed to extract their proteins for further proteomic analysis. The sample processing steps usually follow standard protocols developed for shotgun-proteomics workflows that include microbial cell lysis, extraction, solubilization and preseparation of proteins, specific cleavage of proteins into peptides, and peptide purification and separation immediately prior to MS analysis (Gundry et al. 2009). However, depending on the infectability of the material, all the steps preceding peptide analysis should be carried out in a laboratory approved for working with infectious agents.

Cell Lysis and the Preparation of Whole Cell Protein Extracts

Microbial cell lysis provides access to cytosolic and the majority of membrane proteins, and therefore is a crucial step for efficient extraction of expressed proteins and

their analysis by shotgun-proteomics methods. Such whole cell protein extracts are usually obtained by rupturing cells in lysis buffers containing protease inhibitors by using physical methods, such as ultrasonication, bead beating, French press, freeze-thaw, thermal lysis, pressure cycling, and (bio)chemical lysis procedures, involving murolytic enzymes like lysozyme, detergents, chaotropes, and other reagents. However, in case of biochemical and chemical methods, the compatibility of (bio) chemical reagents with the analytical technique must be considered. For example, although lysozyme is very effective in lysing Gram-positive bacteria by hydrolyzing glycosidic linkages in the bacterial wall peptidoglycan, it may interfere with the identifications of peptides obtained by global digestion of protein extracts.

In general, the obtained lysates are cleared by centrifugation to remove cellular debris, and the supernatant or "supernate" is considered a whole cell protein extract composed of a complex mixture of proteins, other cell constituents such as lipids, nucleic acids, polysaccharides, low molecular mass metabolites, and all additives. These additives include buffers, chaotropes, detergents, or cocktails of proteinase inhibitors, which are added to aid in protein extraction and preserve the integrity of a proteome.

Cell lysis can also include a combination of chemical and diverse physical methods. For example, Lee et al. (2006) demonstrated rapid lysis of bacterial cells using both thermal and mechanical lysis directly on a chip through a combination of the laser irradiation and agitation with magnetic beads. This and many other microfluidic devices for cell lysis were recently reviewed by Nan et al. (2014). In another example, Napoli et al. (2014) performed cellular lysis of bacteria through the frictional action of glass beads added to the sample solution combined with pressure waves provided by a probe sonication of a cells/glass beads mixture. However, sonication becomes problematic for lysis of pathogenic microorganisms due to safety concerns, and is not well adapted for automated, high-throughput liquid-handling platforms. An approach aimed to overcome such concerns was proposed and tested by Tanca et al. (2013) in their comparative study of sample preparation workflows. They extracted proteins from *Escherichia coli* by subjecting cells to lysis in buffered solutions of surfactants for 30 min at 95 °C by using a thermo-mixer at 500 rev/min.

For highly pathogenic microorganisms which should be handled in the biosafety level 3 (BSL-3) laboratory, Tracz et al. (2013) grew bacteria with required biocontainment precautions and after harvesting and resuspending cells in sterile water, they were then gamma-inactivated. Further, microbial suspensions were incubated at 95 °C for 5 min and vortexed with glass beads to rupture cells and release proteins. Similar safety precautions were also used by Jabbour et al. (2010b) and Wade et al. (2011) by pelleting the cells from cultures by centrifugation, washing, resuspending in a buffer, and lysing them thermally by a 1-h long heating at 95 °C. In addition, a portion of each lysed sample was plated and incubated for 5 days to ensure no growth prior to removing samples from the BSL-2 or BSL-3 laboratory. However, for lysing enterohemorrhagic and enteroaggregative *E. coli* strains, Jabbour et al. (2014) used the bead beating technique.

The choice of a lysis method may also be tailored for a specific group of microorganisms. For example, François et al. (2014) prepared a total protein extract from *Staphylococcus aureus* by resuspending harvested cells in a lysis buffer containing calcium and magnesium chlorides and protease inhibitors. By adding the murolytic enzyme lysostaphin that cleaves crosslinking pentaglycin bridges in the cell wall of *Staphylococci,* they released protoplasts that immediately underwent lysis due to hyposmotic shock. The presence of a high-molecular DNA in such samples causes high viscosity that may be reduced by adding DNase; however, this contaminates the sample and may complicate sample processing workflows.

Preparation of Subcellular Fractions

Among subcellular proteomes investigated for identification of bacterial subspecies, attention was concentrated on surface and membrane-associated proteins, especially outer membrane proteins (OMPs) of Gram-negative bacteria, surface layer (S-layer) proteins of Gram-positive bacteria, flagella, and extracellular proteins (ECPs).

Outer Membrane Proteins

After cell lysis by ultrasonication or any other method, cell debris is usually removed by centrifugation and the resulting supernatant is assumed to contain the total cellular protein fraction composed of both membrane and the soluble cytosolic proteins. Therefore, in some applications it is advantageous to separate these proteins, for example, by ultracentrifugation, to obtain the pellet corresponding to the membrane fraction. For example, Jabbour et al. (2010b) and Wade et al. (2011)—after thermal lysis and removing cell debris by centrifugation—ultracentrifuged the obtained supernatants at 100,000 g to pellet membrane proteins which they resuspended in a buffered solution of N-lauroylsarcosinate. Because OMPs of Gram-negative bacteria are insoluble in sarcosine solutions, ultracentrifugation of such mixture allows for pelleting OMPs.

A more streamlined procedure for OMP isolation was used by Damron et al. (2009). They simply suspended harvested cells in a buffered solution of sarcosyl with protease inhibitors and lysed cells by sonication on ice. The lysate was then clarified by low-speed centrifugation, and the supernatant was centrifuged at 40,000 g to obtain a pellet containing OMPs.

Among filtration methods, there is a growing popularity of using ultrafiltration centrifugal devices, for example, Microcon$^{(R)}$-type filters (EMD Millipore, Billerica, MA, USA), which allow for removal of lower molecular mass contaminants, buffer exchange, and sample concentration.

Surface Layer Proteins

Surfaces of many microbial cells are coated with a layer of proteins (known as "S-layer") that have an important role in the cell's growth, survival, and interaction with the host organism, and are present in a high copy number. In addition, such proteins are easily available for solvent extraction and could be used for sequence-based identification and typing of microbial cells.

For example, for the identification and typing of *Lactobacillus* spp. used as probiotic bacteria in dietary supplements and milk products, the extraction of S-layer proteins was carried out from the water washed bacterial cells by incubation with 5 M lithium chloride or guanidine hydrochloride solutions (Johnson et al. 2013; Podlesny et al. 2011). After the removal of cells by centrifugation and filtration, the extract may be concentrated by ultrafiltration. The precipitated S-layer proteins are suspended in 1 M lithium chloride to dissociate any proteins which are soluble, and the purified S-layer proteins are pelleted by centrifugation (Goh et al. 2009). Alternatively, S-layer proteins—which are characterized by a high isoelectric point (pI > 9)—may be purified by a cation-exchange chromatography (Podleśny et al. 2011).

Preparation of Flagella

Flagella are isolated from bacteria growing on plates by scraping and suspending in a suitable buffer while those cultivated on liquid media are directly harvested by centrifugation. However, centrifugation may cause cell surface damage through collisions resulting in shear forces on the bacterial cell surface; therefore, it should be performed at low speeds or even avoided. For example, Cheng et al. (2013) harvested a full loopful of enteric bacteria and gently suspended them in a lysozyme solution, followed by vigorous vortexing to shear off flagella and centrifugation to remove cells. The supernatant was filtered through a 0.2 µm pore size low protein binding membrane of a syringe filter to retain and wash flagella with deionized water. Finally, the isolated flagella were on-filter trypsinized by exposing them for a couple of hours to a trypsin solution.

However, Sun et al. (2013) did not use lysozyme in their protocol on isolation of flagella from *Shewanella* cells that produce a single polar flagellum. Therefore, after a vortexing step to shear off flagella and removing the cells by centrifugation, they passed the supernatant containing flagella through a 0.45-µm-pore filter that did not retain them. Consequently, they used ultracentrifugation to pellet purified flagella and re-suspend them in water for further analysis.

Extracellular Proteins

ECPs include proteins that are actively transported to the bacterial outer surroundings through the cytoplasmic membrane, as well as those that are simply shed from

the bacterial surface. Therefore, they are prepared from spent media obtained after harvesting cells by centrifugation. The cell-free media are usually sterilized by filtration, and the ECPs are routinely isolated by precipitation with trichloroacetic acid (TCA), followed by washing with acetone to remove TCA (Sun et al. 2014; Enany et al. 2014; Halbedel et al. 2014). However, ultrafiltration may also be used for concentrating ECPs, for example, by using centrifugal ultrafiltration devices (Jabbour et al. 2014).

Processing of Bacterial Proteins for Bottom-Up Proteomics Analysis

The conventional method of proteome analysis involves gel separation of proteins as the final purification step that is followed by in-gel digestion and mass spectrometric analysis of released peptides (Tonella et al. 2001). This sample preparation method is still widely used in proteomics of bacteria (Hartmann et al. 2014) and although it has many advantages, it is a relatively lengthy and labor-intensive procedure. Therefore, the gel-free, shotgun protein digestion methods are frequently used for faster and more efficient processing of proteins for LC-MS/MS analysis of peptides. However, the shotgun protocols have to deal with highly contaminated samples because proteins extracted from bacterial cells usually contain other cell constituents and reagents, including those used for breaking interactions involved in aggregation of membrane proteins that facilitate their solubilization (see Section "Cell Lysis and the Preparation of Whole Cell Protein Extracts"). The presence of such substances may interfere with further processing and LC-MS analysis; therefore, they have to be removed from the sample before downstream processing, for example, by using solid-phase extraction or precipitation approaches. However, due to the low molecular mass of many reagents and cellular metabolites in comparison to the M_r of proteins, size-exclusion chromatography or ultrafiltration are frequently used to purify protein extracts, especially in spin-column or spin-filter formats, to minimize the time required for sample processing.

Cell Shaving

Surface proteins play a crucial role in the interaction between cells and their environment, and the outermost cell components can be digested for strain identification without previous cell rupturing. In recent years novel approaches have been developed for analysis of such proteins that include, among others, membrane washing, two-phase partitioning, and protein shaving (Zhang et al. 2013a). Protein shaving is based on the direct digestion of live, intact cells under isotonic conditions, so surface-exposed domains of membrane proteins, named the "surfome," are "shaved" by a protease and the released peptides can be analyzed by LC-MS/MS. This way,

the problems with attempting to solubilize the entire membrane are avoided. Methods and approaches used in surfomics for fast identification of surface proteins have been reviewed by Olaya-Abril et al. (2014).

Recently, Karlsson et al. (2012) applied a lipid-based immobilization technique in the microfluidic format to immobilize intact cells of *Helicobacter pylori* and to obtain peptides from their surface-exposed outermost proteins by shaving them with a trypsin solution. The released peptides were successfully analyzed for strain-level discrimination of analyzed samples.

Protein Digestion Methods

Protein digestion is usually carried out through hydrolysis of the amide bonds catalyzed by chemical reagents, such as cyanogen bromide cleaving at methionine residue, acid catalyzed hydrolysis at aspartic acid (Fenselau et al. 2011), the cleavage at tryptophan and tyrosine residues induced by electrochemical oxidation (Basile and Hauser 2011), or enzymatically with endopeptidases. There are many proteolytic enzymes differing by their specificity for cleaving bonds between individual amino acid residues in a protein. However, trypsin—a serine protease which cleaves at the carboxyl side of arginine and lysine—is the most commonly used protease for protein digestion in shotgun proteomics. Such cleavage specificity gives tryptic peptides a structure that is particularly amenable to informative fragmentation, following ionization and collisional activation in a mass spectrometer. Nevertheless, a combination of highly selective proteases may improve protein and proteome coverage by creating complementary peptides (Wiśniewski and Mann 2012).

In general, the digestion process has to be optimized to achieve maximum efficiency based on a number of parameters affecting the enzymatic reaction that include: (i) solubilization and denaturation of proteins, (ii) reduction of disulfide bonds, (iii) alkylation of reduced cysteines, and (iv) digestion conditions.

Solubilization and Denaturation of Proteins

Adequate solubilization and proper unfolding of proteins in complex microbial extracts are crucial for providing a protease access to cleavage sites. It is especially important in regard to membrane proteins that comprise approximately a quarter of all open-reading frames (ORFs) in typical bacterial genome. They are usually underrepresented in LC-MS proteomics experiments due to poor solubility and lower abundance in comparison to typical cytoplasm proteins. Therefore, the use of diverse solubilization reagents like urea, detergents, and organic solvents has shown to improve digestion efficiency measured as the number of identified peptides and/or sequence coverage of proteins (Mayne et al. 2014).

Detergents are considered the best protein solubilizers, but sodium dodecyl sulfate (SDS) and other conventionally used surfactants are detrimental for LC-MS

analysis and have to be completely removed before analysis. Therefore, surfactant replacement strategies have been developed and are used in many laboratories. The most popular among them are based on filter-aided sample preparation protocols (FASP, Manza et al. 2005; Jabbour et al. 2007, 2010c; Wiśniewski et al. 2009). However, many others methods could be used to remove SDS, for example, ethyl acetate extraction (Yeung et al. 2008), potassium dodecyl sulfate (KDS) precipitation (Zhou et al. 2012), or detergent removal with spin columns (Antharavally et al. 2011; Bereman et al. 2011).

Zhou et al. (2012) compared four in-solution protocols for digestion of whole cell lysates from *Shewanella oneidensis* MR-1. In the first step, they denatured proteins using (1) 8 M urea at 37 °C for 1 h, (2) 50 % trifluoroethanol at 60 °C for 2 h, (3) 1 % SDS at 95 °C for 5 min, and compared them to denaturation with 4 % SDS at 95 °C for 5 min, followed by the FASP protocol that includes SDS exchange by urea prior to sample digestion on a standard ultrafiltration device (Wiśniewski et al. 2009). Samples were then reduced using dithiothreitol (DTT) followed by cysteine alkylation by iodoacetamide (IAA) and after dilution were digested with trypsin. SDS was removed by the KDS precipitation method with KCl. They found only minor differences in sample digestion efficiency among these four methods because LC-MS/MS analyses allowed for identification of more than 4000 peptides from ca. 1000 proteins in each case. This proves that a postdigestion precipitation method could be used as an alternative to predigestion SDS removal by the ultrafiltration-based FASP.

A number of LC- and MS-compatible surfactants, for example, ProteaseMAX, Invitrosol, Rapigest, and PPS Silent Surfactant have also been developed and evaluated to improve protein digestion efficiency. Structures of these commercially available surfactants have an acid labile moiety and, therefore, can be easily degraded prior to LC-MS into components that do not interfere with peptides analysis. For example, Wu et al. (2011) investigated three surfactant-assisted shotgun methods for their applicability to membrane proteome analysis of *E. coli* using acid labile surfactants, sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)-methoxyl]-1-propanesulfonate (RapiGest), PPS, and SDS. They found RapiGest as a preferred reagent for LC-MS/MS analysis of tryptic digests based on the higher number of identified peptides (5799 unique peptides) in comparison to SDS and PPS methods. However, in the study of whole cell protein extracts obtained from *E. coli* cells, Tanca et al. (2013) found that SDS-based buffer outperformed RapiGest in terms of protein extraction yield, and the number of MS identifications and proteome coverage. Therefore, they further tested SDS extracts in five different MS sample preparation workflows, among them, the spin-column detergent removal, followed by in-solution digestion and the FASP method. Although the number of proteins identified among the five tested protocols was comparable (between 1007 and 1104), the FASP dramatically outperformed the competing workflows in the number of identified peptides. For example, with FASP they identified, on average, 7.7 peptides per protein, while the SDS spin-column workflow gave only 4.6 peptides per protein. This indicates the superiority of the FASP procedure for strain identification due to much better proteome coverage.

Waas et al. (2014) investigated the effect of eight commercially available MS-compatible surfactants, two organic solvents, and two chaotropes on the enzymatic digestion efficiency of membrane protein-enriched extract. They found that Progenta™ anionic surfactants—easily cleaved with trifluoroacetic acid (TFA) into small organic molecules that do not exhibit surfactant activity or interfere with analysis by mass spectrometry—outperform other surfactants when tested alone. However, in combinations with guanidine and acetonitrile, all surfactants improved their performance to near similar levels. Nevertheless, the highest number of unique peptides (exceeding 5000) was observed with Invitrosol™, a proprietary surfactant blend manufactured by ThermoFisher Scientific (Waltham, MA, USA), which does not interfere with protease activity and is compatible with reversed-phase (RP) LC-ESI-MS analysis.

The other group of surfactants proven useful for solubilization and digestion of membrane-bound proteins are volatile surfactants like perfluorooctanoic acid that can be easily evaporated prior to LC-MS analysis. As an alternative to surfactants, trifluoroethanol has proven useful for concurrent protein extraction and denaturation for mass-limited samples where sample cleanup is usually detrimental to sensitivity (Wang et al. 2005; Fleurbaaij et al. 2014).

Reduction of Disulfide Bonds and Alkylation of Reduced Cysteines

Thorough protein digestion requires protease access to as many proteolytic sites as possible and is aided by the inclusion of good protein denaturing agents combined with reduction and blocking of free sulfhydryl groups by the alkylation step. Proteins are usually reduced with DTT and cysteines are alkylated with IAA at room temperature to form carbamidomethylated derivatives. Because IAA is unstable in light, it must be prepared immediately before alkylation of reduced proteins and protease digestion for MS analysis. However, in some cases the overalkylation with IAA may modify lysine, histidine, and N-terminal residues (Boja and Fales 2001). Therefore, to avoid these side effects caused by IAA, some researchers suggest alternate approaches, such as the use of 4-vinylpyridine to alkylate cysteine sulfhydryl groups of proteins after previous reduction with tris(2-carboxyethyl)phosphine (Erde et al. 2014). Generally, the concentrations of reagents are selected in consideration of the enzyme optimal activity and overalkylation side effects, and they are removed before MS analysis by ultrafiltration, solid-phase extraction, or in-line RP chromatography.

Protein Digestion Conditions

Trypsin, a work horse in bottom-up proteomics, is the protease of choice as it has a high specificity and is stable under a wide range of conditions, including 40 % acetonitrile and 2 M urea. However, its cleavage sites are not always predictable due to frequent miscleavages caused by skipping a cleavable residue (Lys or Arg) when

the successive Lys/Arg are present, or due to low trypsin digestion efficiency when these residues are followed by Pro. Miscleavages may also occur due to incomplete protein denaturation or post-translational modifications (PTMs) on amino acid residues near protease cleavage sites. In addition, auto-proteolysis can generate pseudotrypsin exhibiting chymotrypsin-like specificity. Hence, the modified trypsin, for example, through dimethylation of lysine residues, is commonly used which has better cleavage specificity and maintains optimal activity at higher temperatures. However, trypsin preparations usually contain some contaminating chymotrypsin; therefore, commercial products known as "sequencing grade" are treated with N-tosyl-phenylalanyl chloromethyl ketone (TPCK) to inhibit chymotrypsin activity. Nevertheless, to avoid auto-digestion, trypsin is used at low concentrations and the reaction is typically carried out at 37ºC for a few hours or even overnight before termination. Therefore, many approaches have been developed focusing on increasing the speed, yield, and robustness of the digestion process through optimization of reaction conditions, immobilization of trypsin on solid supports, or by addition of other proteolytic enzymes. For example, Glatter et al. (2012) found superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin alone to yield fully cleaved peptides while reducing the abundance of miscleaved peptides. The overview of the available techniques and digestion methods for shotgun-proteomics applications can be found in recent literature (e.g., Switzar et al. 2013a; Vuckovic et al. 2013).

Various reagents have been reported as enhancers used to accelerate protease digestion, as well as to improve the digestion efficiency for membrane proteins. For example, Masuda et al. (2008) compared 27 enhancers, including surfactants, organic solvents, and chaotropic agents, and examined their influence on the protease activity of trypsin and protease Lys-C as well as on the solubility of membrane proteins. They found that bile salts, like sodium deoxycholate even at 0.01% concentration, increased trypsin activity more than fivefold; hence they developed a new protocol based on the use of this surfactant for protein extraction, solubilization, and trypsin activation. Their protocol, which included extraction of cholic acid from the acidified sample with ethyl acetate (phase transfer) before LC-MS analysis, improved substantially the efficiency of protein identification for membrane-enriched fractions of *E. coli* (Masuda et al. 2008).

To shorten the digestion time of *E. coli* protein extracts down to 15 min, Masuda et al. (2009) used immobilized trypsin in a spin-column format. Moreover, they increased the digestion efficiency even further by the presence of sample solubilizers, that is, lauroylsarcosine and deoxycholate that act as a natural trypsin activity-enhancing agent present in bile acids secreted into a small intestine. Overall, by using this approach they identified 1453 proteins, including 545 membranes proteins.

Recently, Erde et al. (2014) used 0.2% deoxycholic acid to enhance trypsin performance during the FASP digestion of a whole cell protein extract from *E. coli* cells and showed that this modified protocol, referred to as enhanced or "eFASP," increased tryptic digestion efficiency for both cytosolic and membrane proteins.

Modification of protein digestions using physical methods has also contributed to improved digestion efficiency and proteomic coverage. Covalent and dynamic immobilization of trypsin on micro- and nanoparticles, the use of pressure cycling

technology, high-intensity ultrasound, and the microwave heating have improved the kinetics of tryptic digestion by reducing digestion time and enhancing the cleavage specificity, especially for hydrophobic and membrane proteins (Vaezzadeh et al. 2010).

Sample Digestion Strategies

The currently available digestion strategies and recent developments in the acceleration of the digestion process allowing for reduction of the digestion time from hours to minutes or even seconds have been reviewed by Switzar et al. (2013a).

In recent years, the FASP method (Manza et al. 2005; Jabbour et al. 2007, 2010c; Wiśniewski et al. 2009) has emerged as a key tool for processing microbial protein extracts for strain identification (Jabbour et al. 2010a, b, 2014; Wade et al. 2010, 2011). It enables the integration of all sample processing steps required for efficient on-filter enzymatic cleavage of proteins and removal of contaminants by using a filtration unit as a "one-pot" proteomics reactor, thereby reducing the risk of sample loss. Nevertheless, these commercially available units should be passivated before the use to avoid peptide losses from low copy number proteins. For example, Erde et al. (2014) found that overnight incubation of both filter units and collection tubes in the passivation solution of a nonionic surfactant Tween-20 increased dramatically the peptide recovery from small samples (up to 300%).

Although diverse types of ultrafiltration devices are used, generally, the 30 kDa units are best suited for FASP because they retain small proteins ($M_r < 10$ kDa)—due to the large Stokes radii of proteins unfolded in urea that prevents them from passing the filter—and pass more larger peptides (with $M_r > 1500$ Da) than the 10 kDa filters. In addition, the centrifugation time needed to concentrate samples is 3–4 times shorter than that with the 10 kDa units (Wiśniewski et al. 2011).

To increase the proteome coverage, Wiśniewski and Mann (2012) suggested a consecutive sample digestion procedure carried out in a filtration unit proteomic reactor and developed a protocol, enabling consecutive digestion of the sample with two or more enzymes, referred to as multienzyme digestion (MED)-FASP. In this "extended" FASP method, peptides are liberated by centrifugation after each digestion step and the remaining material is subsequently cleaved with the next proteinase. Therefore, orthogonal populations of peptides are created from the same sample that can be jointly or separately analyzed using LC-MS/MS to increase substantially the number of identified peptides in comparison to the single enzyme digestion protocol applied to the same amount of sample. For example, they found that consecutive use of endoproteinase Lys-C and trypsin enabled in some cases to double the number of identified unique peptides (Wiśniewski and Mann 2012). The application of MED-FASP to analysis of *E. coli* ATCC 25922 strain whole cell lysates—by using digestion with endoproteinase LysC, followed by filter washes and trypsin digestion—allowed the identification of 8206 ± 270 unique peptides in the LysC fraction, and 10,728 ± 319 tryptic peptides per sample (Wiśniewski and Rakus 2014).

The FASP protocol for shotgun proteomics of whole cell lysates was also extended to a high-throughput sample preparation procedure based on simultaneous processing of samples in 96-well filter plates (Switzar et al. 2013b). Their protocol enabled all sample preparation steps, including cell lysis, buffer exchange, protein denaturation, reduction, alkylation, and proteolytic digestion to be carried out for a large number of samples. The protocol would be suitable for diagnostic analysis, for example, in a clinical laboratory or for processing large numbers of fractions resulting from prefractionation of microbial proteomes in a research lab. They pointed out that the usage of a single plate for all sample preparation steps following cell lysis reduces potential samples losses, increases sensitivity, and allows for automation.

Yu et al. (2012) combined FASP, used for an efficient depletion of detergents, with the ultrafast and efficient microwave-assisted on-filter enzymatic digestion by transferring proteins mixed with trypsin on filter units to a microwave oven where they were digested for less than 1 min. Also, Chang et al. (2013) used the FASP method for processing the *Acinetobacter baumannii* whole cell protein extract, followed by a 15-min-long microwave-assisted protein digestion with trypsin.

However, according to Reddy et al. (2013) the faster reaction rate is not caused by the microwave quantum effect but the thermal one. Therefore, both microwave and conventional heating at high temperatures (50 °C) can be used to accelerate digestion reactions. For example, Tracz et al. (2013) trypsin-digested whole cell protein extracts from pathogenic strains of *Yersinia, Francisella,* and *Bacillus* at 53 °C for a couple of hours and used thousands of released and confidently identified peptides for successful bacterial identifications.

Liquid Phase Separation and Ionization of Peptides Followed by Acquisition of Tandem Mass Spectra

In classical bottom-up methods, separated proteins are in-gel trypsinized, and the released peptides are identified by mass mapping or by analyzing product ion mass spectra obtained through the collision-induced dissociation or postsource decay (Chalmers and Gaskel 2000). In the shotgun approach, peptides are released during proteome-wide digestion of microbial proteins with proteolytic enzymes and in some applications they are directly analyzed using MALDI time-of-flight (TOF) MS for peptides mass fingerprinting (PMF) of microbes, or peptides from dominating proteins are sequenced using MALDI-MS/MS technologies. For example, Warscheid and Fenselau (2003) investigated the PMF concept for analysis of small acid-soluble proteins in *Bacillus* species by on-probe shotgun trypsin digestion of spores from this genus. The released peptides were also identified by tandem-MS techniques for distinguishing *B. cereus, B. thuringiensis, B. subtilis, B. globigii,* and *B. anthracis* Sterne strains. More recently, Balážová et al. (2014) demonstrated that microwave-accelerated shotgun tryptic digestion of cellular material combined with MALDI-TOF MS profiling of released peptides allowed for subspecies differentiation of *Staphylococcus* and *Bacillus* strains. However, substantial improvements in

the scope of sequence coverage and reliability can be achieved through separation of peptides by LC or capillary electrophoresis (CE) prior to ESI-MS/MS analysis (Wolters et al. 2001).

Liquid Chromatography-ESI-MS

Several strategies have been developed to fractionate peptides prior to MS analysis that include separation based on one-dimensional (1D) nano-LC and multidimensional separation systems. In the former approach, the resolution of peptide separation can be increased through the use of RP columns with smaller particle sizes, for example, below 2 μm in diameter, and submicroliter flow rates (Fröhlich and Arnold 2009). This technique gives higher efficiency but requires higher pressure separation, and is therefore referred to as ultrahigh-pressure liquid chromatography (UPLC). However, nanospray is more sensitive than approaches using higher flow rates because electrospray is a concentration-sensitive process. Consequently, the use of a narrower column and lower flow rates will cause the elution of peptides as narrower peaks with higher maximal concentrations. In addition, the use of longer columns operated at higher temperatures may increase both high-resolution and high-peak capacity separations even further.

For example, Hebert et al. (2014) identified more than 34,000 peptides with unique sequences over a 70-min run by using a 35-cm long RP column with 75 μm internal diameter. This column was packed with 1.7 μm C18 particles and was operated at 60 °C by using the mobile phase containing 5% of dimethyl sulfoxide, in addition to the standard components, that is, formic acid/water and formic acid/acetonitrile. Eluting peptide cations were electrospray ionized and analyzed on a hybrid mass spectrometer (quadrupole-orbitrap-quadrupole-ion trap, Q-OT-qIT; Orbitrap Fusion, Thermo Scientific, San Jose, CA, USA).

The most popular multidimensional separation systems use: (i) a combination of peptide separation according to their isoelectric point by isoelectric focusing on immobilized pH gradient, followed by RP LC separation according to their hydrophobic properties and MS/MS analysis (Vaezzadeh et al. 2010; Geiser et al. 2011b), (ii) off-gel electrophoresis and RP LC-MS/MS (Geiser et al. 2011a), or (iii) multidimensional liquid chromatography (MDLC). In the latter group of methods, the most commonly used is the multidimensional protein identification technology, termed Mud-PIT, which was introduced by Washburn et al. (2001). Mud-PIT consists of two orthogonal separation systems—strong cation exchange (SCX) and RP—coupled online in an automated fashion and offering the possibility to analyze highly complex peptide mixtures in a single experiment. Most commonly, and as originally published, an RP-precolumn is followed by an SCX-precolumn, and finally the main RP-separation column; thus forming a triphasic column packed into an ESI-emitter tip directly coupled to a mass spectrometer; however, there are many variations of this basic format (Lohrig and Wolters 2009). Recently, a detailed protocol has been described for the construction of a simple and flexible online

RP-SCX-RP LC system and its implementation for deep proteome profiling on a common shotgun-proteomics platform (Lam et al. 2014).

In addition to the online MDLC format, offline approaches are quite popular and each of them has its advantages and disadvantages, that is, reduced labor time in case of online separation and the flexibility of offline fraction collection. However, online methods are not optimal for peptide separation due to the elution of peptides with a solvent step gradient during ion-exchange chromatography. Therefore, offline techniques based on a continuous gradient ion-exchange separation of peptides, which are subsequently analyzed by RP LC coupled with ESI-MS/MS, represent a better choice for the comprehensive analysis of the bacterial proteome. By using this approach, Jaffe et al. (2004) found almost 10,000 unique tryptic peptides corresponding to 81 % of the predicted ORFs for a small, wall-less bacterium *Mycoplasma pneumoniae*.

Although very high proteome coverage can be achieved, in the past it usually required a long data acquisition time. For example, Hendrickson et al. (2010) reported detection of 1671 proteins representing 64 % of all genome predicted proteins of *Methylobacillus flagellatus*. However, they achieved it by analyzing five prefractions, resolved by using 2D capillary high-performance LC (HPLC) analysis that consisted of a seven-part step gradient from the cation-exchange portion of the biphasic column, followed by the reverse phase elution and MS analysis. This gave a total of 35 separate HPLC runs per technical replicate with 60 min effective acquisition time per run.

Capillary Electrophoresis-ESI-MS

ESI-MS/MS allows for online detection and identification of peptides separated by CE (Janini et al. 2003). This approach is rarely used for microbial identification purposes; however, Hu et al. (2005 and 2006) described a successful application of this technique for identification of microbial mixtures using a quadrupole ion trap operated in a selective tandem-MS mode. They trypsin-digested bacterial proteins and analyzed released peptides with CE-MS/MS by targeting species-unique tryptic peptide ions. For that purpose they first created a small DB of proteotypic tryptic peptides derived from abundant proteins that are species-specific biomarkers for targeted strains. Isolated ions of such peptides were analyzed by using a selective reaction monitoring approach. The overall identification success for this method was 97 % on the basis of analysis of 34 clinical samples with a total analysis time of 8 h that included a 6-h long cultivation step. Moreover, they shortened the time-consuming digestion process to 15 min by the application of microwave-assisted proteolysis (Lin et al. 2005).

Recently, Fleurbaaij et al. (2014) developed a CE-ESI-MS/MS bottom-up proteomics workflow for sensitive and specific peptide analysis with the emphasis on the identification of β-lactamases in various Gram-negative bacterial species even

from single colonies. They demonstrated the ability of the system to successfully assess multidrug-resistant bacterial clinical isolates.

Liquid Chromatography MALDI-MS/MS

The separation of complex peptide mixtures using LC columns is usually coupled to mass spectrometric analysis by electrospraying column effluent directly into the mass spectrometer. However, peptides separated by nanoscale LC may be coupled to a collector that deposits microfractions onto a MALDI plate, thus allowing for the MALDI-MS/MS analysis of the fractions by instruments with TOF/TOF ion optics or/and LTQ-Orbitraps (Yang et al. 2007; Baeumlisberger et al. 2011). For example, Lasaosa et al. (2009) found that the MALDI-based platform led to a significantly increased number of peptides identified from a tryptic digest of the cytosolic proteome of the bacterium *Corynebacterium glutamicum;* probably due to the fact that the size of the unique peptides identified by MALDI was, on average, 25 % larger and more hydrophilic than the unique peptides identified by ESI (Yang et al. 2007).

Generally, there are several benefits associated with the LC-MALDI-MS/MS approach. First, the collection of MS/MS data is decoupled from the chromatographic separation, so the sample can be reanalyzed using optimized MS/MS parameters. Second, the relative insensitivity to interfering compounds in the sample matrix and/or mobile phases allows carrying the chromatography under optimized conditions. Third, this approach provides the ability to archive the sample plate (Fernández-Puente et al. 2014).

In conclusion, nano-LC combined with further improvements in MS sensitivity and speed will continue to reduce whole proteome analysis time for microbial strains by producing tens of thousands of peptide sequence-to-spectrum matches (PSMs) in less than 1 h (Hebert et al. 2014). However, LC-MALDI-MS/MS analyses may be better suited for specific applications requiring sample archiving.

Database (DB) Construction and Searching

The prevailing approach for peptide, protein, and microbial strain identification in shotgun proteomics is based on decoding amino acid sequences by using combined information of the tryptic peptide mass and its fragmentation spectrum matched against DB sequences. Therefore, the success of identifying any ionizable peptide depends on the availability of suitable DB reference sequences, and by no means can it be assumed that sets of reference genomes/proteomes available in the public DBs are complete or fully representative for any isolated strain. Therefore, the use of an appropriate DB is crucial for subspecies typing and identification of strains.

Bacterial DBs

The construction of protein sequence DB plays a crucial role in proteomic workflows; however, the DB should contain all possible sequences while on the other hand, if the DB is too large, the search engine may introduce false positive identifications (Vaudel et al. 2014).

There are almost 15,000 bacterial strains with sequenced genomes, including 4000 with complete genome sequences, available in public DBs, as of fall of 2014, and chromosome and plasmid-encoded protein sequences predicted from these genomes can be downloaded from the National Institutes of Health (NIH) National Center for Biotechnology Information (NCBI, ftp://ftp.ncbi.nih.gov/genomes/Bacteria) or from the Universal Protein Resource (UniProt) Knowledgebase (UniProtKB; www.uniprot.org). However, microbial proteomes in these DBs vary greatly in terms of their curation, completeness, and comprehensiveness; hence, the use of most recent versions translated from complete genome sequencing projects is strongly recommended. Amino acid sequences in these DBs represent a translation of nucleotide sequences in computationally determined ORFs that potentially encode proteins. ORF begins with an initiation codon and ends with a stop codon and has the potential to encode a single polypeptide expressed as a protein; however, many may not actually do so. In addition, different bioinformatics approaches for automatic annotation of genes are currently used and this affects the quality of protein lists used in proteomics. For example, different annotation tools may predict different translational start sites (TSS) for ORFs that will affect the N-terminal peptides generated during in silico digestion (de Souza et al. 2010 and 2011; Armengaud et al. 2013). Furthermore, a protein should be understood as one of many isoforms representing the expressed gene and may differ from a polypeptide specified by a nucleotide sequence due to co-translational modifications or PTMs of a nascent polypeptide. Co-translational modification refers to the removal of N-terminal methionine by N-methionyl aminopeptidase and affects the majority of bacterial proteins. PTMs comprise both the proteolytic processing of a polypeptide, for example, to generate appropriate targeting signals, and covalent modifications of its amino acids (Hesketh et al. 2002; Bonissone et al. 2013; Zhang et al. 2013b). Therefore, the available DB searching algorithms, in fact, identify ORFs, not proteins. Moreover, during analysis of an unknown microbial strain the confirmation of the full amino acid sequence or "100% coverage" of a potential protein would be required for the identification of an ORF, because sequences of orthologous proteins from a closely related strain may only differ due to an SAV. Consequently, the true identification of proteins is rarely achievable during high-throughput analyses of microbial proteomes.

In the early studies on identification of bacteria using shotgun proteomics, Dworzanski et al. (2004) constructed a prototype proteome DB from genome sequences downloaded from the NCBI site. They used a computational Gene Locator and Interpolated Markov Modeler (Glimmer) developed by Salzberg et al. (1998) to identify protein-coding ORFs and translated them into amino acid sequences

of all putative proteins. All these sequences were used for assembling a microbial proteome DB in a FASTA format.

A sequence in FASTA format begins with a single-line description distinguished from the sequence data by a greater-than (“>”) symbol and ends with a carriage return. Although the description is generally considered as a free form, software applications such as search engines assume that the first word or string after the “>” symbol is a real sequence identifier and use it for processing while the remainder of the line is a supplementary description. Therefore, Dworzanski et al. (2004) modified header lines of each protein in a DB, by using a header replacer script written in Perl, and added abbreviated strain names in header lines, so the search engine was recognizing and assigning PSMs directly to reference DB strains instead to particular proteins in each proteome. Consequently, the search engine SEQUEST was recognizing each proteome as a single “pseudo-polyprotein” and could be used for ranking all peptide-to-strain matches while retaining complete information about protein sources with each peptide. Although the above DB could be searched directly, the search efficiency may be substantially improved by in silico digestion of all sequences to create an indexed peptide sequence DB derived from all DB proteomes.

Recently, Tracz et al. (2013) described a similar approach by tricking Mascot to assign PSMs directly to reference DB strains instead of proteins. They achieved it by creating a custom database, named “Genome AA,” containing protein sequences deduced from 2,026 completed bacterial genomes available from the NCBI Reference Sequence (RefSeq) DB. However, each entry in the GenomeAA DB consisted of the strain name followed by a “pseudo-polyprotein” created by concatenation of all individual protein sequences separated only with the letter code J. Therefore, to preserve the integrity of peptide termini, trypsin digestion rules used by the search engine were always supplemented with information to cleave on the C- and N-terminal sides of the letter code “J.” Consequently, Mascot searches against this DB report PSMs to reference strains represented by DB proteomes, instead of particular proteins.

In proteogenomic studies, six-frame translated nucleotide sequences from investigated genomes are used (Armengaud et al. 2013). However, DBs used for strain identification are usually downloaded from NCBI or UniProKB as FASTA formatted protein sequences. Nevertheless, they may be additionally cured. For example, Dworzanski et al. (2006, 2010), Jabbour et al. (2010a, b, c,) and Deshpande et al. (2011) continued to create prototype microbial DBs by adding abbreviated strain names to header lines for each downloaded protein, as described above. These abbreviated strain names were also used as specific codes that linked strains to taxonomic information derived from the NCBI taxonomy DB (http://www.ncbi.nlm.nih.gov/Taxonomy/). Finally, they indexed the DB by performing in silico digestion of proteins using a TurboSEQUEST utility program (Thermo Scientific) by assuming (trypsin) endoprotease digestion rules and allowing up to two missed cleavages per peptide; however, only peptides with M_r in the 700−3500 Da range were accepted.

It is also important to append any protein DB with sequences of common laboratory contaminants. For instance, the following FASTA formatted DBs of

contaminants are available via the Internet: (i) the common Repository of Adventitious Proteins, cRAP, can be downloaded from the Global Proteome Machine FTP site (ftp://ftp.thegpm.org/fasta/cRAP) or (ii) a contaminants.fasta file containing common contaminants is available at http://maxquant.org/downloads.htm and could be appended to any target DB and used as a control for environmental and common laboratory contaminants.

DBs of Virulence Factors, Toxins, and Antibiotic Resistance Determinants

In addition to the identification of bacteria, it is also helpful to subtype isolated strains in regard to their functional capabilities such as virulence, antibiotic resistance, or production of toxins which are of high epidemiological, clinical, and agricultural or biosecurity importance. However, the search engines usually disregard this type of information contained in well-annotated DBs or it is difficult to retrieve it in an easy-to-interpret format. Therefore, based on inputs from publicly available sequences, it is advantageous to create customized DBs that are configured to facilitate subtyping of strains based on the presence of sequences associated with specific factors, for example, responsible for virulence or antibiotic resistance. Although some researchers prefer to create their own DBs customized for specific needs, there are also a few well-annotated sequence DBs targeting virulence and antibiotic resistance proteins which are available for downloading via the Internet (Chen et al. 2012; Winnenburg et al. 2008; Gupta et al. 2014).

Virulence factors (VFs) help pathogens to evade host-specific defensive mechanisms to establish infection. They include bacterial toxins, secreted effectors, for example, hydrolytic enzymes that may contribute to the pathogenicity of the bacterium, cell surface proteins that mediate bacterial attachment, and cell surface carbohydrates and proteins that protect a bacterium, among others. There are a few DBs available with protein sequences of such VFs. For example, a DB of protein VFs (VFDB) in the FASTA format was compiled based on information from more than 2000 related publications and can be downloaded from the http://www.mgc.ac.cn/VFs/main.htm Website (Chen et al. 2012). It contains sequences of 460 VFs, 24 pathogenicity islands and ca. 2500 VF-related proteins (as of November, 2014) gathered from 429 chromosomes and 93 plasmids of pathogenic bacterial strains belonging to 26 bacterial genera. The other DB of VFs, named "Victors Virulence Factors" DB currently includes 5173 VFs from strains of 125 microbial species known as pathogenic to humans and animals (50 bacterial species, 54 viruses, 13 parasites, and 8 fungi). A FASTA file with protein sequences of all these VFs is available for download from the http://www.phidias.us/victors/download.php website. The data within Victors are manually curated and comes from peer-reviewed literature and existing DBs (e.g., NCBI RefSeq). The other DB with protein sequences available

for download is known as a pathogen–host interaction DB (PHI-base) (Winnenburg et al. 2008) that contains curated information on genes proven to affect the outcome of PHIs. It catalogs experimentally verified pathogenicity, virulence, and effector genes from fungal, fungus-like eukaryotic microorganisms (*Oomycete*), and bacterial pathogens infecting animal, plant, and insect hosts. PHI-base is therefore an invaluable resource in the discovery of these genes in medically and agronomically important pathogens. PHI-base contains 3012 entries with protein sequences translated from so-called pathogenicity genes (if the effect on the phenotype is qualitative) or virulence/aggressiveness genes (if the effect is quantitative) or effector genes (either activate or suppress plant defense responses) and can be downloaded in the FASTA format at http://www.phi-base.org/.

Antibiotic resistance (AR) Gene-ANNOTation (ARG-ANNOT) DB was developed by Gupta et al. (2014) and consists of a single file with amino acid sequences of existing and putative antibiotic resistance-associated proteins in a FASTA format that can be downloaded from http://www.mediterranee-infection.com/article.php?laref=282titre=arg-annot. They collected information about 1689 AR-associated genes from published works and online resources, and sequences of these gene products were retrieved from the NCBI GenBank DB. AR-associated proteins in ARG-ANNOT DB are linked to diverse antibiotics classes, including aminoglycosides, beta-lactamases, fosfomycin, fluoroquinolones, glycopeptides, macrolide-lincosamide-streptogramin, phenicols, rifampicin, sulfonamides, tetracyclines, and trimethoprim. There are also other available DBs like the Antibiotic Resistance Genes DB (ARDB, Liu and Pop 2009) or MvirDB—a microbial DB of protein toxins, virulence factors, and AR genes for bio-defense applications—that integrates DNA and protein sequence information from other sources (Zhou et al. 2007), however, they were not recently updated.

A number of web-services are available for identification of known or predicted bacterial toxins, for example, BTXpred, which makes available a FASTA formatted file of 185 bacterial toxins (http://www.imtech.res.in/raghava/btxpred/supplementary.html); and a DB of Bacterial ExoToxins for Human (DBETH, http://www.hpppi.iicb.res.in/btox/) with FASTA files of "Human Pathogenic Bacterial Exotoxin Fasta Sequences" and "Human Pathogenic Bacterial Exotoxin Homologs."

Chang et al. (2013) created the β-lactam-resistance protein DB of *A. baumannii* (abbreviated as "BRPDAB") and used it to develop an accurate and rapid shotgun-proteomics method for the identification of β-lactam-resistant *A. baumannii* pathogens. They used a serious of gene ontology (go) terms (Ashburner et al. 2000) such as beta-lactamase activity (go:0008800), penicillin binding (go:008658) or response to antibiotics (go:0046677 used as a synonym to antibiotic susceptibility/resistance), names of all β-lactam antibiotics and the name of a bacterium "*A. baumannii*" to identify in the Uniprot DBs proteins associated with the resistance of this pathogens to antibiotics. They downloaded these sequences and incorporated them into the FASTA formatted BRPDAB.

Creation/Correction of Microbial Protein DBs Through Re-sequencing and Analysis of Genomes

Despite the availability of thousands of completely sequenced genomes, many species are still represented in sequence DBs by only a single or a few strains. Therefore, for analysis of strains from DB underrepresented species, improved DBs are needed to compensate for the missing sequence variations reflecting intraspecies strain diversity that may affect the identification of organisms at the subspecies level. Such DBs could be constructed by the "maturation" of sequences, for example, by N-methionine excision, removal of N-terminal signal peptides based on annotations in the DB, or cleavage site predictions determined with the help of suitable algorithms, such as SignalP (Petersen et al. 2011) or Phobius (Käll et al. 2007b). Additional DB improvements can be achieved by correcting some sequencing errors such as incorrect predictions of TSS during an in silico-driven annotation process to make the N termini of homologs as consistent as possible within the DBs (Sato and Tajima 2012).

For example, the identification of protein variants could be improved by using the multistrain MS prokaryotic DB builder (MSMSpdbb) (de Souza et al. 2010). In this approach, a combined protein DB of closely related microorganisms is created that provides two important advantages. First, it allows for streamlining the initial DB searching by combining groups of phylogenetically close organisms, and second, it provides protein annotation improvements by correcting sequence TSS which are frequently incorrectly annotated, especially for older submissions. Recently, Bland et al. (2014) characterized 534 N termini of the marine bacterium *Roseobacter denitrificans* and found that 10% of them were incorrectly annotated in regard to TSS. They also found five previously un-annotated proteins and eight proteins with multiple translational starts, thus showing the value of empirical evaluation of every sequenced organism for maximum annotation accuracy (Bland et al. 2014).

However, the most reliable solution to overcome the problem of relatively large sequence deviation of an unknown isolate from reference strains should be based on de novo sequencing on protein or nucleic acid levels or ultimately by performing whole-genome sequencing of additional strains from the underrepresented species. Unfortunately, de novo peptide sequencing is still impractical; therefore, both mRNA (RNA-seq, Wang et al. 2012) and genomic DNA sequencing have been used to generate customized DBs for MS identifications in proteomics studies. Because the RNA-seq approach is more appropriate for metaproteomic approaches, DNA sequencing of strains from species underrepresented in public DBs is better suited for expanding the potential sequence variation repertoire for high-resolution discrimination of unknown strains. Furthermore, the Food and Drug administration (FDA) authorization for the first next-generation sequencer, Illumina's MiSeqDx (Collins and Hamburg 2013), will allow not only the development of new human genome-based tests but will also open the way to high-throughput sequencing (HTS) being used in clinical microbiology.

There are two major approaches that have been used: de novo assembly from raw sequence reads and the reference-guided assembly if the closest reference genome is available. However, HTS technologies are error prone; for instance, the Illumina reversible dye-terminator sequencing technology (HiSeq) caused substitutions (Meacham et al. 2011) while ion semiconductor sequencing technology (Ion Torrent, Life Technologies) produced indel errors associated with homopolymer regions (Loman et al. 2012). In addition, despite many computational advances, the complete and accurate genome assembly from second-generation short-read data remains a major challenge. Therefore, instead of de novo genome assembly, the better strategy for proteomics would be re-sequencing based on mapping reads to the whole genome sequence of a strain from the same species followed by searches for single nucleotide variations (SNVs) (Caboche et al. 2014).

Recently, Wu et al. (2014) described a very efficient strategy to overcome sequence variations between the reference genome and the closely related species based on mapping sequencing reads utilizing the error-tolerant FANSe mapping algorithm (Zhang et al. 2012). FANSe corrects the SNVs for the genome deviating from the reference genome ~5 %, and exports them as corrected proteome sequences that can be used in searching peptide fragmentation spectra, and thus efficiently improves peptide and protein identification in nonmodel bacteria without complete genomic sequence. FANSe is a seed-based algorithm which uses the entire information from a sequencing read divided into small seeds of 6–8 nucleotides and aligns all of them to the reference genome sequence. The adjacent seeds mapped to the same segment are combined if they fulfill certain criteria and are used to define so-called hotspots. The alignment for each hotspot is scored and refined based on the least number of mismatches. Consequently, by reducing the number of hotspots FANSe achieves the increased sensitivity, that is, the proportion of actual positives which are correctly identified as such while maintaining a reasonable speed.

For example, sequencing of 1350 bp of 16S rDNA of an environmental isolate, Wu et al. (2014) found 100 % identity to the reference sequence of *Bacillus pumilus* SAFR-032; the only *B. pumilus* strain with complete genome sequence in public DBs. However, 16S rDNA sequence may not distinguish separate strains; therefore, they decided to re-sequence the whole genome and identified 158,407 SNVs. Among these SNVs, 143,263 were identified as substitutions, 221 insertions, and 349 as deletions in protein-coding sequences (CDS). In total, 4.93 % of the mappable region was different in comparison to the reference genome of *B. pumilus* SAFR-032, that is, in the expected range of differences between the same species strains (Goris et al. 2007). This correction allowed them to identify 14.2 % more tryptic peptides from the isolate and they will use this corrected proteome as a reference for the identification and discrimination of other strains from *B. pumilus*. In conclusion, this approach is suitable for the preparation of a set of reference proteomes for DB searching of MS/MS fragment ions derived from the unknown strain proteome for subspecies identification and strain typing.

Finally, it should be remembered that during proteomic analyses, only a fraction of genome predicted proteins and proteotypic peptides are identified and there are a number of reasons why this happens. First, peptides from undetected proteins

may fall into a category of false negatives due to bioanalytical factors inherent to bottom-up proteomics: post-translationally modified peptides, peptides too short, too long or from small and low-abundance proteins are difficult to observe. Second, some predicted proteins are not real, due to incorrect genome interpretation including annotations marked as putative or hypothetical. For example, Hendrickson et al. (2010) noted that many of the nondetectable proteins of *M. flagellatus* may represent artifacts of genome annotation while a portion of the nonexpressed proteins appear to correspond to silent genomic islands. Third, some proteins must be true negatives, that is, they are not expressed under the growth conditions used because the expression of many genes is tightly regulated and/or inducible only under specific conditions. For example, Ansong et al. (2009) showed that as much as a third of the *Salmonella enterica* serovar Typhimurium strain's proteome has to be regulated at the translational level by the single virulence regulator Hfq. Nevertheless, nearly 40 % of predicted proteome was covered by peptide identifications in this work.

Custom DBs of E. coli and Salmonella Flagellins

E. coli bacteria are short rods with flagella that rotate to allow movement in liquid environments. The flagellar filament is the largest portion of the flagellum and consists of repeating subunits of the protein flagellin that induces immune responses. These immune responses have been widely utilized for serological typing of *E. coli* strains, which produce 53 distinct sequence types of flagellar H antigens. Recently, Cheng et al. (2013) developed an MS-based typing method of flagellar H antigens (MS-H). For this purpose, they constructed a FASTA-formatted DB of *E. coli* H types using the sequences and serotype information found in the NCBI nr protein DB. In this *E. coli* flagellin DB redundant sequences were collapsed into a single entry with the H-type listed in each sequence description headerline. If the H-type was not specified in the NCBI nr DB, they compared it against sequences with known H serotypes and assigned the top-scoring one. In some cases, the H-type was manually assigned (based on literature search) to sequences with missing H-type in NCBI annotation, or with incorrect H-type listed in the NCBI entry. Incorrect H-types were also discovered by finding outliers in a phylogenetic analysis of all *E. coli* flagellin sequences in the DB.

The final curated *E. coli* flagellin DB can be downloaded at http://www.biomedcentral.com/1756-0500/7/444 as a FASTA file (KC_Flagellin_20130425.fasta). This DB contains 195 unique sequence entries representing all 53 known *E. coli* H serotypes, that is, averaging close to 4 sequences per serotype. However, some serotypes were represented by only one entry (H4; H15; H23; H24; H30; H32; H39; HH43; H51; and H56) while the most common types, such as H6, H11, and H7 were represented by 10, 12, and 16 flagellin sequences, respectively (Cheng et al. 2014a).

For typing flagellin H-antigens of multiphasic *Salmonella* reference strains, Cheng et al. (2014b) created also a curated *Salmonella* flagellum DB containing 385 entries of flagellin sequences available in the literature and the NCBI nr DB. However, this DB is not available for downloading.

Search Engines

Mass spectrometric analysis of peptides released by shotgun digestion of microbial proteins generates high-resolution and high accuracy data sets of product ion spectra that can be used for decoding their amino acid sequences by three classes of approaches. First, by spectral library searches which compare the acquired spectra with a library of previously identified spectra; second, by de novo sequencing to infer the sequence directly from the mass differences of fragment ions in the spectra; and third, by DB searches which compare how well an acquired spectrum matches to a theoretical spectrum of a peptide deduced from protein sequence in the DB (Cottrell 2011; Ma and Johnson 2012). In the latter case, the search engine constructs a theoretical spectrum for each candidate peptide sequence and compares them to experimentally observed fragment ion spectra.

Although de novo approaches would be the best for decoding sequence information from peptide fragmentation spectra, they show sufficient reliability to infer only short sequence tags and thus currently cannot provide a full solution to the identification problem. Nevertheless, they are used in so-called error-tolerant searches that relax the specificity, for instance, by removing molecular mass constraint and thus allowing for matches to DB sequences when there are sequence variations due to mutations or PTMs.

Therefore, the most popular approach to interpret such MS/MS spectra in a high-throughput manner uses DB searches with software tools known as "search engines" to find the best PSMs. As input, a search engine takes MS/MS spectra and searches them against reference proteomes of strains that are expected to be related to the sample with a twofold purpose: first, to find PSMs which confidently decode tandem mass spectra; and second, to quantify the contribution of DB reference microbes to the decoded spectral data set.

There are many well-established software applications for searches with uninterpreted fragmentation spectra against DB proteomes that include SEQUEST (Eng et al. 1994), Mascot (Perkins et al. 1999), X!Tandem (Craig and Beavis 2004), MyriMatch (Tabb et al. 2007), OMSSA (Geer et al. 2004), and Andromeda (Cox et al. 2011) among many others listed in the review article by Nesvizhskii (2010). In addition, due to different approaches used in search engine algorithms, one can maximize the number of peptide identifications by using multiple search engines and combining the results. For example, the PSM gains (at 1 % error level) observed by starting with Mascot and adding SEQUEST search results may exceed 38 %, and by adding MyriMatch and X!Tandem to the combination, the gain can reach 53 % (Shteynberg et al. 2013). These outcomes were obtained by modeling each

search result with PeptideProphet (Keller et al. 2002) and combining them with the iProphet tool (Shteynberg et al. 2011) that uses linear discriminant analysis to obtain more accurate PSM scores. Nevertheless, to maximize the number of correct PSMs it is important to run DB search engines using appropriate search parameters.

Setting Search Parameters

The search parameters include, among others, ions' mass tolerances appropriate for the type of instrument used and expected peptide modifications. For example, 5-ppm precursor mass tolerances for a high-resolution mass spectrometer and 0.5-Da fragment tolerance for the ion trap fragmentation. The expected peptide modifications include: (i) static (fixed) that apply to all amino acid residues in a sample, for example, cysteine modification due to the alkylation step, and (ii) dynamic (variable) which may or may not be present at each amino acid site.

Most variable modifications of amino acids are dependent on the sample processing and may include the oxidation of methionine and tryptophan; deamidation of asparagine and glutamine to their acidic counterparts, aspartate and glutamate (Yang and Zubarev 2010); carbamylation of free amino groups; and diverse modifications of N-terminal amino group. For example, the common artifact of using gel electrophoresis during the sample preparation is formation of cysteine propionamide (C[+71]). Furthermore, cysteine residues are usually carbamidomethylated (C[+57)]) by treatment with IAA to block free sulfhydryl groups. In addition, overalkylation with IAA frequently also gives modified lysine [K+(57)], histidine [H+(57)], and N-terminal residues (+57) and (+114) and may affect even substantial fraction of peptides (Boja and Fales 2001).

Commonly used protocols include urea for the solubilization and *denaturation* of proteins. However, urea in solution is in equilibrium with ammonium cyanate which decomposes to isocyanic acid reacting with protein primary amino groups and resulting in their carbamylation (Lippincott and Apostol 1999). This modification (~NH-CO-NH_2) gives a mass increment of 43 Da per modified amino group. Therefore, long-term exposure of proteins to high urea concentrations can lead to unfavorable heterogeneity in downstream MS analyses due to carbamylation of lysine, arginine, and N-terminal residues. Consequently, a variable modification for carbamylation of arginine and lysine residues should be taken into account whenever urea is used for sample processing. To minimize the extent of carbamylation, urea solutions should always be used fresh and all operations performed at a temperature below 30 °C. In addition, it is recommended to add methylamine to the urea solution prior to use. However, one should also investigate whether replacing urea by other chaotropic agents, such as sodium deoxycholate or surfactants is appropriate (Proc et al. 2010).

There are also other common biological modifications that should be taken into account. For example, although N-terminal acetylation is rare in bacteria, acetylated N termini are common in archaea and may affect even 15 % of their proteins (Falb et al. 2006). Therefore, the use of appropriate data-mining procedures may

increase the number of identified peptides. In addition, it should be remembered that not all peptides in a sample are represented in the DB, while even spectra derived from non-peptide background constituents can be matched to peptides by a search engine.

In conclusion, the setting of proper search parameters is not trivial because taking into account the above-mentioned modifications will enlarge the search space and thus may prolong the search time substantially. Therefore, the best solution would be to estimate the prevalence of known modifications before setting the parameters for a conventional search engine. For example, the software tool "Preview" (Kil et al. 2011) performs a fast full protein DB searches with a set of product ion spectra in a fraction of time needed by a conventional search engine. It reports: (i) the amount and type of nonspecific digestion, (ii) assays the prevalence of known modifications, and (iii) recognized modifications. Such information not only allows choosing the most appropriate search parameters to maximize the number of correct matches, but also provides timely feedback for the laboratory on sample preparation artifacts, thus improving the overall efficiency and reproducibility of the shotgun-proteomic approach.

DB Searches

It is crucial that matches to all reference proteomes are reported instead of a subset of best hits as commonly done by many search engines. Therefore, the acquired fragmentation spectra could be searched separately against each reference proteome or subsets of combined proteomes, depending on the reporting capabilities of the search engine or experimental needs. For example, SEQUEST, the first software developed for searching MS/MS spectra against sequence DBs and commercially available from Thermo Scientific (San Jose, CA, USA) and Sage-N Research (Milpitas, CA, USA), reports up to 100 matches. Hence, depending on the preliminary information about the sample, the searches should be arranged appropriately. However, searches against a large DB increase the error rate and contribute to the increased rate for false-positive identifications (Cargile et al. 2004). Therefore, there are advantages of using a two-step matching process by performing the initial search against a large DB, followed by a focused DB search against DB strain proteomes with a statistically significant number of matches assigned during the initial search (Jagtap et al. 2013). Moreover, during an initial search to produce a focused DB—it is best to enable only the most common modifications (e.g., oxidized methionine and deamidated asparagine).

In the second step, a smaller and better manageable DB may be used, which may be generated by selecting as the target a set of microbial proteomes representing only one phylum, family, or even genus, and appending these forward sequences with decoy DB sequences. In addition, the smaller DB should include sequences of commonly found contaminants (see Section "Bacterial DBs") and could be searched with an extended list of expected peptide modifications.

Processing of DB Search Results

It is important to remember that an identified peptide may be a false positive regardless of its uniqueness. Moreover, a peptide that is unique throughout the protein sequence DB may be the result of a sequencing error. Therefore, quality assessments of PSMs have to be based on a solid statistical ground by using postprocessors such as PeptideProhet (Keller et al. 2002), Percolator (Käll et al. 2007a; http://www.matrixscience.com/help/percolator_help.html), or q-ranker (Spivak et al. 2009) to apply an optimal scoring function for a particular data set (Granholm and Käll 2011).

DB search algorithms attempt to match every experimental spectrum to DB peptides and report parameters to determine correctness of each PSM. For example, the information contained in each output file generated by a search engine SEQUEST includes: (i) PSMs, (ii) peptide assignments to reference microbial proteins or (if headers were appropriately modified) proteomes in the DB, referred here as "peptide-to-bacterial" (PTB) strain assignments, and (iii) parameters estimating the correctness of PSMs (Xcorr, ΔCn, Sp, RSp, ΔM). However, a better way to express the accuracy of such assignments would be to calculate probabilities that each PSM is correct. For example, Dworzanski et al. (2004) interpreted the above SEQUEST matching parameters using discriminant function (DF) analysis. They arrived at probability scores for PSMs by modeling distributions of correctly and incorrectly identified peptides from a training data set obtained from analysis of a known bacterial strain.

Among many other computational ways to determine such probabilities, the PeptideProphet algorithm (Keller et al. 2002) gained a wide acceptance in the field of proteomics. It may be used as a standalone application or as part of a suite of software tools for the analysis of tandem MS data sets known as the Trans-Proteomic Pipeline (Deutsch et al. 2010). PeptideProphet was also incorporated into BacID/ABOid software and applied for the selection of correct PSMs used for discrimination of diverse microbial strains (Dworzanski et al. 2006, 2010; Jabbour 2010a, b, c, 2014; Wade et al. 2010, 2011). In this approach, BacID/ABOid retrieves and organizes both SEQUEST and PeptideProphet output files by creating a binary matrix of PTB assignments which can be generated using all raw data, or any subset of PSMs selected to ensure high confidence results. The final PTB matrix is created by "filtering out" not only the low-quality PSMs but also identifications matching common contaminants and sequences from the decoy DB, and retaining only a sequence unique set of peptides which are then combined and archived for further processing using a comma separated value (CSV) file format.

Recently, Koskinen et al. (2011) described the approach that "seeks to present DB search results in a more logical format", that is, by creating a minimal set of proteins, grouped into families on the basis of shared peptide matches and by using hierarchical clustering with scores of non-shared peptide matches as a distance metric. This approach is very similar to that used by BacID/ABOid software (Dworzanski et al. 2006, 2010; Deshpande et al. 2011) for presenting DB search results of unknown microbial strains represented as "pseudo-polyproteins." Unfortunately,

the BacID/ABOid software is not available to the scientific community; therefore incorporation of this approach into a family of DB processing tools by Mascot will allow researchers to illustrate how families of strains are related and thus making it easier to make taxonomic or diagnostic decisions.

Subspecies Differentiation and Strain-Level Typing of Bacteria Based on Searching Protein DBs with Peptide MS/MS Spectra

The availability of commercial LC-MS software tools for full characterization of microorganisms and their physiological capabilities have lagged behind the technological advances in MS instrumentation. Although significant effort has been put into development of bioinformatics tools to identify mixtures of proteins, software applications that focus specifically on the identification of microbial strains and characterization of its proteome have been lacking. While commercial search engines combined with available data-mining methods can be used to identify microorganisms, the majority of these tools do not have the ability to take into account intricate phylogenetic relationships among strains which are an important part of characterizing both isolates and microbial mixtures. Therefore, customized DBs and data-mining approaches have been developed by a few research groups to overcome these shortcomings.

Searches of fragmentation mass spectra from trypsinized microbial proteins against DB of reference proteomes return PSMs identifying peptide sequences and can be used for revealing the distribution of PSMs among DB species. Further processing of such assignments allows to: (i) deduce the identity of an isolated organism based on analysis of taxon-specific and taxon-shared sequences, and (ii) uncover intraspecies relatedness based on genomic similarities revealed by analysis of the multidimensional structure of peptide conservation profiles across DB strains. However, there are no standardized approaches on how to perform such analyses; therefore, I will outline only the methods most frequently used by diverse research groups.

Classification and Identification of Bacteria Based on the Number of Shared Peptides

The need for rapid detection, identification, and classification of pathogenic microorganisms is vital for clinical, epidemiological, agricultural, and public health emergencies that include a potential biological terrorist attack. Therefore, the efforts to achieve such objectives were substantially intensified after the October 2001 anthrax attack in the USA. Many methods were proposed for this purpose and some

of them were based on mass spectrometry, for example, analysis of microbial cell pyrolysis products, lipid extracts, nucleic acids, proteins, or amino acid sequences of protein digestion products, that is, peptides.

The use of protein sequences for the identification of species is not new (Sanger 1959) and was underscored by Frederic Sanger during his Nobel Prize Lecture in 1952. His idea was next revitalized with the advent of high-throughput proteomics era by C. Fensealu, P. Demirev, J. Yates, and others (Yates 1998; Demirev et al. 1999; Fenselau and Demirev 2001).

One of the bottom-up proteomic methods aimed for rapid identification and classification of microbes based on the concept of the number shared peptides was invented by J. P. Dworzanski and L. Li (Dworzanski et al. 2004). They coupled LC/MS/MS analysis of peptides obtained by trypsin digestion of whole cell bacterial extracts with searching an in-house created DB obtained by translating available genomic sequences with the ORF finding software Glimmer. The analysis of peptide sequences and their matches to proteomes of reference bacteria in the DB allowed them to identify selected bacterial samples down to the species and strain levels. Furthermore, they could identify the isolates regardless of the culture growth phase and with no prior knowledge of the test sample (Dworzanski et al. 2006). This procedure was next automated by using algorithms BacID/ABOid developed by J.P. Dworzanski and implemented by S. Deshpande in Visual Basic and Perl (Dworzanski et al. 2006; Deshpande et al. 2011) and applied for analysis of diverse agents of biological origin (ABO) (Dworzanski et al. 2010; Jabbour et al. 2010a, b, c, 2014; Wade et al. 2010, 2011).

Peptide-to-Taxa Assignments: Determination of the Closest Neighbor

The shotgun-proteomic analysis of an unknown strain followed by DB searches and validation of determined PSMs gives a peptide profile of an unknown strain (u). That type of peptide profile can be represented as a column vector with each component indicating that the specified sequence is encoded in its genome. On the other hand, each of these peptides may match only one DB reference proteome (unique peptides) or many (shared peptides). Thus, each peptide is characterized by a "phylogenetic" profile across DB reference strains and may be represented as a row vector with each component taking a value of either one or zero, where one/zero indicates the presence/absence of the exact matching peptide sequence in the corresponding DB proteome. These row vectors form a matrix of assignments that may be visualized as a virtual array of peptides assigned to theoretical proteomes of DB strains (Fig. 5.1, where 1/0 are represented as closed/open circles, respectively). This way, the results of MS/MS analysis may be represented as a binary map of PTB strain assignments where similar peptide profiles per reference strains indicate a correlated pattern of relatedness among such DB strains while similar "phylogenetic" profiles of peptides across strains suggest that they originate from homologous proteins (Dworzanski et al. 2006).

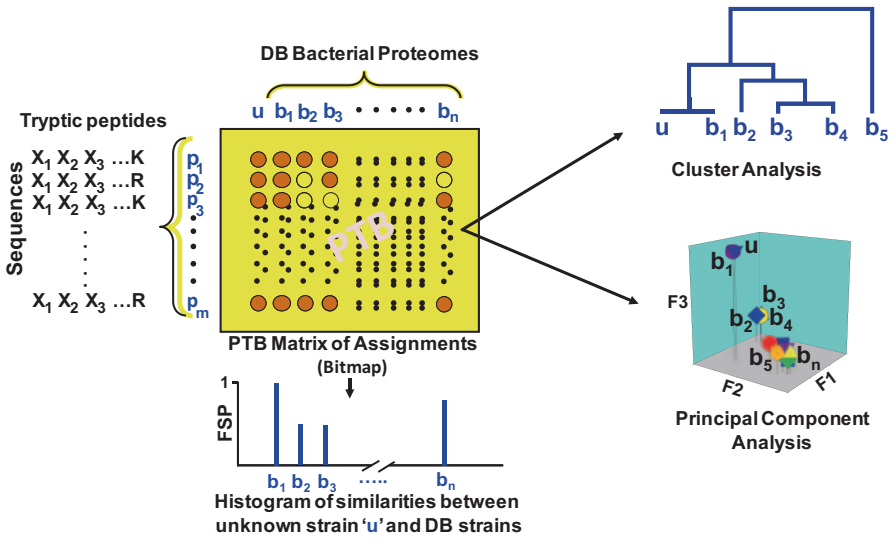


Fig. 5.1 Schematic representation of mapping tryptic peptides sequences (p_1, p_2, p_3, …, p_m) identified by shotgun-proteomics analysis of an unknown *(u)* strain to database *(DB)* proteomes of reference strains (b_1, b_2, b_3, …, b_n) and analysis of the created matrix of peptide-to-bacterial *(PTB)* strain assignments using multivariate statistical methods to reveal the closest DB neighbor. "FSP" in the "Histogram of similarities" stands for fractions of shared peptide sequences between *u* and each DB bacterial proteome. (Reprinted with permission from Dworzanski et al. (2006, pp. 76–87). Copyright 2006 American Chemical Society)

Dworzanski et al. (2004) carried out 1D HPLC-MS/MS analysis of tryptic digests derived from protein extracts of selected bacterial strains with fully sequenced genomes and used a statistical scoring algorithm to rank MS/MS spectral matching results for bacterial identification. Peptides with scores exceeding a threshold probability value were accepted and assigned to the bacterial proteomes represented in the DB. Because they used modified header lines of each protein in a DB (see Section "Bacterial DBs" for details), SEQUEST was recognizing each proteome as a single "pseudo-polyprotein" and assigning PSMs directly to reference DB strains instead to particular proteins in each proteome.

All the PTB strain assignments reported by SEQUEST were then organized as PTB matrices allowing for easy transformations and presentation of results that included: (i) ranking of assignments in the form of histograms showing the number of matching peptides per reference proteome (similarity scores) or (ii) displaying the distribution of unique peptides to further improve identification by the removal of "degenerate peptides," that is, peptides shared by reference proteomes (Dworzanski et al. 2004).

The selection of unique peptides was carried out by assuming that a DB strain proteome with the highest number of matching peptides is deemed to be the most likely candidate of a true match. With this assumption, deconvolution can be performed iteratively by selecting the highest scoring bacterium and filtering out shared

peptides from histogram bins associated with all the remaining bacteria, which generates a new histogram of peptide matches per strain. A subsequent step involves the removal of peptides from the second highest scoring organism in the newly assembled histogram, and so on. Such a deconvolution filter acts as the "Occham razor" that removes shared peptide sequences from the PTB matrix, usually associated with orthologous proteins, and reveals the minimum set of strains capable of explaining all accepted PSMs.

Initially, the approach developed by Dworzanski et al. (2004) was focused on the identification of bacteria with fully sequenced genomes and therefore represented in the DB. However, they also reported that although unique peptides from the correctly identified strain can explain all high scoring PSMs, it is not the case for a strain not represented in the DB. In such cases, the Occham razor-type filter reveals the nearest DB neighbors of an unknown strain, reflecting taxonomical position of an unknown microorganism.

Recently, Tracz et al. (2013) reported a novel variation of the above method for bacteria identification implemented by using the Mascot search engine. In this approach, they compared the number of peptides shared between the unknown and DB strains by tricking Mascot to report such assignments. They achieved it by creating "pseudo-polyproteins" of concatenated protein sequences of each DB strain proteome (see Section "Bacterial DBs" for details), so searches against such a DB report PSMs to DB strains instead to particular proteins.

In proof-of-concept experiments, they analyzed whole cell protein extracts from selected *Bacillus, Francisella, Yersinia,* and *Clostridium* strains with complete genome sequences, or their close neighbors with the same status, which were chosen as surrogates for highly pathogenic species (Fig. 5.2). To speed up the sample preparation process, the reduced and C-alkylated proteins were digested with trypsin at elevated temperature and analyzed with a nano-LC-LTQ Orbitrap mass spectrom-

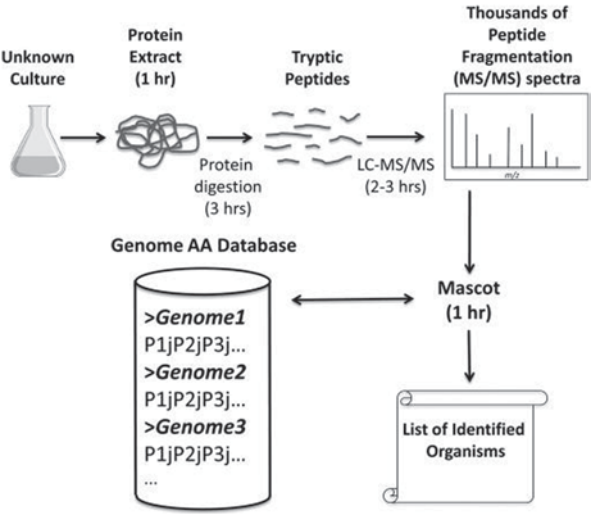


Fig. 5.2 Schematic of a shotgun-proteomics "genome identification" method that in less than 8 h (postculture) allows for strain identification. This method involves: (i) protein extraction and in-solution trypsin digestion, (ii) analysis of tryptic peptides by LC-MS/MS, and (iii) using MS data to search against a novel DB of genomes represented by concatenated proteins of genome-predicted proteomes. (Reprinted with permission from Tracz et al. (2013, pp. 54–57). Copyright 2013 Elsevier B.V.)

eter. The acquired tandem mass spectra were searched against "Genome AA" DB, and the search results were exported from the Mascot Website interface in a CSV file format.

The results file contains all accepted PSMs and DB strains ranked according to Mascot scores reflecting, among others, the total number of peptide matches per strain. These numbers are graphed as black bars in Fig. 5.3 depicting results from identification of strains by LC-MS/MS. However, peptides assigned to the highest scoring strain (so-called "red bold" matches in the Mascot jargon) could be divided into strain specific or "unique" peptides and those shared with other strains and called "degenerate." Mascot flags the latter peptides when they are assigned to any other strain in the report as "not bold red"; thus allowing to filter out matches to degenerate peptides from all remaining strains. This process is repeated in regard to the second highest ranking strain and so on, allowing counting only unique matches and preparing a minimal list of strains contributing to the pool of identified peptides. The numbers of such peptides per strain are presented as gray bars in Fig. 5.3 and allow for clear identification of analyzed *Francisella* strains as *Francisella tularensis* LVS and *Francisella philomiragia* subsp. philomiragia. Note that the analyzed strain of *F. tularensis* LVS is represented in the DB while the strain of *F. philomiragia* subsp. philomiragia (ATCC 251015) is represented in the DB only by a different strain of this subspecies, that is, strain ATCC 25017.

Under these circumstances the highest scoring strain, LVS, was correctly identified because among all unique peptides, the matches to other closely related strains were lower than 0.2%. However, in case of ATCC 25015^T strain matches to other strains were substantially higher because 37 (2.6%) of unique peptides matched *Francisella noatunensis* subsp. orientalis str. Toba 04 and 30 (2%) matched the *Francisella sp.* TX077308 strain. This indicates minor sequence differences between strains ATCC 25015 and ATCC 25017^T, and proves that LC-MS/MS can potentially discriminate isolates from the subspecies philomiragia.

Tracz et al. (2013) pointed out the advantages of their approach such as the lack of any prior knowledge of the analyzed microorganism, and the capability of generating organism-specific sequence data. In addition, their method can provide relative protein expression levels, including the confirmation of virulence factor expression, which has relatively low cost of consumables per sample; and a relatively fast turnaround time (<8 h postculture). Moreover, it can be easily implemented in a typical proteomics laboratory.

Analysis of Subproteomes

Although subproteome analyses, by definition, are limited in scope, they usually provide comprehensive representation and coverage of specific protein types in comparison to whole cell proteome approaches. Among subproteomes investigated for subspecies identification of bacteria, the most attention attracted surface and membrane proteins, especially OMPs of Gram-negative bacteria, surface layer (S-layer) proteins of Gram-positive bacteria, and ECPs.

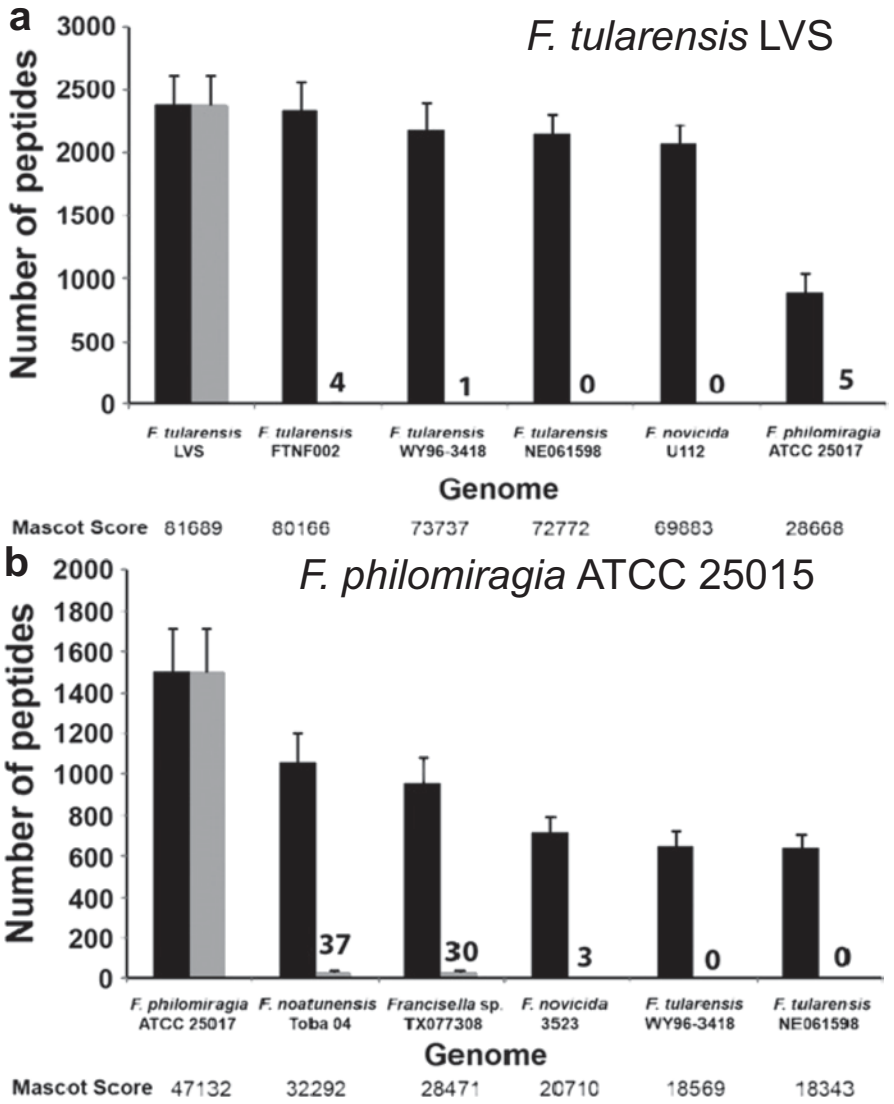


Fig. 5.3 Representative results from identification of bacterial species by LC-MS/MS. DB search results are plotted for *Francisella tularensis* LVS (**a**) and *Francisella philomiragia* ATCC 25015 (**b**). The total number of shared peptides *(black bars)* and the number of strain unique peptides *(gray bars)* for identified bacterial genome-predicted proteomes were sorted by Mascot scores. (Reprinted with permission from Tracz et al. (2013, pp. 54–57). Copyright 2013 Elsevier B.V.)

Surface proteins, including OMPs play a critical role in processes leading to pathogenicity by mediating interaction with a host, evasion of the immune system, efflux of antibiotics, and import of nutrients. Due to their location, they interface the cell and the environment and are candidate targets for developing protective

strategies (vaccines and therapeutics) as well as detection and identification strategies for microbial strains. For example, Jabbour et al. 2010b showed that shotgun-proteomics analysis of OMPs from the *Yersinia pestis* CO92 strain provided unambiguous strain-level identification with all identified tryptic peptides matching the correct DB reference strain, while the remaining DB reference strains of *Y. pestis,* that is, 91001, Antiqua, Nepal 516, Kim, and *Yersinia pseudotuberculosis* IP 32953 were ranked as distant matches based on the number of shared peptides. In addition to strain identification, the results of the same analysis also provided a list of proteins known as being associated with established *Yersinia* virulence factors, like plasmid-encoded plasminogen activator protease precursor and the toxin protein.

Karlsson et al. (2012) analyzed surface-exposed proteins of fully sequenced *Helicobacter pylori* strains J99, ATCC 26695, and the type strain of this species, CCUG 17874^{T}, by using shotgun-proteomics method applied to intact cells immobilized in the flow channel of a microfluidic device called lipid/protein interaction (LPI)-FlowCell. The released and identified peptides were matched to 38 reference strains with complete genome sequences, including 26 *H. pylori* and 12 strains from other species of the *Helicobacter* genus. They showed that this method worked well for discriminating different strains of *H. pylori,* including the strain not represented in the DB.

Wade et al. (2011)investigated the discrimination of pathogenic and nonpathogenic strains of *Francisella tularensis* and *Burkholderia pseudomallei* by using shotgun-proteomics analyses. They found that LC-MS/MS analysis of trypsinized, OMP-enriched subproteomes of these microorganisms combined with data processing that included the BACid software application that allowed for confident subspecies identification and discrimination between pathogenic and nonpathogenic strains of the same species. For example, they analyzed the OMP extract of a highly virulent strain *F. tularensis* subsp. *tularensis* Schu S4, considered a potential bioterrorism agent because it causes the severe disease called type A tularemia, and the analysis of the attenuated strain of *F. tularensis* subsp. *holarctica,* known as the only live vaccine strain (LVS). These strains represent two of multiple recognized *F. tularensis* subspecies that differ in virulence and lethality following infection, that is, *tularensis,* causing the most severe disease, moderately virulent subspecies *holarctica,* followed by *mediasiatica* and *novicida* causing infections only in immunocompromised individuals (Steiner et al. 2014). Genomic analysis suggests that the subspecies of *Francisella tularensis* have evolved by vertical descent, through unidirectional gene losses from the highly virulent strain of *F. tularensis* subsp. *tularensis* which gave the less virulent *F. tularensis* subsp. *holarctica* strains. Furthermore, the attenuated LVS strain also evolved from the *holarctica* strain through gene losses, because complementation of LVS with genes *pilA* and FTT0918 restored its virulence to the level of virulent *holarctica* strains (Forslund et al. 2006; Salomonsson et al. 2009).

Shotgun proteomics analyses took advantage of such differences by allowing not only distinguishing different *Francisella tularensis* subspecies but also for confident discrimination between similar strains. For instance, analysis of the strain Schu S4 allowed for correct identification of this strain at the subspecies level *(tularensis).*

The analysis also discriminated this strain from other *F. tularensis* subsp. *tularensis* strains, that is, FSC 198, WY96198, and identified strain unique peptides from proteins associated with known virulence factors, like type-IV pili fiber building block protein (Lindgren et al. 2009).

Sequence variability is a common feature in surface and secreted proteins of microorganisms because such variability may confer increased fitness allowing the pathogen to use alternative receptors and infect different tissues or even different species. In most cases the variability probably reflects antigenic variation, which allows the pathogen to evade protective immunity in an infected host. It is commonly assumed that conservation of a limited number of residues is sufficient to promote correct protein folding and/or to confer a specific function, while other residues may vary and cause changes in antigenic properties of the protein. For example, Jabbour et al. (2014) reported that ECPs that include both actively secreted and those originating from leaking through or shedding cellular membranes could be used for the characterization of pathogenic *E. coli* strains. These included enterohemorrhagic *E. coli* (EHEC) that cause hemorrhagic colitis and enteroaggregative (EAEC) strains, like the serotype O104H4 that caused the fatal outbreak which occurred in Germany in 2011. They found shotgun-proteomics analysis of ECPs very useful and practical for differentiation among EHEC and EAEC strains due to the increased number of strain-unique peptides identified in comparison to their results obtained with whole cell protein extracts.

Confirmation of the Taxonomic Position of an Unknown Strain

Generally, analysis of a strain not represented in the DB indicates that a single DB strain cannot explain all accepted PSMs. However, a similar output could also be obtained by analysis of a mixed-culture sample or by contamination with other microbial proteins/peptides in the analytical laboratory, for example, through sample carryover. Therefore, to exclude the risk of cross-contamination or sample carryover, the profiles of all identified peptides represented as a binary matrix of PTB assignments can be further analyzed to infer taxonomic positions of contributing strains. The approach devised by Dworzanski et al. (2006) is based on the lowest common ancestor (LCA) strategy of inferring taxonomic position from peptide sequences by mapping them to "pseudo-super-proteomes" of DB strains grouped into hierarchical taxonomic units. A very similar strategy was later incorporated into the MEGAN algorithm (Huson et al. 2007) and other software tools for analysis of metagenomic data and metaproteomic data, like the Unipept web application that—based on submitted tryptic peptides—returns an interactive tree map by providing an insight into the sample biodiversity (Mesuere et al. 2012).

In this approach all DB strains are classified in accordance with the established taxonomy of prokaryotic microorganisms where similar bacterial strains are grouped into species while groupings of very similar species form genera. These species/genus levels in the taxonomic position within the classification scheme is reflected in the binomial name of bacteria. However, groupings do not stop at this

level, but also include higher taxonomic arrangements of organisms into hierarchical classifications based on similarities. Namely, similar genera are placed in the same family; similar families in the same order; similar orders in the same class; similar classes in the same phylum; and finally all bacterial and archaeal phyla form the domains (or "kingdoms") of *Bacteria* and *Archaea,* respectively. Consequently, the classification of an unknown strain involves mapping of its peptides to taxa represented by "pseudo-super-proteomes" composed of DB strains grouped into the descending taxonomic ranks: phyla, classes, orders, families, genera, and species in accordance with the NCBI taxonomic classification hierarchy (Federhen 2012).

According to this peptide-centric LCA algorithm, a peptide is assigned to a lower level taxon only if its sequence is unique to this taxon; otherwise it remain assigned only to the higher level taxon and the process proceeds from domains to phyla, classes, orders, families, genera, species, and subspecies levels. For example, a peptide is assigned to a given species only if it does not match with any other species contained in the sequence DB; conversely, if the sequence is shared among several species contained in the DB, all belonging to the same genus, the sequence is unambiguously assigned only at the genus level. This way, widely conserved peptide sequences are always assigned to high-order taxa and highly variable provide the most accurate results for discrimination at the subspecies level. This type of analysis takes into account the error rate determined for the accepted set of PSMs and can be executed in a few seconds by a software application (ABOid, Deshpande et al. 2011). Moreover, this approach could also be applied to metaproteomic analyses of microbial mixtures.

The above described procedure is quite useful because it focuses the final classification process on a group of reference strains that are closest relatives of the isolated one. For example, shotgun-proteomic analysis of the whole cell protein extract of a poisonous strain isolated from the Indonesian rice dish followed by the above classification method indicated that it can be classified as *Firmicutes* → *Bacilli* → *Bacillales* → *Bacillaceae* → *Bacillus* → *B. cereus* group strain with the highest number of unique assignments matching the *B. cereus* ATCC 14579 strain. The correctness of this identification was confirmed by DDH analysis, and sequencing of the 16S rRNA and *gyrB* phylogenetic markers. In addition, this strain was serotyped based on the polymorphism of flagellar H-antigen as H-10 (Dworzanski et al. 2010). Bitmap representation of the PTB matrix of 599 peptide sequences from this strain assigned to the nearest DB strains is shown in Fig. 5.4. The displayed DB strains and peptide sequences were rearranged and analyzed by two-way hierarchical cluster analysis (HCA) with PermutMatrix (http://www.atgc-montpellier.fr/permutmatrix/) using Euclidean distances and unweighted pair group averages as the aggregation method (Caraux and Pinloche 2005). Dworzanski et al. (2010) found that the rice isolate shared 526 peptides (FSP = 0.88) with the closest DB neighbor strain (*B. cereus* ATCC 14579) while other members of the *B. cereus* group, and especially a clade of *B. anthracis* strains, were more distant (FSP = 0.82). Furthermore, the remaining *Bacillaceae* strains shared only 3–6 % of peptides with the serotype H-10.

The bitmap representation of PTB matches simplifies comparative analysis of strains by focusing on peptides with high discriminative power. For example,

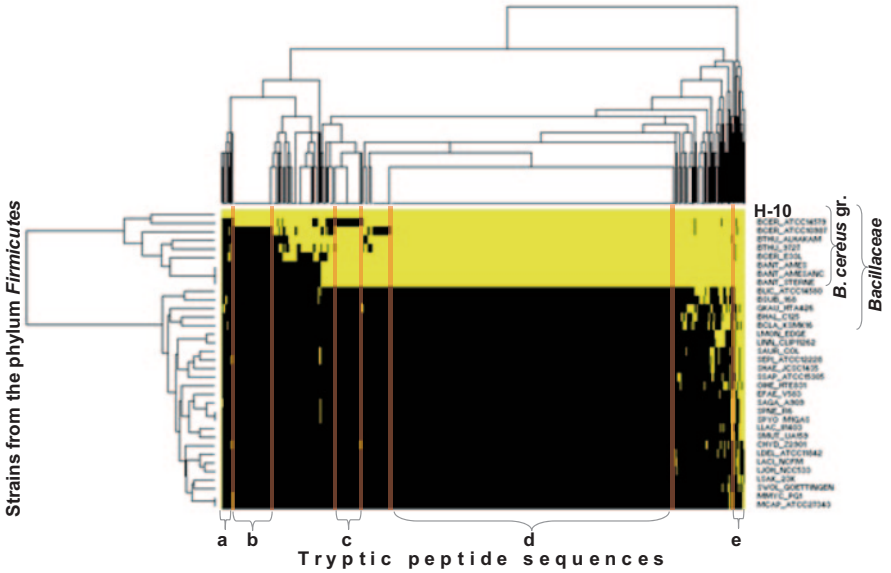


Fig. 5.4 Bitmap representation of the clustered data matrix of 599 peptide sequences from the *B. cereus* serotype H-10 isolate assigned to the nearest neighbors in the DB. Each *yellow (white)* cell represents the presence and each *black* cell the absence of a peptide-to-bacterium match. Two-way HCA was performed with PermutMatrix (Caraux and Pinloche 2005) using Euclidean distances and unweighted pair group averages as the aggregation method. The dendrogram of bacterial strains shows that the H-10 strain clusters with the *B. cereus* group of bacteria and forms a subcluster with a type strain *B. cereus* ATCC 14579. The dendrogram of peptides allows visual selection of sequences. For instance, clusters marked "a" through "e" indicate groupings of peptides with different discriminative/diagnostic power. Abbreviations of DB bacterial strains: XYYY_Z…Z, where X represents the first letter of a genus name, YYY represent the first three letters of a species name, and Z…Z represent the strain name. (Reprinted with permission from Dworzanski et al. (2010, pp. 145–155). Copyright 2010 American Chemical Society)

peptides marked as cluster "d" in Fig. 5.4 represent the majority of identified peptides while they only discriminate between the *B. cereus* group and remaining DB strains. On the other side, clusters "a" and "c" reveal sequences that discriminate serotype H-10 and its closest DB neighbor while cluster "e" indicates peptides with low discriminatory power. Indeed, peptides grouped in the latter cluster were derived from proteins with highly conserved sequences, that is, ribosomal proteins, elongation factors, and chaperones.

Relationship Between the Fraction of Shared Peptides (FSP) and Conservation of the Genome/Proteome

Currently, public DBs list multiple genome sequences for many microbial species and this increasing number of complete genome sequences together with next-

generation sequencing capabilities available in many laboratories provides a wealth of new data for analysis of genomic similarities. Among many attempts to use such data to find similarities between strains, currently the best approach seems to be by quantifying the DNA conservation of bacterial genomes. Accordingly, the relatedness between two bacterial strains can be determined by comparing sequences of all homologous genes or their protein products through the computation of sequence-derived parameters that estimate ANI or AAI indices that correspond to the traditional DDH standard of the current species definition (Konstantinidis and Tiedje 2005a, b; Goris et al. 2007). Several programs are available for calculating the ANI; for example, JSpecies can be found at the Website: http://www.imedea.uib.es/jspecies/ (Richter and Rosselló-Móra 2009).

Despite their taxonomic value as a robust and universal measure of strain similarities, these indices are not applicable to nonsequenced species, such as clinical, food, or environmental isolates. Therefore, shotgun-proteomics methods that can indirectly measure or can estimate an AAI and DDH indices from experimentally determined FSP values could be applied for strain-level discrimination and typing of bacteria.

Inter-relationships between FSPs determined from shotgun-proteomic experiments and widely used genome conservation measures, that is, DDH and the ANI/AAI indices were estimated by Dworzanski et al. (2010). Their approach was based on the Kimura (1969) model for the estimation of amino acid substitution rates for homologous proteins. However, they extended this model to short DNA segments (used for the determination of DDH values) and (tryptic) peptides viewed as expression products of DNA segments, by making the following assumptions. First, they assumed that "homologous proteins" in the Kimura model could be substituted by "pseudo-polyproteins" of closely related strains. Second, "amino acid substitutions" arising from genomic mutations in strains (e.g., SNPs) could be replaced on the (tryptic) peptidome level by "no longer shared peptides" between "pseudo-polyproteins" representing bacterial strains. Consequently, differences between strains manifested as amino acid substitutions and quantified as AAI indices, are reflected at the DNA level by DDH values, and on the peptidome level by FSPs determined from shotgun-proteomics experiments.

With the above assumptions the time *(t)* since the divergence of any two strains from a common ancestor can be expressed as $t = -2.3 \log (\mathrm{AAI})/2k_{aa}$, where k_{aa} is the rate of substitution per amino acid per time. However, by substituting "amino acid sites" in each proteome with "peptide sites" T_p of length L (where L represents the number of amino acid residues), the "fraction of identical amino acid sites" (AAI) could be substituted by the "fraction of identical peptide sites" (FSP index), and the time since the strain divergence could be calculated from the equation $t = -2.3 \log (\mathrm{FSP})/2k_p$, where k_p refers to the rate constant for peptide substitutions. Obviously, for any given pair of microorganisms, the time since divergence is independent from the similarity measures used to express it. Hence, by equating time, expressed using the above shown equations, the relationship between FSP and genome/proteome conservation index can be estimated in the exponential form as $\mathrm{FSP} = (\mathrm{AAI})^L$, where the peptide length L is equivalent to the ratio of substitution rates ($L = k_p/k_{aa}$) (Dworzanski et al. 2010).

In accordance with this model, the fraction of peptides shared between two microbial proteomes is always lower than the AAI index value and depends on the peptide length. These inter-relationships are depicted in Fig. 5.5 for peptides with 8, 15, and 30 amino acids that represent a typical range of peptide lengths identified in shotgun experiments.

Due to logarithmic relationships between FSP and AAI indexes, this model predicts that relatively small differences in the amino acid identities are associated with substantially decreased values of the FSP index. Indeed, as shown above (see Section "Confirmation of the Taxonomic Position of an Unknown Strain") for a *B. cereus* strain isolated from food (H-10), the FSP with its nearest neighbor was 0.82, or 82%; however, for more distant strains from the same genus, the FSP values dropped to only a few percent. Furthermore, this model predicts that for proteomes characterized by 94% sequence identity on the amino acid level (AAI), that is assumed as a cutoff value for strains belonging to the same species (Konstantinidis and Tiedje 2005b), the FSP for 15 amino acid residues long peptides is only 40% and much lower for longer peptides (Fig. 5.5). Therefore, the FSP index is char-

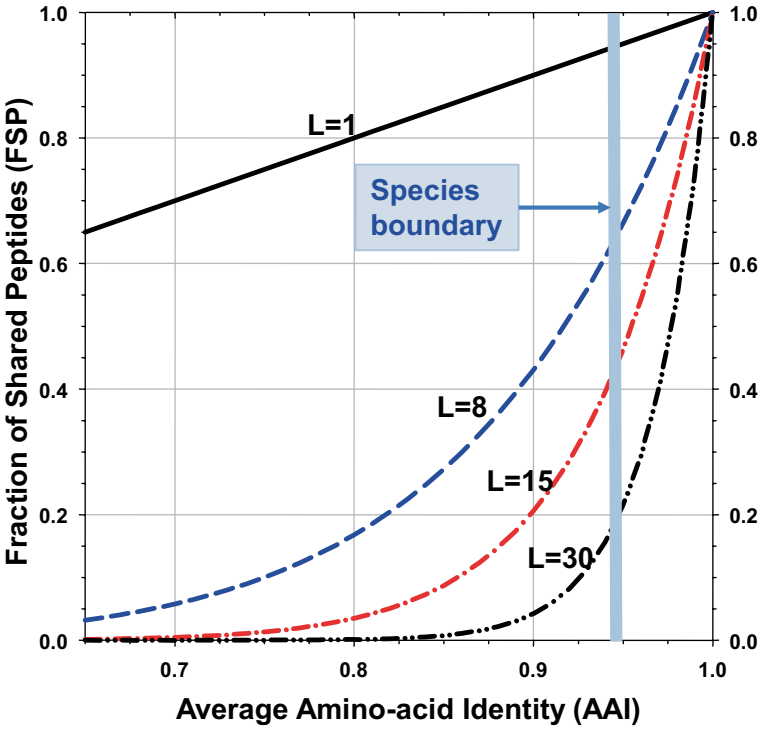


Fig. 5.5 Relationships between proteome conservation expressed as the averaged amino acid identity (AAI) index and the fraction of shared peptides (FSP) calculated for peptides of different length (L) by using the equation $FSP=(AAI)^{L}$. (Reprinted with permission from Dworzanski et al. (2010, pp. 145–155). Copyright 2010 American Chemical Society)

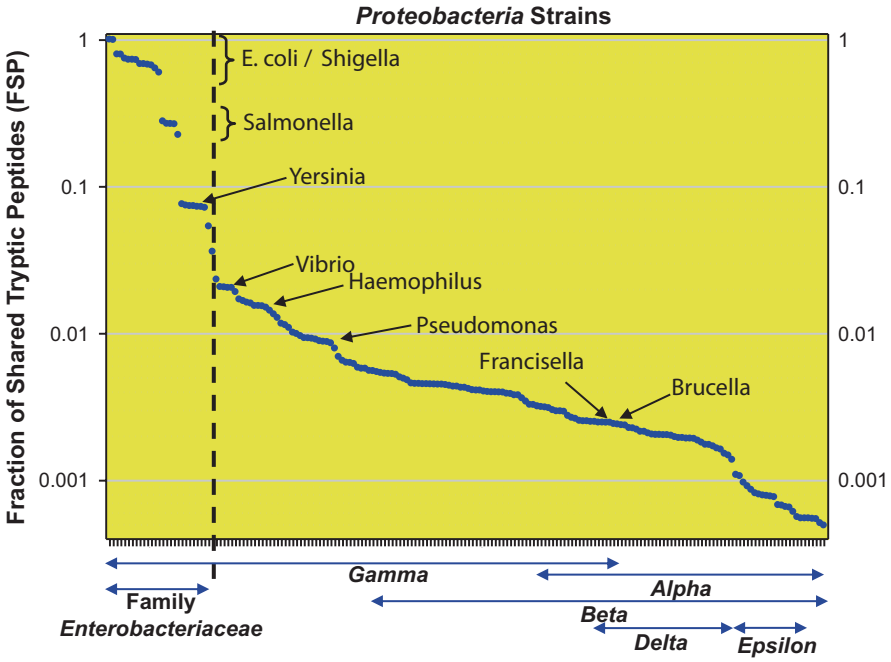


Fig. 5.6 Theoretical proteome similarities between *E. coli* K-12 and other DB alpha-, beta-, gamma-, delta-, and epsilon-proteobacterial strains, expressed by tryptic peptide FSPs calculated as Dice indices that take into account the proportion of shared (common) peptides to the averaged number of unique peptides found in both proteomes. Note that only bacterial strains from the same family (*Enterobacteriaceae*) share more than 2 % of tryptic peptides

acterized by a good resolving power required for discrimination of closely related strains, as demonstrated by whole proteome similarities between *E. coli* K-12 strain and other DB strains from the phylum *Proteobacteria* (Fig. 5.6). These theoretical FSPs were calculated as Dice similarity indices based on in silico digestion of reference DB proteomes following trypsin specificity rules, allowing up to two missed cleavages per peptide, and counting only tryptic peptides with M_r in the 700−3500 Da range (Dworzanski et al. 2010).

Conceptually, the overall genomic similarity between two strains expressed as a DDH value is equivalent to the FSP index because the matched peptides between two strains reflect DNA segments which would potentially form perfect hybrid pairings. Therefore, for consistency with the FSP term, the DDH value could be considered as a fraction of shared DNA segments between strains. On the basis of available data, relationships between these similarities were approximated by a linear function ($DDH = 1.597 \times FSP - 0.707$, $R^2 = 0.78$) and used to calibrate proteomic similarities against DDH values (Dworzanski et al. 2010). According to this equation, the DDH cutoff of 70 %, which is used for species discrimination (Wayne et al. 1987), is equivalent to experimentally determine proteomic similarities of 88 % (FSP, 0.86−0.90). Accordingly, strains with the FSP values higher than

88 % should be treated as one species. However, the FSP values used in the above work were obtained from trypsinized whole cell protein extracts which may over-represent peptides derived from highly conserved, high copy number proteins like those involved in the information processing. Therefore, these FSP values may be biased toward a higher FSP values in comparison to complete proteomes or some subproteomes.

Indeed, based on in silico digestion of all predicted tryptic peptides between reference proteomes of *B. anthracis* Sterne or *B. cereus* ATCC 14579 and other *Bacillaceae* strains with sequenced genomes, Dworzanski et al. (2010) found that theoretical FSPs were substantially lower than their experimental values. For example, for a pair of theoretical proteomes with a calculated FSP of 0.7, the experimentally determined values were found in the range 0.83−0.89. However, in the case of tryptic peptides released from surface-exposed proteins of *H. pylori* strains J99 and 26695, Karlsson et al. (2012) found intraspecies FSPs between these and more than 20 other *H. pylori* strains to be in the range of 0.65–0.82, that is, closer to the expected theoretical value.

It is well known that HGT, gene duplications, indels, and nucleotide substitutions are major evolutionary processes shaping microbial genomes, and closely related organisms engage in genetic exchange more frequently than distantly related ones. Recently, Caro-Quintero and Konstantinidis (2014) quantified HGT between bacterial genomes representing different phyla and found that inter-phylum HGT may affect up to ~16 % of the total genes. However, ribosomal and other conserved protein-coding genes were subjected to HGT at least 150 times less frequently than genes encoding metabolic enzymes or ATP-binding cassette transporters (ABC transporters). Therefore, sequences of the latter genes and their products have more discriminatory power for strain differentiation that is reflected in lower FSP values.

Turse et al. (2010) carried out investigations aimed to find FSPs between bacterial strains as a function of separating them evolutionary distances determined from 16S rDNA sequences with CLUSTAL W. In the first stage ("proof of concept") they performed LC-MS/MS analyzes of trypsin digested whole cell protein extracts from *Shewanella* strains and phylogenetically distant strains of *S. enterica* subsp. *enterica* and *Deinococcus radiodurans*. Although *Shewanella* and *Salmonella* strains are both classified as *Gamma-Proteobacteria*, Deinococcus is much more distant from both of these genera because it belongs to the separate phylum, *Deinococcus-Thermus*. They found that with increasing evolutionary distances between bacteria, the determined FSPs decrease exponentially, that is, in a fashion expected from relationships between FSPs and evolutionary similarities expressed as AAI indexes. For example, FSPs between most genetically distant *Shewanella* strains was only 6 %, while strains from this genus shared less than 1 % peptides with the *Salmonella* strain.

In the second stage Turse et al. (2010) analyzed four Columbia River environmental isolates designated as HRCR-1, 2, 4, and 5, which based on 16S rDNA sequences, showed phylogenetic affiliation with *Shewanella oneidensis* MR-1 or *Shewanella putrefaciens* CN32 strains. These findings were confirmed by the determined FSPs calculated from LC-MS/MS spectra acquired during analyses of these

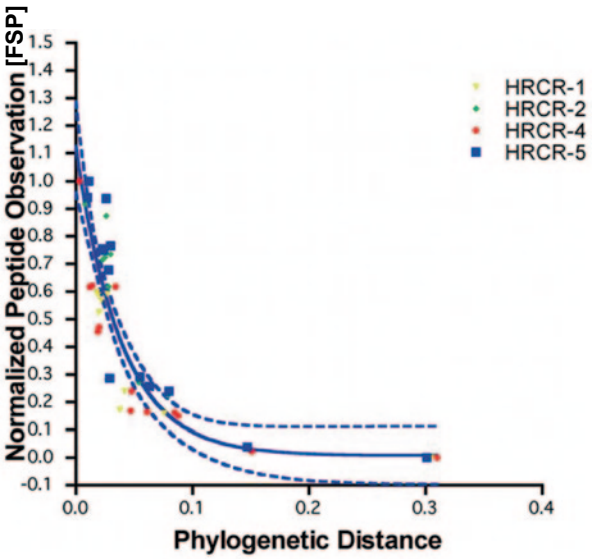


Fig. 5.7 The fraction of shared peptides (FSP is denoted here as *Normalized Peptide Observation*) between the environmental *Shewanella* isolates *(HRCR-1 through 5)* and DB strains plotted against phylogenetic (evolutionary) distances determined from 16S rDNA sequences. Reproduced from Turse et al. (2010). Open access journal

isolates. They also found that in all cases FSPs plotted against evolutionary distances were decreasing exponentially (Fig. 5.7). Note that according to the terminology used by Turse et al. FSPs are called "normalized peptide observations" indicating that the number of observed shared peptides was normalized, that is, divided by the total number of identified peptides determined by analyzing the actual reference strain under identical conditions.

Karlsson et al. (2012) analyzed surface proteins of *H. pylori* strains J99 ATCC 26695, and CCUG 17874^T by "shaving" surface-exposed domains of these proteins directly from intact cells immobilized in the flow channel of a microfluidic device. The released and identified peptides were matched to 38 reference strains with complete genome sequences, including 26 of *H. pylori* and 12 strains from other species of the *Helicobacter* genus. In the above-mentioned study the authors compared genomic similarities between *Helicobacter* strains based on the number of shared peptides with the well-established methods based on analysis of DNA sequences: (i) the ANI index (Konstantinidis and Tiedje 2005b) calculated using both BLAST (ANIb) and MUMmer algorithms; and (ii) tetra-nucleotide frequency correlation coefficient (TETRA, Bohlin et al 2008) that bypasses the complexity of performing multiple sequence alignments and avoids the ambiguity of choosing individual genes by inferring evolutionary relationships between species directly from their complete genomic sequences.

The ANI values between the same species strains are typically 94% or greater while between strains of distinct species exhibit values below 94%. The ANI

values observed between *H. pylori* strain J99 and the other *H. pylori* strains were at a similar level of ca. 94%, that is, typical for intraspecies diversity, with the exception of *H. pylori* Shi 470, which had a lower ANI value of 93% (Fig. 5.8). However, the FSP values with strain J99 showed slightly better resolution for other *H. pylori* strains (FSPs at the level 0.75–0.69) with the lowest FSP value (0.686) for the Shi 470 strain. Also, the comparison of ANI and FSP values between *H. pylori* 26695 and other *H. pylori* strains showed a similar trend; however, ANIs were slightly higher (95%), with the exception of strain J99. When comparing to other species of *Helicobacter,* such as *Helicobacter acinonychis* and *Helicobacter hepaticus,* the ANI values for J99 dropped to approximately 89% and 66%, respectively, which was reflected in a lower peptide matches per strain that is equivalent to FSPs dropping down to 0.52 and 0.07, respectively (Fig. 5.8). These values correlated well with TETRA results between genomes which also have been shown to be high (>0.99) when ANI and FSP values are high, although stronger correlation was observed for interspecies genome comparisons, for example, in the case of *H. hepaticus* and other 11 strains outside the *H. pylori* species (Karlsson et al 2012).

The numbers of peptides shared between the *H. pylori* strains J99 and 26695 and strains of *H. pylori* for which genome sequences exist were also compared to the

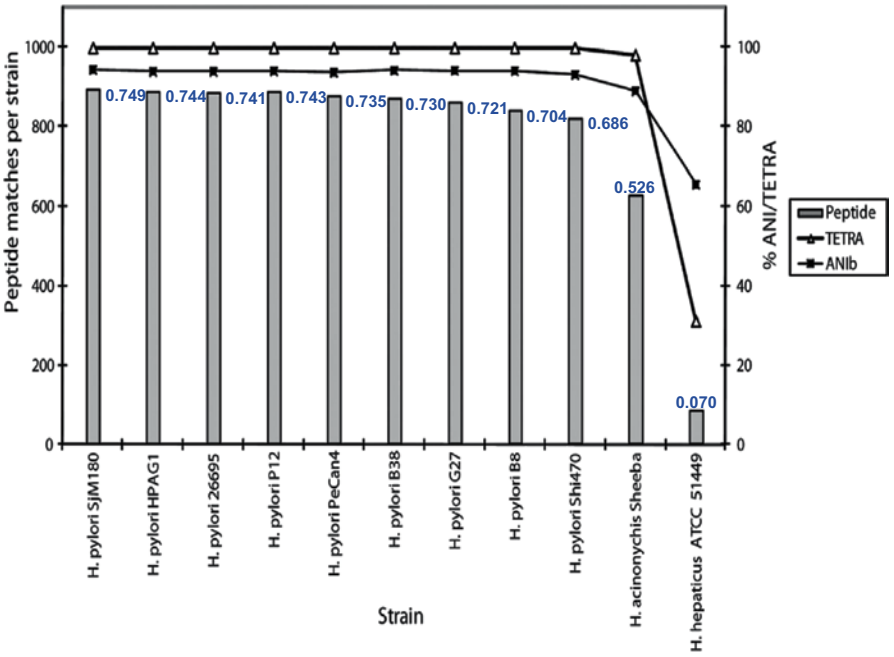


Fig. 5.8 Peptide matches per *Helicobacter* strains for the *H. pylori* J99 sample, compared to whole-genome analyses using TETRA and ANI. The peptide matches per strain are shown as bars accompanied by the FSPs values, and the TETRA (multiplied by 100) and ANI indices are depicted by lines connected by symbols as indicated by the legend box. (Reprinted with permission from Karlsson et al. (2012, pp. 2710–2720). Copyright 2012 American Chemical Society)

results from multilocus sequence typing (MLST) analyses. Karlsson et al. (2012) carried such sequence analyses of internal fragments for the seven housekeeping genes by using *H. pylori* MLST Website (http://pubmlst.org/helicobacter/). They found that similarities of the concatenated MLST sequences of *H. pylori* strains in relation to the reference strains ranged from 94.6 to 97.0%, with no direct correlation between the number of strain-specific peptides and MLST sequence similarities for *H. pylori* strains. However, interspecies comparisons showed that the decrease in the number of strain-specific peptides was accompanied by a marked decrease also in MLST sequence similarities.

In conclusion, the FSP index provides a sensitive metric for measuring genomic relatedness between microorganisms that outperforms commonly used methods for quantifying genome conservation between microbial strains.

Genomic Interrelationships Among Unknown Strains Revealed by Shotgun Proteomics

Shotgun-proteomics analysis of strains isolated from clinical, food, or environmental matrices usually indicates that many DB strain proteomes could explain the determined PSMs. In general, this situation is analogous to a protein inference problem frequently encountered in bottom-up proteomics, although in this case we are focusing on "pseudo-polyproteins" representing strains instead of regular proteins. Therefore, there are two basic ways of finding the solution. The first approach is based on the parsimony principle and seeks to find the minimal list of DB strains that could explain all identified peptides (Tracz et al. 2013). The second approach is based on the creation of a maximal exploratory list of strains containing all DB strains matching at least one peptide; equivalent to selecting the whole matrix of PTB strain assignments (Dworzanski et al. 2010). However, the optimal solution could rely on using the "trimmed" matrix of PTB assignments, obtained by keeping only reference strains from the closest taxonomic units, for example, on the species, genus or family level.

In proteomics the most popular is the first approach, that is, the construction of minimal explanatory list of proteins and several tools, including ProteinProphet (Nesvizhskii et al. 2003) and IDPicker (Ma et al. 2009) are able to extract such lists automatically from the identified peptides. However, for a strain typing purposes all reference proteomes matching an isolate could be used as coordinates representing their similarities to an unknown strain (Dworzanski et al. 2010).

For example, let us assume for the sake of clarity that LC-MS/MS analysis of an unknown (U) strain s1 returned four confidently identified peptides p1 through p4 which were assigned to the closest DB neighbors represented by reference strains b_1–b_5, as shown schematically in Fig. 5.9.

As discussed in Section "Peptide-to-Taxa Assignments: Determination of the Closest Neighbor", the results of such PTB matches are arranged into the presence/absence assignment matrix and, in general, similarities between the analyzed

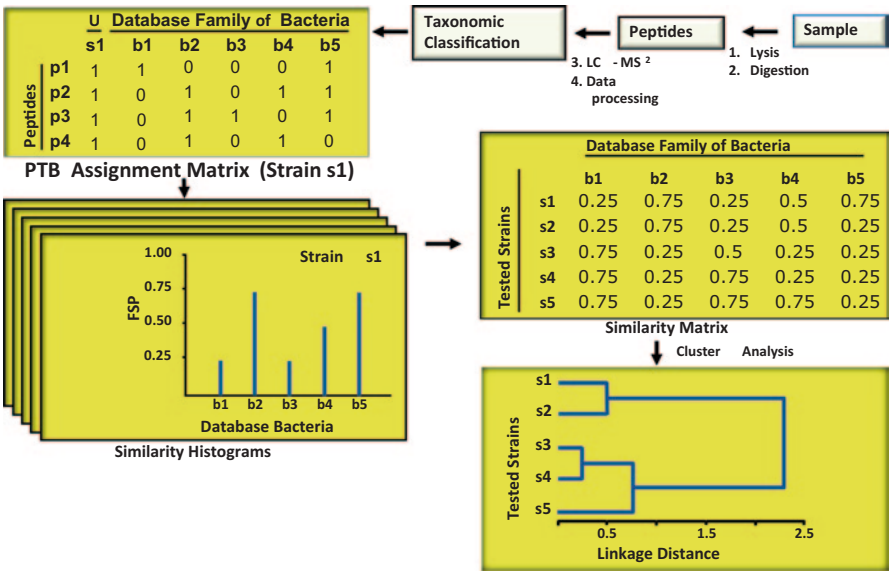


Fig. 5.9 Schematic representation of the sample and data processing workflow for shotgun-proteomics-based analysis of unknown strains (s1, s2, s3, s4, s5) revealing genomic inter-relationships among them. *U* unknown strain, *PTB* peptide sequence-to-bacterial strains assignments. (Reprinted with permission from Dworzanski et al. (2010, pp. 145–155). Copyright 2010 American Chemical Society)

microbial isolate and *n* reference strains $(b_1, b_2, b_3, \ldots, b_n)$ in the DB are measured as FSPs that may be presented as a similarity histogram. Moreover, these FSP indices are also considered as elements of a row vector representing that isolate in an *n*-dimensional vector space of reference strains. In Fig. 5.9 the similarity histogram for s1 indicates that this strain is not identical with any reference strain; however strains b_2 and b_5 are its closest relatives in this micro-DB, sharing 75 % of peptides (FSP = 3/4) while b_1 and b_3 are the least similar (FSP = 1/4). In the case of analyzing numerous isolates (e.g., strains s1–s5), each isolate is characterized by a set of FSP values which are elements of a row vector; and all such row vectors form a similarity matrix that can be analyzed using multivariable analysis methods, such as HCA to reveal genomic relatedness among unknown strains, for example, s1–s5.

Dworzanski et al. (2010) used this approach for phylogenomic analysis of isolates from poisonous food samples. The results of their analysis are shown as the upper diagram in Fig. 5.10 and are contrasted with a dendrogram obtained by cluster analysis of the DDH data, lower diagram, for the same strains.

The topologies of both dendrograms are very similar. Moreover, both trees closely resemble clusters and subclusters of strains revealed by HCA of concatenated nucleotide sequences of *gyrB* genes superimposed on both trees and marked as *gyrB* "Groups 1−3" to facilitate a three-way comparison of analyzed strain groupings. For instance, these topologies indicate that strains belonging to *gyrB* "Group 1" include *B. anthracis* Sterne and ten food isolates, and the same pattern was inferred from both DNA hybridization results and the proteomics data. As can be noted, two

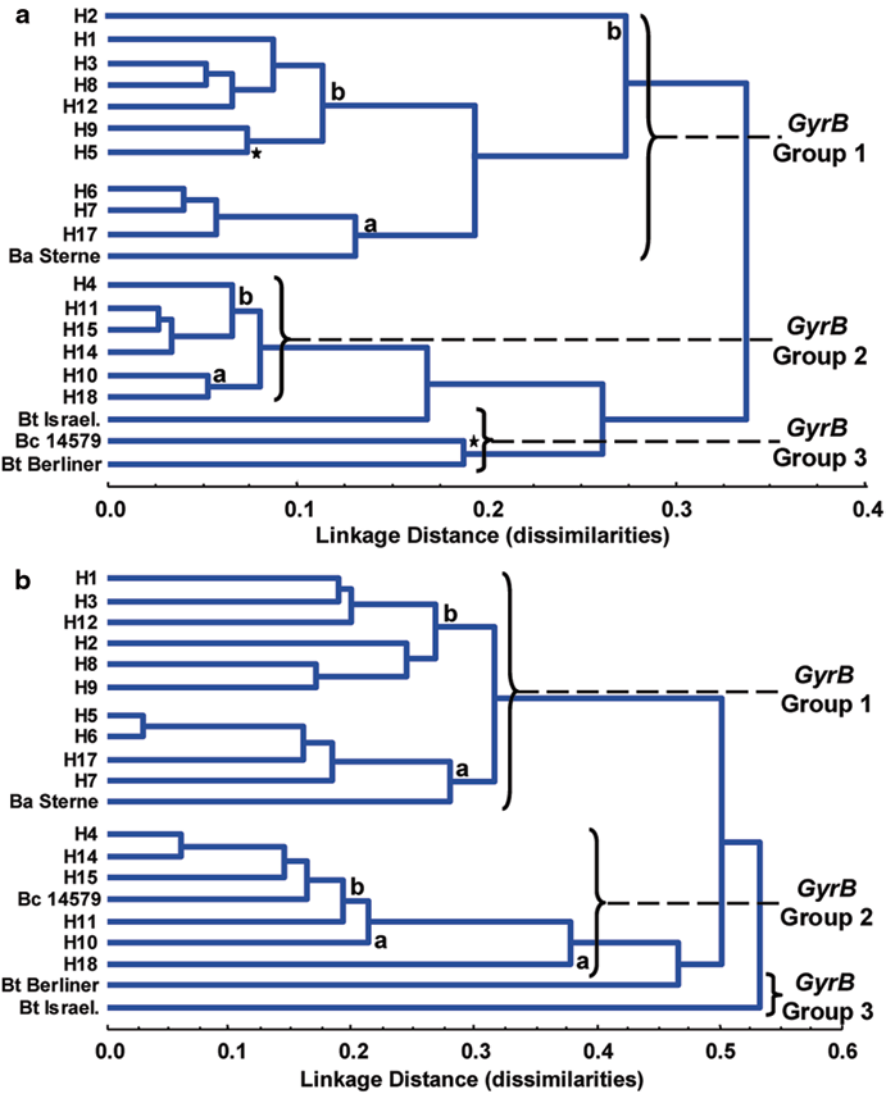


Fig. 5.10 Relatedness among *B. cereus* strains isolated from poisonous food samples (serotypes H1 through H18) and selected *Bacillus* type strains determined by hierarchical cluster analysis of distance matrices obtained from **(a)** proteomic and (**b)** DNA−DNA reassociation data. *gyrB* groups 1−3 stand for clusters of H-serotypes revealed by the analysis of concatenated nucleotide sequences of *gyrB* genes. (Reprinted with permission from Dworzanski et al. (2010, pp. 145–155). Copyright 2010 American Chemical Society

distinct subgroupings emerge from "Group 1." The subcluster marked as "a" indicates strains highly similar to *B. anthracis,* while the subcluster "b" agglomerates strains only moderately similar to this reference strain. It is interesting to note that serotypes H1, H3, and H12 of the "b" subcluster are known as cereulide-producing

strains. However, the comparison of strains grouped as members of *gyrB* subclusters "a" and "b" shows a biologically interesting disagreement between proteomics and DDH-based data. On the basis of proteomic similarities, strain H5 (marked with an asterisk in Fig. 5.10) was assigned to subcluster "b" while it was placed, together with serotype H9, into subcluster "a" on the basis of both *gyrB* sequences and hybridization values. Nevertheless, phylogenetic trees built using sequences of many housekeeping proteins and the *B. cereus* virulence factor sphingomyelinase indicate a substantial similarity of H5 with serotypes H3 and H12 and thus support findings revealed by proteomic similarities.

Overall, the obtained data indicate that proteomic similarities, DDH and *gyrB* sequencing provide very similar strain classification results, thus validating the proteomics-based approach developed by Dworzanski et al. (2010). Therefore, proteomic similarities expressed as FSP values could potentially replace DDH, as well as the gyrB or 16S rRNA sequencing in revealing phylogenomic affiliations and interrelationships among the *B. cereus* group.

Discrimination of Microbial Strains Based on Typing of Flagellin and Surface Layer Proteins

Flagellar filaments are composed of as many as 20,000 structural subunits of a 40–60 kDa protein flagellin—expressed by many bacteria, including pathogenic strains of *E. coli* and *Salmonella* spp.—and is characterized by highly variable sequences associated with the surface-exposed domains (also known as H antigens), and the conserved sequences that are crucial for filament assembly. These filaments are acting as propellers allowing cells to be motile and thus to respond to environmental stimuli; however, flagella may also contribute to bacterial pathogenicity and host immune responses (Ramos et al. 2004).

Typing of *E. coli* and *Salmonella* Strains Based on Flagellin H Antigen Sequences

Typing of *E. coli* Strains Antigenicity of flagellar H antigens and lipopolysaccharides (O antigens) were used for serotyping of *E. coli* strains for decades and this approach is widely adopted in classification of strains for taxonomic and epidemiological purposes. Moreover, serotyping based on the examination of 53 distinct H antigens is regarded as the gold standard for classification of isolates, especially during the investigation of outbreaks caused by *E. coli* pathogenic strains. However, serotyping of surface antigens is associated with some difficulties because on the one side, flagellum expression may depend on several environmental factors and on the other, diagnostic H-sera are not commercially available and therefore differ in quality. In addition, preparatory steps and serological protocols involved are laborious and lengthy because, in addition to multistep agglutination reactions, may

involve extra procedures, like motility induction. The procedures involved usually take a few days to complete, therefore, molecular methods capable of replacing or to support the serotyping have been developed. They take advantage of sequence polymorphism of flagellin encoding gene *fli* C (Prager et al. 2003) or its product, that is, flagellin H-antigen (Cheng et al. 2013) to provide clear cut classification with very good correlation to serotyping.

The shotgun-proteomics-based approach based on flagellin sequencing was recently reported by Cheng et al. (2013). In this approach, referred to as "MS-H," Cheng and co-workers isolated flagella and typed the *E. coli* H antigens by searching fragmentation spectra of flagellin tryptic peptides against a custom flagellin DB of 195 unique sequence entries representing all 53 known *E. coli* H serotypes (Section "Custom DBs of *E. coli* and S*almonella* Flagellins"). More importantly, they also developed a new procedure for flagella isolation and sample processing prior to LC-MS/MS analysis. This procedure includes a simplified workflow of vortexing bacterial cells to shear off flagella, combined with their isolation by filtration that is followed by on-filter trypsin digestion (see Section "Preparation of Flagella") and LC-MS/MS analysis. The H-serotype assignments to 41 clinical isolates of *E. coli* carried out by proteomics and serological methods showed that they were concordant in 92.7% of cases. Interestingly, the discrepancies included two strains which were untypeable by serological methods while the MS-H approach assigned their types as H7 and H21. One of these strains was previously typed as H7 and later became untypeable by agglutination, while the correctness of the second assignment (H21) to the sero-untypeable strain was confirmed by DNA sequencing of *fli* C.

The sequence coverage of flagellin depends on many factors, and one of them relates to the amount of flagellin digest used for the MS-H procedure. For example, LC-MS/MS analyses of 0.15 µg of flagellin from serotype O157:H7 with a quadrupole-TOF instrument were associated with 60% of sequence coverage which was increased up to 88% for a 7.5 µg sample. Therefore by replacing a quadrupole TOF instrument in the LC-MS/MS system with a higher resolution Orbitrap, they found that both diagnostic specificity and sensitivity parameters for MS-H method reached 100%. The example of complete concordance between serotyping and proteomics results obtained by searching MS data against a curated *E. coli* flagellin DB (Custom DB) is shown in Table 5.1. In addition, the comparison of top hits returned by searches against the custom and public DBs shows the superiority of using the curated DB for strain typing based on flagellin sequences. Consequently, Cheng et al. (2013) concluded that MS-H generates results much faster and with greater simplicity in comparison to antibody-based agglutination or primer-based PCR methods and pointed out that the MS-H method should be particularly useful during *E. coli* outbreak by providing rapid presumptive H-type classification of strains.

Typing of Salmonella Strains Cheng et al. (2014b) explored the MS-H platform also for typing *Salmonella* flagella by using the same sample preparation method as for *E. coli* samples (Section "Preparation of Flagella"), followed by LC-MS/MS analysis of peptides, and searching their fragmentation spectra against a curated *Salmonella* flagellum DB containing 385 entries. However, *Salmonella* flagellins

Table 5.1 Top hits produced by searching *E. coli* flagellin MS data against a curated *E. coli* flagellin custom DB and the public DBs: Swiss-prot and NCBI nr[a]. (Cheng et al. 2014a)

Strain number	Confirmed serotype	Custom DB (195 sequences) top hit	Swiss-prot (331,337 sequences) top hit	NCBInr (25,303,445 sequences) top hit
E169	H1	H1	*Shigella* flagellin	flagellin [*E. coli*]
E170	H2	H2	*E. coli* Elongation factor	flagellin [*E. coli*]
E171	H3	H3	*Salmonella* flagellin	flagellin [*E. coli*]
E172	H4	H4	*E. coli* K12 flagellin	flagellin [*E. coli*]
E173	H5	H5	*E. coli* K12 flagellin	*E. coli* flagellar protein FliC
E174	H6	H6	*Shigella* flagellin	FliC [*E. Coli*]
EDL933	H7	H7	*Shigella* flagellin	flagellin [*E. coli*]
E176	H8	H8	*Shigella* flagellin	flagellin [*E. coli*]
E177	H9	H9	*Shigella* flagellin	flagellin [*E. coli*]
E659	H10	H10	*E. coli* K12 flagellin	flagellin [*E. coli*]

[a] An Orbitrap system was used with 30 ppm peptide mass tolerance, 0.5 Da MS/MS tolerance, one missed tryptic cleavage for all DB searches. Oxidation on methionine and deamidation on glutamine and asparagine were chosen as a possible modification.

are more diversified in comparison to *E. coli,* therefore the Kauffmann-White-Le Minor serotyping scheme for designation of *Salmonella* serotypes recognizes 119 *Salmonella* flagellum H antigens composed of combinations of distinct antigenic factors. The antigenic portion of the *Salmonella* flagellar structure is encoded by two genes—*fliC* with homologs in other enteric bacteria and *Salmonella* specific *fljB—which encode two types of flagellins, known as phase 1 and phase flagellins, respectively.* Although diphasic cells express only one type of flagellar protein at a time, some serovars always express only one flagellar antigen and are considered monophasic (e.g., *S. enterica* subspecies IIIa, IV, VII and *Salmonella bongori*). Nevertheless, in rare instances *Salmonella* may be also triphasic by expressing one-third, plasmid-encoded flagellar H antigen, thus providing a mechanism for the generation of new serovars through the horizontal transfer and recombination of flagellin genes (Li et al. 1994; McQuiston et al. 2004). Flagellar antigens that are immunologically related are also known as "antigen complexes" and exhibit very similar sequences (Ranieri et al. 2013).

To validate the MS-H approach for typing *Salmonella* strains, Cheng et al. (2014b) analyzed 24 serovars from 43 strains that included 25 diphasic, one triphasic, and 17 monophasic isolates; and obtained identification results for the first strain in only a few hours after sample preparation from the culture based on sequence coverage and the associated identification confidence scores. They found that all 17 monophasic flagella were correctly and reproducibly identified, however, complications were noticed during the characterization of phase 2 factor 1 complexes (1,2; 1,5; 1,6; and 1,2,7) and phase 1 antigen groups ("r," "i," and "r, i") due to their extremely close sequence similarities (McQuiston et al. 2004). In addition,

a phase 3 antigen z49 of serovar Infantis (6,7:r:1,5:z49) was not identified because the z49 sequence was not available for comparison. Overall, for 25 diphasic strains, there was 75 % accuracy for phase 1 antigens and 69 % accuracy for unstable phase 2 antigens; however, the results were 100 % accurate at the antigen cluster/complex level (Cheng et al. 2014b). In conclusion, with the increasing number of sequenced flagellar genes, the resolution of the MS-H method for some diphasic strains should also be improved in the near future.

Typing of *Lactobacillus* Strains Based on Surface Layer (S-Layer) Protein Sequences

Cell envelopes in numerous bacteria and archaea are covered by a porous layer of proteins. Moreover, for the majority of bacteria this proteinaceous surface layer is de facto composed from numerous identical protein subunits with M_r in the range of 25–200 kDa, and with a copy numbers exceeding 5×10^5 subunits (Sleytr and Messner 1983), thus making them an attractive target for extraction (see Section "Surface Layer Proteins") and sequence-based discrimination of microbial strains.

S-layers have been found in numerous *Lactobacillus* species, such as *L. helveticus*, *L. brevis*, and the former *L. acidophilus* group, that is, *L. acidophilus, L. amylovorus, L. crispatus, and L. gallinarum.* Moreover, phylogenetic trees based on *Lactobacillus* S-layer protein sequences provide much better strain resolution than those constructed on the basis of 16S rRNA or the elongation factor Tu sequences (Hynönen and Palva 2013). Therefore, Podleśny et al. (2011) took advantage of these sequence differences between the S-layer proteins by using a proteomics-based approach to identify and type strains isolated from a Canadian dairy product. They also compared proteomics results with genomic data obtained by sequencing genes encoding 16S rRNA, the RNA polymerase alpha subunit *(rpoA),* phenylalanyl-tRNA synthase alpha subunit *(pheS),* translational elongation factor Tu *(tuf),* and Hsp60 chaperonins *(groEL)* and found them in full agreement. For instance, the sequence analysis of 16S rRNA gene from the isolated strain confirmed the affiliation of an isolate with the *Lactobacillus acidophilus* group bacteria, while the MLSA data revealed the close relationships with *L. helveticus* and *L. gallinarum.* However, the determination of the partial sequences for *pheS* and *groEL* showed higher similarity with *L. helveticus* (98 %) than with *L. gallinarum* (*phes*, 96 %, *groEL* 94 %). On the contrary to these lengthy genomic procedures, the nano-LC-linear quadrupole ion trap-Fourier transform ion cyclotron resonance (LTQ-FT-ICR) MS analysis of tryptic peptides from S-layer proteins combined with searching the NCBI nonredundant DB allowed not only for high confidence identification of the source organism as *L. helveticus,* but also for typing and strain rankings based on the number of matched peptides. These data placed "surface layer protein precursor" protein—encoded by the gene *slp*—from *L. helveticus* R0052 as the best match which suggests that this strain is the nearest neighbor among six *L. helveticus* strains available in the DB. Moreover, 53 unique peptides (71 % sequence coverage)

matched this surface protein from the strain R00052 while the number of matches to the remaining five *L. helveticus* strains (JCM1003, GCL1001, CP790, M4, and DPC4571) was in the range of only 14–22 peptides. This proteomics-based strain-level classification was finally validated by sequencing the *slp* gene encoding surface layer protein of the isolate and showing its 99.8 % sequence identity with the corresponding slp gene of *L. helveticus* R0052 (Podleśny et al. 2011).

In conclusion, LC-MS/MS analysis of surface layer proteins proved that the proteomics method is the appropriate molecular tool for the identification of S-layer-possessing lactobacilli at the subspecies level.

Discrimination of Strains Based on Antibiotic Resistance

The term "antibiotic resistance" implies that isolates are not inhibited by the usually achievable concentrations of a drug and may fall in the range where specific microbial resistance mechanisms are likely. In general, the resistance to a given antibiotic may be intrinsic or acquired. Therefore, the correct identification of a pathogen could be used to predict its intrinsic resistance as a naturally occurring trait characteristic for a given subspecies, species or genus. However, the conventional identification process provides no information about the acquired resistance derived either from genetic mutations or acquisition of foreign DNA from other bacteria and therefore it has to be determined experimentally by measuring the ability of an isolate to grow in the presence of commonly used antibiotics.

The automated systems for simultaneous microbial identification and antimicrobial susceptibility testing are commercially available. However, although the microbial identification may be performed in less than 1 h, for example, by MALDI-TOF-MS-based systems, the time of full panel antimicrobial susceptibility testing usually requires up to 24 h (Machen et al. 2014).

Although the antibiotic resistance could be detected by analysis of specific genes, the question remains: are these genes functional and will they be expressed? The bottom-up proteomics approach can easily address these issues by searching for specific proteins associated with antibiotic resistance (see Section "DBs of Virulence Factors, Toxins, and Antibiotic Resistance Determinants"). More importantly, the mass spectra acquired during proteomic analysis may be used to provide information both on strain identity and the expression of genes associated with antibiotic resistance.

For example, Chang et al. (2013) developed a rapid shotgun-proteomics method for the identification of β-lactam-resistant *A. baumannii* pathogenic strains based on searching a custom DB of resistance-associated proteins, referred to as "BRPDAB" (see Section "Creation/Correction of Microbial Protein DBs Through Re-sequencing and Analysis of Genomes"). They disrupted bacterial cells with a bead-beater homogenizer and processed the protein extract using a FASP method (see Section "Sample Digestion Strategies") combined with a 15-min long microwave-assisted

protein digestion with trypsin. The released peptides were analyzed using a nano-LC-ESI-MS/MS platform and the acquired fragmentation spectra were searched against the BRPDAB DB with SEQUEST.

They used data from shotgun-proteomics analyses of both multidrug resistant strain MDRAB1, and sensitive to antibiotics strains ATCC17978 and ATCC19606, to identify strain-specific peptides for *A. baumannii* which were added to the BRP-DB DB. By combining all the β-lactam resistance-related proteins and *A. baumannii* specific proteins in the same DB, they used the same search results both for the identification of *A. baumannii* and the evaluation of its antibiotic resistance potential. To validate this approach they analyzed 20 clinical isolates and found: (i) all of them correctly identified as *A. baumannii* strains; and (ii) all the 20 *A. baumannii* strains as potentially antibiotic resistant due to detection of at least two β-lactam-resistance associated proteins in each isolate. For example, all the clinical isolates expressed AmpC cephalosporinase, known as a strong antibiotic resistance enzyme that hydrolyzes most β-lactams, including penicillin, monobactam, and cephalosporins. Nineteen strains expressed carbapenem-associated resistance protein, while the *Acinetobacter*-derived cephalosporinase-53 and beta-lactamase OXA-69-like protein (named for its greater activity against oxacillin) were identified in extracts from 7 and 6 clinical isolates, respectively. Moreover, the entire procedure, including LC-MS/MS analysis and DB searching only requires 5–6 h to simultaneously identify *A. baumannii* strains and their antibiotic resistance mechanisms.

Overall, the shotgun-proteomics findings were consistent with the minimal inhibitory concentration (MIC) determination results because all 20 *A. baumannii* clinical isolates were found resistant to carbapenem, monobactam, cephalosporin, and to a combination treatment of penicillin and β-lactamase inhibitors. The results obtained demonstrate that by augmenting the custom DB with strain-specific unique peptide sequences, it is possible to obtain simultaneously both strain-level identification of *A. baumannii* clinical isolates and their antibiotic resistance mechanism information within 5–6 h. Therefore, the approach developed by Chang et al (2013) could be used for a rapid, sensitive, and specific detection of β-lactam-resistant strains of *A. baumannii*.

The bottom-up proteomic method based on CE-ESI-MS/MS of tryptic peptides was also used for the detection of a class of β-lactamases called carbapenemases in multidrug-resistant Gram-negative bacteria (*Klebsiella pneumoniae, E. coli*, and *Enterococcus cloacae*) from 27 clinical isolates (Fleurbaaij et al. 2014). For this purpose, bacteria harvested from liquid growth media, or even picked from single colonies were resuspended in 50% solution of trifluoroethanol in deionized water and lyzed by sonication, followed by protein reduction, alkylation, and the overnight digestion with trypsin. Data from MS analysis were searched against a custom DB composed of bacterial sequences downloaded from the Microbial Proteomic Resource at the University of Bergen Gade Institute Website (http://org.uib.no/prokaryotedb; de Souza et al. 2010) supplemented in-house with various β-lactamase sequences.

Overall, using a CE-ESI-MS/MS platform, Fleurbaaij et al. (2014) identified OXA-48 carbapenemase in 17 samples and demonstrated the *Klebsiella pneumoniae* carbapenemase (KPC) in 10 samples. Moreover, they found that some of these isolates also expressed a number of extended spectrum β-lactamases such as CTX (named for their greater activity against cefotaxime) which were co-expressed in 11 out of 17 OXA-48 positive strains. All these findings were confirmed by a battery of phenotypic and genomic tests (PCR-based test targeting carbapenemase; MIC analysis with meropenem, the phenotypic Hodge test).

However, they pointed out that in the case of PCR methods specific primers are needed, requiring a priori knowledge that may become problematic in case specific mutations occur in the corresponding target sequences. They also performed the MALDI-TOF MS-based ertapenem breakdown assay (Sparbier et al. 2012) with all clinical samples, and while KPC was easily detected with this method (10/10), they only correctly identified three out of 17 (3/17) OXA-48 producers.

Finally, Fleurbaaij et al. (2014) noticed that analysis of as little as 10 ng of a tryptic digest results in the identification of 300–500 unique peptides from 100 to 200 proteins. Therefore, it is obvious that the same analysis can reveal not only β-lactamase resistance but also the identity of bacterial species harboring the resistance phenotype.

It should be noted that although the antibiotic resistance in pathogenic bacteria can even cause death, the antibiotic resistance might be a useful property in case of probiotic strains used as prophylactic agents in the treatment of antibiotic-associated diarrhea. However, even probiotic strains should be free of transmissible genes that can cause the dissemination of antibiotic resistance to pathogenic bacteria and this way may reduce the therapeutic possibilities in infectious diseases. For example, Jacobsen et al. (2007) reported on in vivo transfer of wild-type AR plasmids from food strains of *Lactobacillus plantarum* to *Enterococcus faecalis* strain in the gastrointestinal tract of rats. This and other findings of acquired AR genes in isolates intended for probiotic or nutritional use highlight the importance of antimicrobial susceptibility testing in industrial laboratories for documenting the safety of commercial lactic acid bacteria in our food and the potential role of shotgun proteomics in this process (Klare et al. 2007; Gueimonde et al. 2010).

Concluding Remarks

Currently, the total number of prokaryotic genomes available in public DBs approaches 15,000 and exceeds the number of known species with validly published names (12,391); although, numerous taxa are still underrepresented in public DBs. However, species most important from the pathological, biotechnological, and epidemiological standpoint are represented by many strains, thus assuring a solid foundation for a growing use of bottom-up proteomics methods for the subspecies-level identification and typing of strains. For example, 964 and 150 genome sequences are available for *E. coli* and *B. cereus* strains, respectively. Therefore, the very large

and still-growing number of sequenced microbial genomes makes it likely that identical or very similar sequences from a given species have been investigated.

On the heels of this genomic revolution, bottom-up proteomics methods allow for comparison of microbial genomes through the lens of tens of thousands of peptide sequences, providing high coverage of predicted proteomes on a routine basis. Such comprehensive readout of sequence information from genes that are actually expressed can be used for subspecies identification and sequence-based typing of microbial strains not included in whole-genome DBs. In addition, in a fraction of time needed for the whole-genome sequencing, shotgun-proteomics methods may provide comparable depth of information about genomic-level relatedness among investigated strains, thus bridging the gap between the whole-genome sequencing and other genomic methods.

The principal factor motivating the implementation of shotgun-proteomics methods is a high-information-content output provided by this approach, in comparison to MALDI-TOF-based platforms, allowing not only for high-resolution strain-level identification through finding the nearest-neighbor strains in the DB and assessment of their relatedness, but also for a comprehensive analysis of proteomes.

Such analysis of microbial strain proteomes may be performed simultaneously with strain identification and used for the characterization of strain serological and biological properties affecting pathological potential or disease outcomes, which may be revealed by the identification of virulence and AR-associated proteins as biomarkers of high diagnostic and prognostic value. Therefore, in the era of high-throughput proteomics and online bioinformatics, rapid genome-based proteomic typing of infecting agents, and especially highly virulent and potentially antibiotic resistant resistance strains, holds promise for guiding proper clinical care and to prevent potential local or global outbreaks.

Bottom-up proteomics methods still need refinement of protocols, and improvements in the standardization and availability of bioinformatics tools for comprehensive data analysis on a routine basis. Although recent innovations in mass spectrometric instrumentation have accelerated the speed and sensitivity of proteome analysis (Hebert et al. 2014), further improvements can be obtained by emphasizing the optimization, simplification, and automation of sample preparation, for example, through single-tube proteomics approaches integrating all steps from cell lysis to peptide fractionation (Hughes et al. 2014; Fan et al. 2014), peptide separation techniques, and bioinformatics tools for fast, automated data interpretation for strain-level identification of cultivable bacteria and comprehensive characterization of each isolated microbial strain in the near future.

Acknowledgment The author thanks Dr. A Peter Snyder for fruitful collaborations during various phases of the research partially described here, for critical reading of the manuscript, constructive comments on this article, and checking infelicities of expression.

References

Ansong C, Yoon H, Porwollik S, et al. Global systems-level analysis of Hfq and SmpB deletion mutants in *Salmonella*: implications for virulence and global protein translation. PLoS ONE 2009;4(3):e4809. doi:10.1371/journal.pone.0004809.

Antharavally BS, Mallia KA, Rosenblatt MM, et al. Efficient removal of detergents from proteins and peptides in a spin column format. Anal Biochem. 2011;416:39–44. doi:10.1016/j.ab.2011.05.013.

Armengaud J, Hartmann EM, Bland C. Proteogenomics for environmental microbiology. Proteomics. 2013;13:2731–42. doi:10.1002/pmic.201200576.

Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000:25:25–9. doi:10.1038/75556.

Baeumlisberger D, Rohmer M, Arrey TN, et al. Simple dual-spotting procedure enhances nLC–MALDI MS/MS analysis of digests with less specific enzymes. J Proteome Res. 2011:10:2889–94. doi:10.1021/pr2001644.

Balážová T, Šedo O, Štefanić P, et al. Improvement in *Staphylococcus* and *Bacillus* strain differentiation by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling by using microwave-assisted enzymatic digestion. Rapid Commun Mass Spectrom. 2014;28:1855–61. doi:10.1002/rcm.6966.

Basile F, Hauser N. Rapid online nonenzymatic protein digestion combining microwave heating acid hydrolysis and electrochemical oxidation. Anal Chem. 2011;83:359–67. doi:10.1021/ac1024705.

Bereman MS, Egertson JD, MacCoss MJ. Comparison between procedures using SDS for shotgun proteomic analyses of complex samples. Proteomics. 2011:11:2931–35. doi:10.1002/pmic.201100045.

Bland C, Hartmann EM, Christie-Oleza JA, et al. N-terminal-oriented proteogenomics of the marine bacterium *Roseobacter denitrificans* OCh114 using N-Succinimidyloxycarbonylmethyl) tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP) labeling and diagonal chromatography. Mol Cell Proteomics. 2014;13:1369–81. doi:10.1074/mcp.O113.032854.

Bohlin J, Skjerve E, Ussery DW. Reliability and applications of statistical methods based on oligonucleotide frequencies in bacterial and archaeal genomes. BMC Genomics. 2008;9:104. doi:10.1186/1471-2164-9-104.

Boja ES, Fales HM. Overalkylation of a protein digest with iodoacetamide. Anal Chem. 2001;73:3576–82. doi:10.1021/ac0103423.

Bonissone S, Gupta N, Romine M, et al. N-terminal protein processing: a comparative proteogenomic analysis. Mol Cell Proteomics. 2013;12:14–28. doi:10.1074/mcp.M112.019075.

Caboche S, Audebert C, Lemoine Y, et al. Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data. BMC Genomics. 2014;15:264. doi:10.1186/1471-2164-15-264.

Caraux G, Pinloche S. PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics. 2005;21:1280–1. doi:10.1093/bioinformatics/bti141.

Cargile BJ, Bundy JL, Stephenson JL Jr Potential for false positive identifications from large databases through tandem mass spectrometry. J Proteome Res. 2004;3:1082–5. doi:10.1021/pr049946o.

Caro-Quintero A, Konstantinidis KT. Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. ISME J. 2015;9(4):958–67. doi:10.1038/ismej.2014.193.

Chalmers MJ, Gaskel SJ. Advances in mass spectrometry for proteome analysis. Curr Opin Biotechnol. 2000;11:384–90.

Chang CJ, Lin JH, Chang KC, et al. Diagnosis of β-lactam resistance in *Acinetobacter baumannii* using shotgun proteomics and LC-nano-electrospray ionization ion trap mass spectrometry. Anal Chem. 2013;85:2802–8. doi:10.1021/ac303326a.

Chatellier S, Mugnier N, Allard F, et al. Comparison of two approaches for the classification of 16S rRNA gene sequences. J Med Microbiol. 2014;63:1311–5. doi:10.1099/jmm.0.074377-0.

Chen LH, Xiong ZH, Sun LL, et al. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucl Acids Res. 2012;40:D641–5. doi:10.1093/nar/gkr989.

Cheng K, Drebot M, McCrea J, et al. MS-H: a novel proteomic approach to isolate and type the *E. coli* H antigen using membrane filtration and liquid chromatography-tandem mass spectrometry (LC-MS/MS). PLoS ONE. 2013;8:e57339. doi:10.1371/journal.pone.0057339.

Cheng K, Sloan A, McCorrister S, et al. Fit-for-purpose curated database application in mass spectrometry-based targeted protein identification and validation. BMC Res Notes. 2014a;7:444. doi:10.1186/1756-0500-7-444.

Cheng K, Sloan A, Meakin J, et al. Sequence-level and dual-phase identification of *Salmonella* flagellum antigens by liquid chromatography-tandem mass spectrometry (LC-MS/MS). J Clin Microbiol. 2014b;52:2189–92. doi:10.1128/JCM.00242-14.

Cole JR, Wang Q, Cardenas E, et al. The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucl Acids Res. 2009;37:D141–5. doi:10.1093/nar/gkp353.

Collins FS, Hamburg MA. First FDA authorization for next-generation sequencer. N Engl J Med. 2013;369:2369–71.

Cottrell JS. Protein identification using MS/MS data. J Proteomics. 2011;74:1842–51. doi:10.1016/j.jprot.2011.05.014.

Cox J, Neuhauser N, Michalski A, et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10:1794–1805. doi:10.1021/pr101065j.

Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004;20:1466–7. doi:10.1093/bioinformatics/bth092.

Damron FH, Napper J, Teter MA, et al. Lipotoxin F of *Pseudomonas aeruginosa* is an AlgU-dependent and alginate-independent outer membrane protein involved in resistance to oxidative stress and adhesion to A549 human lung epithelia. Microbiology. 2009;155:1028–38. doi:10.1099/mic.0.025833-0.

Deloger M, El Karoui M, Petit MA. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. J Bacteriol. 2009;191:91–9. doi:10.1128/JB.01202-08.

Demirev PA, Ho YP, Ryzhov V, et al. Microorganism identification by mass spectrometry and protein database searches. Anal Chem. 1999;71:2732–8. doi:10.1021/ac990165u.

Deshpande SV, Jabbour RE, Snyder PA, et al. ABOid: a software for automated identification and phyloproteomics classification of tandem mass spectrometric data. J Chromatograph Separat Techniq. 2011;S5:001. doi:10.4172/2157-7064.S5-001.

de Souza GA, Arntzen MØ, Wiker HG. MSMSpdbb: providing protein databases of closely related organisms to improve proteomic characterization of prokaryotic microbes. Bioinformatics. 2010;26:698–9. doi:10.1093/bioinformatics/btq004.

de Souza GA, Arntzen MØ, Fortuin S, et al. Proteogenomic analysis of polymorphisms and gene annotation divergences in prokaryotes using a clustered mass spectrometry-friendly database. Mol Cell Proteomics. 2011;10:M110-002527. doi:10.1074/mcp.M110.002527.

Deutsch EW, Mendoza L, Shteynberg D, et al. A guided tour of the trans-proteomic pipeline. Proteomics. 2010;10:1150–9. doi:10.1002/pmic.200900375.

Dworzanski JP, Snyder AP, Chen R, et al. Identification of bacteria using tandem mass spectrometry combined with a proteome database and statistical scoring. Anal Chem. 2004;7:2355–66. doi:10.1021/ac0349781.

Dworzanski JP, Deshpande SV, Chen R, et al. Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. J Proteome Res. 2006;5:76–87. doi:10.1021/pr050294t.

Dworzanski JP, Dickinson DN, Deshpande SV, et al. Discrimination and phylogenomic classification of *Bacillus anthracis-cereus-thuringiensis* strains based on LC-MS/MS analysis of whole cell protein digests. Anal Chem. 2010;82:145–55. doi:10.1021/ac9015648.

Falb M, Aivaliotis M, Garcia-Rizo C, et al. Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey. J Mol Biol. 2006;362:915–24. doi:10.1016/j.jmb.2006.07.086.

Enany S, Yoshida Y, Yamamoto T. Exploring extra-cellular proteins in methicillin susceptible and methicillin resistant *Staphylococcus aureus* by liquid chromatography–tandem mass spectrometry. World J Microbiol Biotechnol. 2014;30:1269–83. doi:10.1007/s11274-013-1550-7.

Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Amer Soc Mass Spectrom. 1994;5:976–89. doi:10.1016/1044-0305(94)80016-2.

Erde J, Ogorzalek Loo RR, Loo JA. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. J Proteome Res. 2014;13:1885–95. doi:10.1021/pr4010019.

Fan C, Shi Z, Pan Y, et al. Dual matrix-based immobilized trypsin for complementary proteolytic digestion and fast proteomics analysis with higher protein sequence coverage. Anal Chem. 2014;86:1452–8. doi:10.1021/ac402696b.

Federhen S. The NCBI Taxonomy database. Nucl Acids Res. 2012;40:D136–43. doi:10.1093/nar/gkr1178.

Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev. 2001;20:157–71. doi:10.1002/mas.10004.

Fenselau C, Laine O, Swatkoski S. Microwave assisted acid cleavage for denaturation and proteolysis of intact human adenovirus. Internat J Mass Spectrom. 2011;301:7–11.doi:10.1016/j.ijms.2010.05.026.

Fernández-Puente P, Mateos J, Blanco FJ, et al. LC-MALDI-TOF/TOF for shotgun proteomics. In: Martins-de-Souza D editor. Shotgun proteomics: methods and protocols, methods in molecular biology. Vol. 1156. New York: Springer; 2014. p 27–38. doi:10.1007/978-1-4939-0685-7_2.

Fleurbaaij F, Heemskerk AA, Russcher A, et al. Capillary-electrophoresis mass spectrometry for the detection of carbapenemases in (multi-) drug-resistant Gram-negative bacteria. Anal Chem. 2014;86:9154–61. doi:10.1021/ac502049p.

Forslund AL, Kuoppa K, Svensson K, et al. Direct repeat-mediated deletion of a type IV pilin gene results in major virulence attenuation of *Francisella tularensis*. Mol Microbiol. 2006;59:1818–30. doi:10.1111/j.1365-2958.2006.05061.x.

François P, Scherl A, Hochstrasser D, et al. Proteomic Approach to Investigate Pathogenicity and Metabolism of Methicillin-Resistant Staphylococcus aureus. In:Yindo Ji (ed.), Methicillin-resistant *Staphylococcus aureus* (MRSA) protocols, Methods in Molecular Biology. Vol. 1085. New York: Springer; 2014. p. 231–50. doi:10.1007/978-1-62703-664-1_14.

Fröhlich T, Arnold GJ. A newcomer's guide to nano-liquid-chromatography of peptides. In: Reinders J, Sickmann A, editors. Proteomics, methods in molecular biology. Vol. 564. Springer, Heidelberg; 2009. p. 123–41. doi:10.1007/978-1-60761-157-8_7.

Geer LY, Markey SP, Kowalak JA, et al. Open mass spectrometry search algorithm. J Proteome Res. 2004;3:958–64. doi:10.1021/pr0499491.

Geiser L, Dayon L, Vaezzadeh AR, et al. Shotgun proteomics: a relative quantitative approach using Off-Gel electrophoresis and LC-MS/MS. In: Walls D, Loughran ST, editors. Protein chromatography: methods and protocols, methods in molecular biology. Vol. 681. Springer, Heidelberg; 2011a. pp 459–72. doi:10.1007/978-1-60761-913-0_27.

Geiser L, Vaezzadeh AR, Deshusses JM, et al. Shotgun proteomics: a qualitative approach applying isoelectric focusing on immobilized pH gradient and LC-MS/MS. In: Walls D, Loughran ST, editors. Protein chromatography: methods and protocols, methods in molecular biology. Vol. 681. Springer, Heidelberg; 2011b. pp 449–58. doi:10.1007/978-1-60761-913-0_26.

Glatter T, Ludwig C, Ahrné E, et al. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J Proteome Res. 2012;11:5145–56. doi:10.1021/pr300273g.

Goh YJ, Azcarate-Peril MA, O'Flaherty S, et al. Development and application of a *upp*-based counterselective gene replacement system for the study of the S-layer protein SlpX of *Lactobacillus acidophilus* NCFM. Appl Environ Microbiol. 2009;75:3093–105. doi:10.1128/AEM.02502-08.

Goris J, Konstantinidis KT, Klappenbach JA, et al. DNA–DNA hybridization values and their relation to whole genome sequence similarities. Int J Syst Evol Microbiol. 2007;57:81–91. doi:10.1099/ijs.0.64483-0.

Granholm V, Käll L. Quality assessments of peptide–spectrum matches in shotgun proteomics. Proteomics. 2011;11:1086–93. doi:10.1002/pmic.201000432.

Gueimonde M, Flórez AB, van Hoek AHAM, et al. Genetic basis of tetracycline resistance in *Bifidobacterium animalis* subsp. *Lactis*. Appl Environ Microbiol. 2010;76:3364–9. doi:10.1128/AEM.03096-09.

Gundry RL, White MY, Murray CI, et al. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. Curr Protoc Mol Biol. 2009;88:10.25.1–10.25.23. doi:10.1002/0471142727.mb1025s88.

Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014;58:212–20. doi:10.1128/AAC.01310-13.

Halbedel S, Reiss S, Hahn B, et al. A systematic proteomic analysis of *Listeria monocytogenes* house-keeping protein secretion systems. Mol Cell Proteomics. 2014;13:3063–81. doi:10.1074/mcp.M114.041327.

Hartmann EM, Allain F, Gaillard JC, et al. Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. In: Vergunst AC, O'Callaghan D, editors. Host-bacteria interactions, methods in molecular biology. Vol. 1197. New York: Springer, 2014. p. 275–85. doi:10.1007/978-1-4939-1261-2_16.

Hebert AS, Richards AL, Bailey DJ, et al. The one hour yeast proteome. Mol Cell Proteomics. 2014;13:339–47. doi:10.1074/mcp.M113.034769.

Hendrickson EL, Beck DA, Wang T, et al. Expressed genome of *Methylobacillus flagellatus* as defined through comprehensive proteomics and new insights into methylotrophy. J Bacteriol. 2010;192:4859–67.doi:10.1128/JB.00512-10.

Hesketh AR, Chandra G, Shaw AD, et al. Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*. Mol Microbiol. 2002;46:917–32. doi:10.1046/j.1365-2958.2002.03219.x.

Hu A, Tsai PJ, Ho YP. Identification of microbial mixtures by capillary electrophoresis/selective tandem mass spectrometry. Anal Chem. 2005;77:1488–95.

Hu A, Chen CT, Tsai PJ, et al. Using capillary electrophoresis-selective tandem mass spectrometry to identify pathogens in clinical samples. Anal Chem. 2006;78:5124–33.

Hughes CS, Foehr S, Garfield DA, et al. Ultrasensitive proteome analysis using paramagnetic bead technology. Mol Syst Biol. 2014;10:757. doi:10.15252/msb.20145625.

Huson DH, Auch AF, Qi J, et al. MEGAN analysis of metagenomic data. Genome Res. 2007;17:377–86. doi:10.1101/gr.5969107.

Hynönen U, Palva A. *Lactobacillus* surface layer proteins: structure, function and applications. Appl Microbiol Biotechnol. 2013;97:5225–43. doi:10.1007/s00253-013-4962-2.

Jabbour RE, Dworzanski JP, Deshpande SV, Wick CH, Zulich AW. Effect of microbial sample processing conditions on bacterial identification using mass spectrometry-based proteomics approach. Proceedings of the 55th Conference of the American Society for Mass Spectrometry, Indianapolis, IN, June 3–7, 2007.

Jabbour RE, Deshpande SV, Wade MM, et al. Double-blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics. Appl Environ Microbiol. 2010a;76:3637–44. doi:10.1128/AEM.00055-10.

Jabbour RE, Wade MM, Deshpande SV, et al. Identification of *Yersinia pestis* and *Escherichia coli* strains by whole cell and outer membrane protein extracts with mass spectrometry-based proteomics. J Proteome Res. 2010b;9:3647–55. doi:10.1021/pr100402y.

Jabbour RE, Deshpande SV, Stanford MF, et al. A protein processing filter method for bacterial identification by mass spectrometry-based proteomics. J Proteome Res. 2010c;10:907–12. doi:10.1021/pr101086a.

Jabbour RE, Deshpande SV, McCubbin PE, et al. Extracellular protein biomarkers for the characterization of enterohemorrhagic and enteroaggregative *Escherichia coli* strains. J Microbiol Methods. 2014;98:76–83. doi:10.1016/j.mimet.2013.12.017.

Jacobsen L, Wilcks A, Hammer K, et al. Horizontal transfer of *tet*(M) and *erm*(B) resistance plasmids from food strains of *Lactobacillus plantarum* to *Enterococcus faecalis* JH2-2 in the gastrointestinal tract of gnotobiotic rats. FEMS Microbiology Ecology. 2007;59:158–166. doi:10.1111/j.1574-6941.2006.00212.x.

Jaffe JD, Berg HC, Church GM. Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics. 2004;4:59–77. doi:10.1002/pmic.200300511.

Jagtap P, Goslinga J, Kooren JA, et al. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. Proteomics. 2013;13:1352–7. doi:10.1002/pmic.201200352.

Janini GM, Zhou M, Yu LR, et al. On-column sample enrichment for capillary electrophoresis sheathless electrospray ionization mass spectrometry: evaluation for peptide analysis and protein identification. Anal Chem. 2003;75:5984–93. doi:10.1021/ac0301548.

Johnson B, Selle K, O'Flaherty S, et al. Identification of extracellular surface-layer associated proteins *in Lactobacillus acidophilus* NCFM. Microbiology. 2013;159:2269–82.doi:10.1099/mic.0.070755-0.

Käll L, Canterbury JD, Weston J, et al. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 2007a;4:923–5. doi:10.1038/nmeth1113.

Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucl Acids Res. 2007b;35 Suppl 2:W429–32. doi:10.1093/nar/gkm256.

Karlsson R, Davidson M, Svensson-Stadler L, et al. Strain-level typing and identification of bacteria using mass spectrometry-based proteomics. J Proteome Res. 2012;11:2710–20. doi:10.1021/pr2010633.

Keller A, Nesvizhskii AI, Kolker E, et al. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem. 2002;74:5383–92. doi:10.1021/ac025747h.

Kil YJ, Becker C, Sandoval W, et al. Preview: a program for surveying shotgun proteomics tandem mass spectrometry data. Anal Chem. 2011;83:5259–67. doi:10.1021/ac200609a.

Kimura M. The rate of molecular evolution considered from the standpoint of population genetics. Proc Natl Acad Sci U S A. 1969;63:1181–8.

Klare I, Konstabel C, Werner G, et al. Antimicrobial susceptibilities of *Lactobacillus, Pediococcus* and *Lactococcus* human isolates and cultures intended for probiotic or nutritional use. J Antimicrob Chemother. 2007;59:900–12. doi:10.1093/jac/dkm035.

Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005a;102:2567–72. doi:10.1073/pnas.0409727102.

Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol. 2005b;187:6258–64. doi:10.1128/JB.187.18.6258-6264.2005.

Koskinen VR, Emery PA, Creasy DM, et al. Hierarchical clustering of shotgun proteomics data. Mol Cell Proteomics. 2011;10(6):M110.003822. doi:10.1074/mcp.M110.003822.

Lam MPY, Law CH, Quan Q, et al. Fully automatable multidimensional reversed-phase liquid chromatography with online tandem mass spectrometry. In: Martins-de-Souza D editor. Shotgun proteomics: methods and protocols, methods in molecular biology. Vol. 1156. New York; Springer; 2014. p. 39–51. doi:10.1007/978-1-4939-0685-7_3.

Lasaosa M, Delmotte N, Huber CG, et al. A 2D reversed-phase x ion-pair reversed-phase HPLC-MALDI TOF/TOF-MS approach for shotgun proteome analysis. Anal Bioanal Chem. 2009;393:1245–56. doi:10.1007/s00216-008-2539-1.

Lee J-G, Cheong KH, Huh N, et al. Microchip-based one step DNA extraction and real-time PCR in one chamber for rapid pathogen identification. Lab Chip. 2006:886–95. doi:10.1039/B515876A.

Li J, Nelson K, McWhorter AC, et al. Recombinational basis of serovar diversity in *Salmonella enterica*. Proc Natl Acad Sci U S A. 1994;91:2552–6.

Lin SS, Wu CH, Sun MC, et al. Microwave-assisted enzyme-catalyzed reactions in various solvent system. J Am Soc Mass Spectrom. 2005;16:581–8. doi:10.1016/j.jasms.2005.01.012.

Lindgren H, Honn M, Golovlev I, et al. The 58-kilodalton major virulence factor of *Francisella tularensis* is required for efficient utilization of iron. Infect Immun. 2009;77:4429–36. doi:10.1128/IAI.00702-09.

Lippincott J, Apostol I. Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea. Anal Biochem. 1999;267:57–64. doi:10.1006/abio.1998.2970.

Liu B, Pop M. ARDB-antibiotic resistance genes database. Nucl Acids Res. 2009;37:D443–7. doi:10.1093/nar/gkn656.

Lohrig K, Wolters D. Multidimensional protein identification technology. In: Reinders J, Sickmann A, editors. Proteomics, methods in molecular biology. Vol. 564. Heidelberg: Springer; 2009. p. 143–53. doi:10.1007/978-1-60761-157-8_8.

Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30:434–9. doi:10.1038/nbt.2198.

Ma B, Johnson R. *De novo* sequencing and homology searching. Mol Cell Proteomics. 2012;11(2):O111.014902. doi:10.1074/mcp.O111.014902.

Ma Z, Dasari S, Chambers MC, et al. IDPicker 2.0: improved protein assembly with high discrimination peptide identification filtering. J Proteome Res. 2009;8:3872–81. doi:10.1021/pr900360j.

Machen A, Drake T, Wang YF. Same day identification and full panel antimicrobial susceptibility testing of bacteria from positive blood culture bottles made possible by a combined lysis-filtration method with MALDI-TOF VITEK mass spectrometry and the VITEK2 system. PLoS ONE. 2014;9(2):e87870. doi:10.1371/journal.pone.0087870.

Manza LL, Stamer SL, Ham AJL, et al. Sample preparation and digestion for proteomic analyses using spin filters. Proteomics. 2005;5:1742–5.

Masuda T, Tomita M, Ishihama Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. J Proteome Res. 2008;7:731–40. doi:10.1021/pr700658q.

Masuda T, Saito N, Tomita M, et al. Unbiased quantitation of *Escherichia coli* membrane proteome using phase transfer surfactants. Mol Cell Proteomics. 2009;8:2770–7. doi:10.1074/mcp.M900240-MCP200.

Mayne J, Starr AE, Ning Z, et al. Fine tuning of proteomic technologies to improve biological findings: advancements in 2011–2013. Anal Chem 2014;86:176–95. doi:10.1021/ac403551f.

McQuiston JR, Parrenas R, Ortiz-Rivera M, et al. Sequencing and comparative analysis of flagellin genes *fliC, fljB,* and *flap* from *Salmonella*. J Clin Microbiol. 2004;42:1923–32. doi:10.1128/JCM.42.5.1923-1932.2004.

Meacham F, Boffelli D, Dhahbi J, et al. Identification and correction of systematic error in high-throughput sequence data. BMC Bioinformatics. 2011;12:451. doi:10.1186/1471-2105-12-451.

Meier-Kolthoff JP, Auch AF, Klenk HP, et al. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14:60. doi:10.1186/1471-2105-14-60.

Mesuere B, Devreese B, Debyser G, et al. Unipept: tryptic peptide-based biodiversity analysis of metaproteome samples. J Proteome Res. 2012;11:5773–80. doi:10.1021/pr300576s.

Miller JM. Whole-genome mapping: a new paradigm in strain-typing technology. J Clin Microbiol. 2013;51:1066–70. doi:10.1128/JCM.00093-13.

Nan L, Jiang Z, Wei X. Emerging microfluidic devices for cell lysis: a review. Lab Chip. 2014;14:1060–73. doi:10.1039/c3lc51133b.

Napoli A, Aiello D, Aiello G, et al. Mass Spectrometry-based proteomic approach in *Oenococcus oeni* (*O. oeni*) enological starter. J Proteome Res. 2014;13:2856–66. doi:10.1021/pr4012798.

Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. J Proteomics. 2010;73:2092–123. doi:10.1016/j.jprot.2010.08.009.

Nesvizhskii AI, Keller A, Kolker E, et al. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75:4646–58. doi:10.1021/ac0341261.

Olaya-Abril A, Jimenez-Munguia I, Gomez-Gascon L, et al. Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. J Proteomics. 2013;97:164–76. doi:10.1016/j.jprot.2013.03.035.

Perkins DN, Pappin DJC, Creasy DM, et al. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20:3551–67.

Petersen TN, Brunak S, von Heijne G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6. doi:10.1038/nmeth.1701.

Podlesny M, Jarocki P, Komon E, et al. LC-MS/MS analysis of surface layer proteins as a useful method for the identification of lactobacilli from the *Lactobacillus acidophilus* group. J Microbiol Biotechnol. 2011;21:421–9. doi:10.4014/jmb.1009.09036.

Prager R, Strutz U, Fruth A, et al. Subtyping of pathogenic *Escherichia coli* strains using flagellar (H)-antigens: serotyping versus fliC polymorphisms. Int J Med Microbiol. 2003;292:477–86. doi:10.1078/1438-4221-00226.

Proc JL, Kuzyk MA, Hardie DB, et al. A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. J Proteome Res. 2010;9:5422–37. doi:10.1021/pr100656u.

Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl Acids Res. 2013;41:D590–6. doi:10.1093/nar/gks1219.

Ramos HC, Rumbo M, Sirard JC. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. 2004;12:509–17. doi:10.1016/j.tim.2004.09.002.

Ranieri ML, Shi C, Switt AIM, et al. Comparison of typing methods with a new procedure based on sequence characterization for *Salmonella* serovar prediction. J Clin Microbiol. 2013;51:1786–97. doi:10.1128/JCM.03201-12.

Reddy PM, Huang YS, Chen CT, et al. Evaluating the potential nonthermal microwave effects of microwave-assisted proteolytic reactions. J Proteomics. 2013;80:160–70. doi:10.1016/j.jprot.2013.01.005.

Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009;106:19126–31. doi:10.1073/pnas.0906412106.

Salomonsson E, Kuoppa K, Forslund AL, et al. Reintroduction of two deleted virulence loci restores full virulence to the live vaccine strain of *Francisella tularensis*. Infect Immun. 2009;77:3424–31. doi:10.1128/IAI.00196-09.

Salzberg SL, Delcher AL, Kasif S, et al. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 1998;26:544–8. doi:10.1093/nar/26.2.544.

Sanger F. Chemistry of insulin. Science. 1959;129:1340–4. doi:10.1126/science.129.3359.1340.

Sato N, Tajima N. Statistics of N-terminal alignment as a guide for refining prokaryotic gene annotation. Genomics. 2012;99:138–43. doi:10.1016/j.ygeno.2011.12.004.

Shteynberg D, Deutsch EW, Lam H, et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol Cell Proteomics. 2011;10(12):M111.007690. doi:10.1074/mcp.M111.007690.

Shteynberg D, Nesvizhskii AI, Moritz RL, et al. Combining results of multiple search engines in proteomics. Mol Cell Proteomics. 2013;12:2383–93. doi:10.1074/mcp.R113.027797.

Sleytr UB, Messner P. Crystalline surface layers on bacteria. Annu Rev Microbiol. 1983;37:311–39.

Sparbier K, Schubert S, Weller U, et al. Matrix-assisted laser desorption ionization–time of flight mass spectrometry-based functional assay for rapid detection of resistance against β-lactam antibiotics. J Clin Microbiol. 2012;50:927–37. doi:10.1128/JCM.05737-11.

Spivak M, Weston J, Bottou L, et al. Improvements to the percolator algorithm for peptide identification from shotgun proteomics data sets. J Proteome Res. 2009;8:3737–45. doi:10.1021/pr801109k.

Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol. 1994;44:846–9. doi:10.1099/00207713-44-4-846.

Steiner DJ, Furuya Y, Metzger DW. Host–pathogen interactions and immune evasion strategies in *Francisella tularensis* pathogenicity. Infect Drug Resist. 2014;7:239–51. doi:10.2147/idr.s53700.

Sun L, Jin M, Ding W, et al. Posttranslational modification of flagellin FlaB in *Shewanella oneidensis*. J Bacteriol. 2013;195:2550–61. doi:10.1128/JB.00015-13.

Sun L, Dong Y, Shi M, et al. Two residues predominantly dictate functional difference in motility between *Shewanella oneidensis* flagellins flaA and flaB. J Biol Chem. 2014;289:14547–59. doi:10.1074/jbc.M114.552000.

Switzar L, Giera M, Niessen WM. Protein digestion: an overview of the available techniques and recent developments. J Proteome Res. 2013a;12:1067–77. doi:10.1021/pr301201x.

Switzar L, van Angeren J, Pinkse M, et al. A high-throughput sample preparation method for cellular proteomics using 96-well filter plates. Proteomics. 2013b;13:2980–3. doi:10.1002/pmic.201300080.

Tabb DL, Fernando CG, Chambers MC. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome Res. 2007;6:654–61.

Tanca A, Biosa G, Pagnozzi D, et al. Comparison of detergent-based sample preparation workflows for LTQ-Orbitrap analysis of the *Escherichia coli* proteome. Proteomics. 2013;13:2597–607. doi:10.1002/pmic.201200478.

Tonella L, Hoogland C, Binz PA, et al. New perspectives in the *Escherichia coli* proteome investigation. Proteomics. 2001;1:409–23. doi:10.1002/1615-9861(200103).

Tracz DM, McCorrister SJ, Chong PM, et al. A simple shotgun proteomics method for rapid bacterial identification. J Microbiol Methods. 2013;94:54–7. doi:10.1016/j.mimet.2013.04.008.

Turse JE, Marshall MJ, Fredrickson JK, et al. An empirical strategy for characterizing bacterial proteomes across species in the absence of genomic sequences. PLoS ONE. 2010;5(11):e13968. doi:10.1371/journal.pone.0013968.

Vaezzadeh AR, Deshusses JM, Waridel P, et al. Accelerated digestion for high-throughput proteomics analysis of whole bacterial proteomes. J Microbiol Methods. 2010;80:56–62. doi:10.1016/j.mimet.2009.10.019.

Vaudel M, Venne AS, Berven FS, et al. Shedding light on black boxes in protein identification. Proteomics. 2014;14:1001–5. doi:10.1002/pmic.201300488.

Vuckovic D, Dagley LF, Purcell A, et al. Membrane proteomics by high performance liquid chromatography–tandem mass spectrometry: analytical approaches and challenges. Proteomics. 2013;13:404–23. doi:10.1002/pmic.201200340.

Waas M, Bhattacharya S, Chuppa S, et al. Combine and conquer: surfactants, solvents, and chaotropes for robust mass spectrometry based analyses of membrane proteins. Anal Chem. 2014;86:1551–9. doi:10.1021/ac403185a.

Wade MM, Zulich AW, Wick CH, et al. Discrimination of pathogenic versus non-pathogenic *Yersinia pestis* and *Escherichia coli* using proteomics mass spectrometry (No. ECBC-TR-771). Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD. 2010. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA522639. Accessed 7 Nov 2014.

Wade MM, Wick CH, Zulich AW, et al. Discrimination of pathogenic vs. nonpathogenic *Francisella tularensis* and *Burkholderia pseudomallei* using proteomics mass spectrometry (No. ECBC-TR-857). Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD. 2011. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA540823. Accessed 7 Nov 2014.

Wang H, Qian WJ, Mottaz HM, et al. Development and evaluation of a micro- and nanoscale proteomic sample preparation method. Proteome Res. 2005;4:2397–403. doi:10.1021/pr050160f.

Wang X, Slebos RJ, Wang D, et al. Protein identification using customized protein sequence databases derived from RNA-Seq data. J Proteome Res. 2012;11:1009–17. doi:10.1021/pr200766z.

Warscheid B, Fenselau C. Characterization of *Bacillus* spore species and their mixtures using postsource decay with a curved-field reflectron. Anal Chem. 2003;75:5618–27.

Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242–7.

Wayne LG, Brenner DJ, Colwell RR, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol. 1987;37:463–4. doi:10.1099/00207713-37-4-463.

Winnenburg R, Urban M, Beacham A, et al. PHI-base update: additions to the pathogen-host interaction database. Nucl Acids Res. 2008;36:D572–6. doi:10.1093/nar/gkm858.

Wiśniewski JR, Mann M. Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. Anal Chem. 2012;84:2631–7. doi:10.1021/ac300006b.

Wiśniewski JR, Rakus D. Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the *Escherichia coli* proteome. J Proteomics. 2014;10:322–31. doi:10.1016/j.dib.2014.08.004.

Wiśniewski JR, Zougman A, Nagaraj N, et al. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6:359–62.

Wiśniewski JR, Zielinska DF, Mann M. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. Anal Biochem. 2011;410:307–9. doi:10.1016/j.ab.2010.12.004.

Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001;73:5683–90.

Wu F, Sun D, Wang N, et al. Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis. Anal Chim Acta. 2011;698:36–43. doi:10.1016/j.aca.2011.04.039.

Wu X, Xu L, Gu W, et al. Iterative genome correction largely improves proteomic analysis of non-model organisms. J Proteome Res. 2014;13:2724–34. doi:dx.doi.org/10.1021/pr500369b.

Yang H, Zubarev RA. Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides. Electrophoresis. 2010;31:1764–72. doi:10.1002/elps.201000027.

Yang Y, Zhang S, Howe K, et al. A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics. J Biomol Tech. 2007;18:226–37.

Yates JR. Mass spectrometry and the age of the proteome. J Mass Spectrom. 1998;33:1–19.

Yeung YG, Nieves E, Angeletti RH, Stanley ER. Removal of detergents from protein digests for mass spectrometry analysis. Anal Biochem. 2008;382:135–7. doi:10.1016/j.ab.2008.07.034.

Yu Y, Xie L, Gunawardena HP, et al. GOFAST: an integrated approach for efficient and comprehensive membrane proteome analysis. Anal Chem. 2012;84:9008–14. doi:10.1021/ac300134e.

Zhang G, Fedyunin I, Kirchner S, et al. FANSe: an accurate algorithm for quantitative mapping of large scale sequencing reads. Nucl Acids Res. 2012;40(11):e83. doi:10.1093/nar/gks196.

Zhang CX, Creskey MC, Cyr TD, et al. Proteomic identification of *Listeria monocytogenes* surface-associated proteins. Proteomics. 2013a;13:3040–5. doi:10.1002/pmic.201200449.

Zhang K, Zheng S, Yang JS, et al. Comprehensive profiling of protein lysine acetylation in *Escherichia coli*. J Proteome Res. 2013b;12:844–51. doi:10.1021/pr300912q.

Zhou CE, Smith J, Lam M, et al. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucl Acids Res. 2007;35:D391–4. doi:10.1093/nar/gkl791.

Zhou JY, Dann GP, Shi T, et al. Simple sodium dodecyl sulfate-assisted sample preparation method for LC-MS-based proteomics applications. Anal Chem. 2012;84:2862–7. doi:10.1021/ac203394r.

Chapter 6
Maximizing the Taxonomic Resolution of MALDI-TOF-MS-Based Approaches to Bacterial Characterization: From Culture Conditions Through Data Analysis

Lin Zhang and Todd R. Sandrin

Introduction

Bioterrorism, infectious disease, and microbial contamination of food and water threaten public health and safety (Bain et al. 2014; Chiu 2014; Ferreira et al. 2014; Murray 2010; Trafny et al. 2014). Infectious bacterial diseases cause nearly 20 million deaths annually (Chiu 2014; Lazcka et al. 2007). Therefore, the development of rapid, reliable, and sensitive methods for microbial identification is critically important in environmental monitoring, clinical microbiology, as well as water quality and food safety.

Bacterial resistance to antibiotics has increased significantly recently (Gentile et al. 2014; Martin-Loeches et al. 2014). Antibiotic-resistant strains, such as methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococcus* (VRE), and multidrug resistant bacteria, for example, bacteria with New Delhi Metallo-beta-lactamase-1 (*NDM*-1) gene (Eells et al. 2013; Epstein et al. 2014; Holland et al. 2014b; Morgan et al. 2014; Epstein et al. 2014) complicate the treatment of infections (Livermore 2012). Rapid determination and detection of antibiotic resistant strains play an important role in therapy (Niederman 2009) and is necessary for preventing transmission of such pathogens (Grundmann et al. 2010). This highlights the needs for rapid approaches to differentiate antibiotic-resistant from antibiotic-sensitive strains of pathogenic microorganisms. Strain-level information is also critically important when identification of strains with increased virulence or expanded host range is sought (Li et al. 2009). As a result, rapid bacterial strain typing, or identifying bacteria at the strain level, has become increasingly important in modern microbiology.

Bacterial taxonomic levels remain highly debated in the literature (Gao and Gupta 2012; Staley 2006). The term "strain" is used in this chapter to refer to a taxo-

T. R. Sandrin (✉)
School of Mathematical and Natural Sciences, New College of Interdisciplinary Arts & Sciences, Arizona State University, Phoenix, AZ, USA
e-mail: Todd.Sandrin@asu.edu

L. Zhang
School of Mathematical and Natural Sciences, Arizona State University, MC 2352, PO Box 37100, Phoenix, AZ 85069, USA

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_6

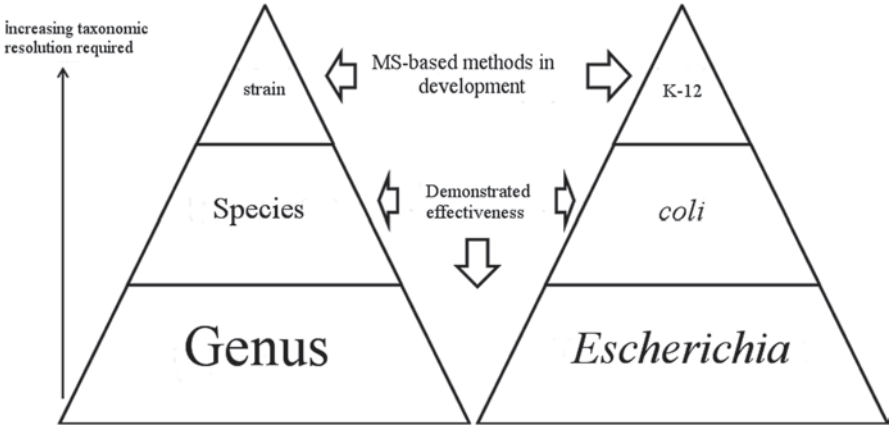


Fig. 6.1 Increasing taxonomic resolution is required to reliably characterize bacteria when moving from the genus level (e.g., *Escherichia*) to the subspecies (e.g., K-12) level. Well-described and effective methods exist for many applications at the genus and species levels, while strain-level applications require additional methodological development and optimization

nomic level more specific and exclusive than species, which includes all subspecies taxa (Fig. 6.1).

Traditionally, methods to type bacteria are classified into two groups: phenotyping and genotyping. Bacterial phenotypes can be determined by assessing the morphology of bacterial colonies on solid media surfaces, gram staining, biochemical/metabolic patterns, immunology-based methods, and antibiotic susceptibility. These methods often do not provide enough information to differentiate closely related strains. Discrimination of strains based on comparison of genetic variation is widely used. Bacteria can be classified using DNA fingerprinting, DNA sequence information, and microarrays (Li et al. 2009). DNA fingerprint-based methods analyze patterns of DNA bands (fragments) which are generated by digestion of genomic DNA using restriction enzymes, amplification of DNA, or by a combination of both. Such methods include pulsed-field gel electrophoresis (PFGE; Spanu et al. 2014), restriction fragment length polymorphism (RFLP; Perez-Boto et al. 2014), repetitive sequencing-based polymerase chain reaction (rep-PCR; Nucera et al. 2013), multiple-locus variable number tandem repeat analysis (MLVA; Shan et al. 2014), and denaturing/temperature gradient gel electrophoresis (DGGE/TGGE; Xiao et al. 2014). Each of these genotyping methods can provide quantitative, accurate information about the unknown bacteria; however, they are time consuming, laborious, and technically demanding. Some methods, such as microarray-based methods, are also particularly expensive (Li et al. 2009).

For more than two decades, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has been shown to be a rapid and effective tool to profile bacteria at the genus and species levels (Dallagassa et al. 2014; Demirev and Fenselau 2008; Freiwald and Sauer 2009; Giebel et al. 2010; Welker and Moore 2011). The utility of this approach to profile bacteria at the strain level has not been as clearly demonstrated, in part, because similar bacteria tend

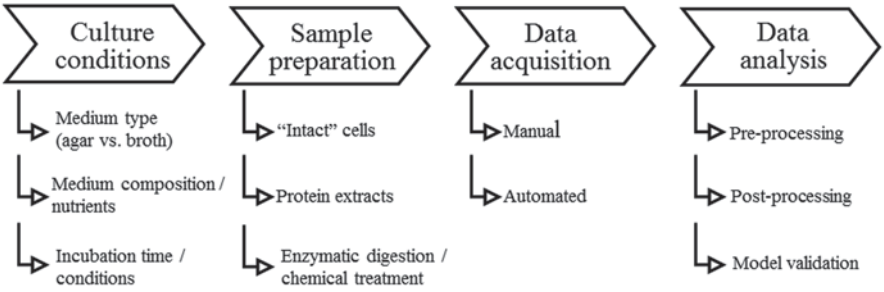


Fig. 6.2 A generic workflow from cultivation through data analysis to characterize bacteria using MALDI-TOF MS. At each step of the workflow, different approaches have been employed and can be optimized to maximize taxonomic resolution

to produce remarkably similar MALDI-TOF profiles that often do not allow facile differentiation of individual strains from one another. The focus of this chapter is on recent developments and the state of the science of maximizing the taxonomic resolution of MALDI-TOF MS-enabled characterization of bacteria. In particular, the chapter focuses on recent approaches employed throughout the MALDI-TOF MS workflow—from culture conditions, sample preparation, data acquisition, through data analysis—that affect and can be optimized to enhance the performance of MALDI-TOF MS-based characterization of bacteria at the strain level (Fig. 6.2).

Overview of MALDI-TOF MS Profiling of Bacteria

Mass spectrometry was first used for microbial characterization by Anhalt and Fenselau in 1975 (Anhalt and Fenselau 1975). By directly inserting lyophilized cells into a double-focusing mass spectrometer (CEC 21-110), *Staphylococcus epidermidis* and *Staphylococcus aureus* produced distinguishable mass profiles. MALDI-MS was first introduced for analysis of high mass peptides and proteins in the late 1980s (Karas et al. 1987). A few years later, protein profiles of lysed and intact bacterial cells, for example *Escherichia coli,* were analyzed by MALDI-TOF MS (Cain et al. 1994; Holland et al. 1996). Results showed that bacteria could be easily distinguished based on these "fingerprint" mass spectra. Since that seminal work, a staggering number and diversity of medically and environmentally relevant bacteria have been profiled using MALDI-TOF MS (e.g., Dallagassa et al. 2014; Ge et al. 2014; Giebel et al. 2008).

Library-Based Approaches

The most commonly employed approach to characterizing bacteria using MALDI-TOF MS involves comparing mass spectra of unknown bacteria to spectra in databases that contain spectra of known reference bacteria (Fig. 6.3a). This library-

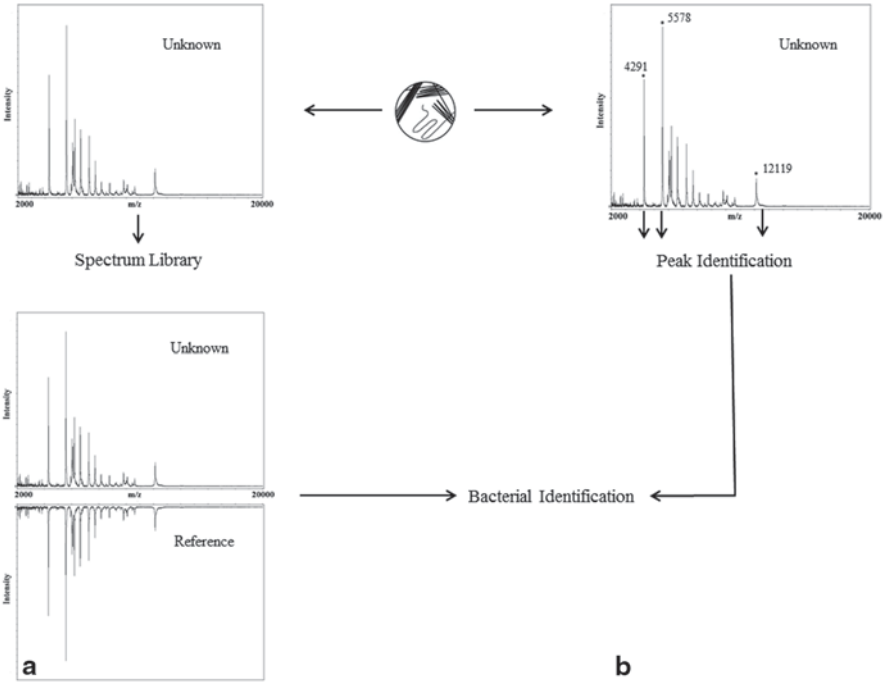


Fig. 6.3 MALDI-TOF MS-enabled characterization of unknown bacterium using **a** library-based approaches and **b** bioinformatics-based approaches

based strategy is popular because of its ease of use and the high speed of data collection (Chiu 2014; Fenselau 2013; Sandrin et al. 2013) . The reproducibility of mass spectra of unknown bacteria must be assessed before matched to mass spectra in the databases, and high reproducibility is critically important, particularly in strain-level applications.

Typically, bacteria are streaked onto agar plates (Grosse-Herrenthey et al. 2008; Pennanec et al. 2010; Stets et al. 2013; Wang et al. 2014) or grown in liquid broth (Han et al. 2014; Wensing et al. 2010; Xiao et al. 2014b). After harvesting from broth or agar media, bacterial cells are inactivated and prepared for MALDI analysis (Freiwald and Sauer 2009). Both intact cells (Han et al. 2014; Niyompanich et al. 2014; Wang et al. 2014) and cell protein extracts (Goncalves et al. 2014; Kopcakova et al. 2014; Stets et al. 2013) have been widely used. In the intact cell approach, cells are often directly mixed with MALDI matrix, such as sinapinic acid or alpha-cyano-4-hydroxycinnamic acid, and the cell-matrix mixtures are deposited onto the target for analysis (AlMasoud et al. 2014; Helmel et al. 2014). Cultures/cell suspensions have also been applied directly to the MALDI target plate and overlaid with matrix (Carbonnelle et al. 2007; Christner et al. 2014; Han et al. 2014). In protein extract-based approaches, cells are lysed using either physical (Fujinami et al. 2011; Sun et al. 2006; Teramoto et al. 2007a) or chemical (AlMasoud et al. 2014;

Wieme et al. 2014b; Zhang et al. 2014b) methods to release the contents of the cells into the supernatant. The supernatant containing the extracted proteins is then either overlaid (Wenning et al. 2014; Wieme et al. 2014b) or mixed (AlMasoud et al. 2014) with MALDI matrix and deposited onto the target. Excellent and comprehensive reviews of various sample preparation methods (Šedo et al. 2011a) as well as a detailed protocol for sample preparation for profiling bacteria using MALDI-TOF MS (Freiwald and Sauer 2009) are available.

Reference spectra are generally collected using the linear detector of the MALDI TOF MS. Both automatic data acquisition using software (e.g., AutoXecute in Bruker's FlexControl software) to control the mass spectrometer (Eddabra et al. 2012; Schumaker et al. 2012; Zhang et al. 2014a) and manual data acquisition (Khot et al. 2012; Schumaker et al. 2012) have been used to collect spectra. Automated data acquisition can enhance the high-throughput nature of MALDI profiling, while manual data acquisition has been shown to yield mass spectra with higher quality and reproducibility. Typically, spectra are collected over a mass range from 2 to 20 kDa (Salaun et al. 2010; Stets et al. 2013; Thevenon et al. 2012), while broader mass ranges, for example, 1 Da–100 kDa (Jackson et al. 2005) and 1 Da–20 kDa (Hettick et al. 2006), and narrower mass ranges such as 500 Da–10 kDa (Keys et al. 2004; Rajakaruna et al. 2009) and 7–10 kDa (Sauer and Kliem 2010) have also been employed. In automatic data acquisition, users must specify threshold values of several instrument operation parameters, including minimum base peak intensity, resolution, signal-to-noise ratio, and the number of peaks accumulated. These operation parameters can be optimized to increase the quality and reproducibility of the resulting mass spectra (Zhang et al. 2014a). After collecting mass spectra from a large collection of bacterial strains, the spectra are processed and analyzed to obtain information (e.g., similarity coefficients and potential biomarker peaks) to characterize unknown bacteria. A reference spectrum of a species is generated by summarizing the processed spectra of technical and/or biological replicates of the species. The mass spectra of unknown bacteria are then compared with the reference spectra for characterization using a variety of metrics, including manufacturer-defined algorithms (e.g., BioTyper scores), Pearson product–moment correlation coefficients, and jackknife values, which are described more fully later in this chapter.

Bioinformatics-Enabled Approaches

In contrast to library-based approaches which typically do not involve identification of the biological nature/origin of particular peaks, bioinformatics-enabled approaches identify peaks in MALDI profiles to characterize unknown bacteria (Fig. 6.3b). Bioinformatics-enabled approaches are commonly applied to bacteria with sequenced genomes. Two methods have been used in bioinformatics-enabled approaches: bottom-up and top-down methods. Bottom-up methods involve digestion of proteins using enzymes, such as trypsin, prior to MS analysis. The enzymes cleave at well-defined sites (e.g., after every arginine and lysine, in the case of tryptic digestion) of the proteins to create complex peptide mixtures. Peptides in

the mixtures can be subjected to collision-induced dissociation (CID) to generate fragments, and the masses of the fragments can be determined. The masses of the peptide fragments are used to identify the proteins by searching databases (e.g., NCBI). The identified proteins are used as biomarkers to identify bacteria (Fenselau et al. 2007; Pribil et al. 2005; Russell et al. 2007). In contrast, top-down methods introduce intact (undigested) proteins into the mass spectrometer. The intact proteins are fragmented into smaller peptides in tandem mass spectrometry (MS/MS). The experimental MS/MS spectra are compared with in silico-generated MS/MS spectra from protein sequences in proteome databases for rapid identification of bacteria (Demirev et al. 2005; Wynne et al. 2009). Top-down methods have also been successfully applied to distinguish a pathogenic *E. coli* strain (O157:H7) from the non-pathogenic strains (non-O157:H7) (Fagerquist et al. 2010). Spectral reproducibility of bioinformatics-enabled approaches is not critical as long as the ions of biomarkers are consistent with the sequences in the database. Thus, bioinformatics-enabled approaches do not require rigorously standardized protocols across laboratories to the extent required by library-based approaches. However, bioinformatics-enabled approaches are rarely applied to microorganisms that do not have fully sequenced genomes and readily available protein/peptide databases.

Successes at the Genus and Species Levels

Genus- and species-level characterization of bacteria using MALDI-TOF MS with library-based and/or bioinformatics-enabled methods has been successfully applied in many areas. For example, in clinical microbiology, MALDI-TOF MS has been used to identify pathogens directly from monomicrobial positive blood cultures (Klein et al. 2012; Martinez et al. 2014; Rodriguez-Sanchez et al. 2014) and urine samples (Ferreira et al. 2010b; Rossello et al. 2014; Wang et al. 2013) . Correct identification rates at the genus and species levels shown in these studies range from 98% (Martinez et al. 2014) down to 70% (Klein et al. 2012). Gram-negative bacteria have been reported to be more readily identified correctly than Gram-positive bacteria (Klein et al. 2012). Schrottner et al. (2014) reported that MALDI-TOF MS can distinguish between two opportunistic pathogens, *Myroides odoratus and Myroides odoratiminus* at the species level, and results were comparable to those obtained with 16S rDNA sequencing. Zhang et al. (2014b) used MALDI-TOF MS to identify *Lactobacilli* isolated from saliva samples of adults with dental caries. Results showed that 88.6% of *Lactobacillis* isolates and 95.5% of *non-Lactobacillis* isolates were correctly identified at the genus level using MALDI-TOF MS. These rates were comparable to those obtained using 16S rDNA sequencing (Zhang et al. 2014b). Hsueh et al. (2014) showed that *Acinetobacter* species isolated from blood samples could be correctly identified using MALDI with commercially available software (Bruker's Biotyper). The correct identification rate for various *Acinetobacter* species ranged from 98.6% down to 72.4% (Hsueh et al. 2014). Both library-based (Ilina et al. 2010) and bioinformatics-enabled methods (Xiao et al.

2014c) have been successfully applied to identify the etiologic agent of stomach ulcers, *Helicobacter pylori*, a Gram-negative, microaerophilic bacterium with high genetic variability. Furthermore, differentiation between *Streptococcus pneumonia* and some closely related species, such as, *Streptococcus pseudopneumoniae*, *Streptococcus mitis* and *Streptococcus ordis*, is difficult and misidentifications occur with routinely employed molecular methods (Werno et al. 2012). MALDI-TOF MS has been reported to facilitate identification of 75 % of *Streptococcus* isolates at the genus and species levels (Wessels et al. 2012). Werno et al. (2012) suggest that rigorous examination of the mass peak profiles can enhance the ability of MALDI-TOF MS to distinguish *Streptococcus pneumonia* from nonpneumococcal isolates . MALDI-TOF MS is not only applicable to aerobic bacteria, but also applied to anaerobic bacteria. Zarate et al. (2014) used MALDI-TOF MS to identify 106 clinical isolates of anaerobic bacteria. The correct identification rate at the genus and species levels was 95.3 %, comparable to that obtained using conventional biochemical tests . The possibility of using MALDI to identify and type anaerobic bacteria has been reviewed recently (Nagy 2014).

Recently, MALDI-TOF MS has been employed in food microbiology. *Enterococcus* species are considered to be secondary contaminants of food and often play roles in food spoilage. Some closely related enterococcal species are difficult to discriminate using 16S rDNA sequencing, while one study has shown that *Enterococcus* can be rapidly identified at the species level using MALDI-TOF MS (Quintela-Baluja et al. 2013). Other food-borne pathogens and spoilage bacteria, such as *Arcobacter* spp., *Helicobacter* spp., *Campylobacter* spp., *Lactobacillus* spp., *Pediococcus* spp., *Leuconostoc* spp., *Streptococcus* spp., *Clostridium* spp., and *Staphylococcus* spp., have also been identified at the species level using library-based MALDI-TOF MS methods (Alispahic et al. 2010 Bohme et al. 2011a; Bohme et al. 2011b; Han et al. 2014; Kern et al. 2013; Regecova et al. 2014; Wieme et al. 2014b). Some *Bacillus* species are pathogens or spoilage agents in food products. MALDI-TOF MS was shown to be more effective than 16S rDNA sequencing to differentiate *Bacillus subtilis* and *Bacillus cereus* from *Bacillus amyloliquefaciens* and *Bacillus thuringiensis* (Fernandez-No et al. 2013). Acetic acid bacteria are involved in the industrial production of vinegar. Andrés-Barrao et al. (2013) characterized 64 strains of acetic acid bacteria belonging to the genera *Acetobacter*, *Gluconobacter,* and *Gluconacetobacter* using MALDI-TOF MS with the SARAMIS™ software package (Spectral Archive and Microbial Identification System; Anagnostec Gmbh, Germany). Results showed that these acetic acid bacteria could be rapidly and reliably identified using fingerprint mass spectra (Andres-Barrao et al. 2013). Bohme et al. (2013) used two technologies, 16S rDNA sequencing and MALDI-TOF MS, to identify pathogens in seafood. DNA sequencing identified only 50 % of the strains at the species level and performed relatively poorly with regard to identification of members of the *Pseudomonas* and *Bacillus* genera. In contrast, MALDI-TOF MS facilitated correct identification of 76 % of the isolated strains and showed a higher rate of correct classification of members of the *Pseudomonas* and *Bacillus* genera.

In addition to the applications in clinical and food microbiology, MALDI-TOF MS has been used to characterize culturable bacterial populations isolated from various environments, including laboratory mice and rats (Goto et al. 2012), horse semen samples (Masarikova et al. 2014), human feces (Samb-Ba et al. 2014), air (Setlhare et al. 2014), as well as recombinant bacteria (Xiao et al. 2014a). All of these studies suggest that MALDI-TOF MS is a rapid, reliable, and alternative method for characterizing bacteria particularly at the genus and species levels.

Strain-Level Characterization: Successes, Challenges, and Strategies

MALDI-TOF MS has shown promise at the strain level. For example, Christner et al. (2014) used MALDI-TOF MS to type 294 *E. coli* isolates collected during a large outbreak in northern Germany. Strain-specific biomarker peaks were reported, and 99 % of the *E. coli* strains were correctly identified using strain-specific biomarkers (Christner et al. 2014). Similarly, Schafer et al. (2014) used MALDI to identify ERIC-genotypes of *Paenibacillus larvae* strains. *P. larvae* is the causative agent of American foulbrood disease in honeybees. Results showed that with a reference database, ERIC I and II types of *P. larvae* strains could be unambiguously identified. In the food industry, *Lactobacillus brevis* strains exhibit varying beer-spoiling abilities. Kern et al. (2014) showed that strain-level identification of 17 *Lactobacillus brevis* strains was achieved in 90 % of 204 spectra (Kern et al. 2014a).

The MALDI-TOF MS fingerprint technique also shows promise in tracking strains isolated from different environmental sources. Siegrist et al. (2007) showed that a limited number of environmental *E. coli* strains could be grouped according to the source from which they were isolated. Similarly, Niyompanich et al. (2014) showed that 6 out of 11 clinical and environmental *Burkholderia pseudomallei* strains were grouped correctly according to their respective sources . Strain-level characterization has also been demonstrated for *Propionibacterium acnes* (Nagy et al. 2013) and *Bacteroides fragilius* (Nagy et al. 2011). Fujinami et al. (2011) reported using MALDI-TOF MS to successfully discriminate 23 *Legionella pneumophila* strains . In addition, MALDI-TOF MS has been used to detect antibiotic resistance associated with identified strains, which is extensively reviewed elsewhere in this book (Chapter(s) XX).

While successes using MALDI-TOF at the strain level have been described, characterization at this level faces many more challenges than applications at higher taxonomic ranks. For example, Zeller-Peronnet et al. (2013) studied the discriminatory power of MALDI-TOF MS to differentiate 24 strains within *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides* species. Results showed that, although individual species could be readily identified, only half of the strains could be correctly identified to the strain level, suggesting that the discriminatory power of MALDI might not be adequate for characterization of these two species at the strain level (Zeller-Peronnet et al. 2013). Several studies have explored whether MALDI-

TOF MS can differentiate MRSA from methicillin-sensitive *Staphylococcus aureus* ones (MSSA), but there is limited consensus (Bernardo et al. 2002; Jackson et al. 2005; Lasch et al. 2014; Majcherczyk et al. 2006; Walker et al. 2002). Conflicting results may be because of low quality and reproducibility of the fingerprint mass spectra. In addition, cultivation conditions and sample preparation methods have also been suggested to affect strain-level differentiation (Goldstein et al. 2013; Balážová et al. 2014a). Besides *Staphylococcus aureus*, MALDI-TOF MS has also showed insufficient discriminatory power for typing strains of other bacteria. Lasch et al. (2014) reported that MALDI could not reliably differentiate *Enterococcus faecium* strains based on clonal complexes and multilocus sequence types. Kern et al. (2014) reported that when highly similar strains of the beer-spoilage bacterium *Pectinatus frisingensis* were incorporated for analysis using MALDI, the correct identification rate for *P. frisingensis* at the strain level decreased from 73 % (using a relatively diverse set of strains) to 60 % (including those that were highly similar to one another; Kern et al. 2014b) . A more comprehensive review of reports on bacterial strain categorization, differentiation, and identification using MALDI-TOF MS is available (Sandrin et al. 2013).

Commercially Available Software

In library-based approaches, bacterial characterization requires comparison of mass spectra of unknowns with those of reference spectra of known bacteria. Though visual inspection can sometimes provide a qualitative assessment of the similarity between mass spectra at the genus and possibly the species levels, software algorithms have been developed to provide more objective and quantitative assessments. Such tools are critically important at the strain level, where spectra of closely related strains are often extraordinarily similar, and reliable discrimination requires sensitive and repeatable measures. Many software solutions, such as custom R packages, Microsoft Excel (Microsoft Corporation, Redmond, WA), MATLAB (MathWorks, Matick, MA), and BioNumerics (Applied Maths, Sint-Martens-Latem, Belgium), have been applied to enhance analysis (Croxatto et al. 2012; Pavlovic et al. 2013; Sandrin et al. 2013).

Currently, two of the most frequently used commercially available software packages are BioTyper (Bruker Daltonics, Billerica, MA, USA) and SARAMIS (bioMérieux, SA, Marcyl'Etoile, France). BioTyper is commonly used in the clinical setting with the MicroFlex LT mass spectrometer (Bruker Daltonics; Billerica MA, USA). SARAMIS is offered by bioMérieux and is routinely used with the Vitek MS system (bioMérieux SA; Marcyl'Etoile, France). Both BioTyper and SARAMIS provide a database environment and spectrum comparison algorithm that allows for quantitative comparison and identification of bacteria within the database. These two software packages also allow users to build custom databases, in which spectra acquired in-house can be added to the database. For quantitative comparison, both BioTyper and SARAMIS require summarization of mass spectra

of biological and/or technical replicates of bacteria. The resulting composite mass spectra are called main spectral projections (MSPs) in BioTyper and a SuperSpectrum™ in SARAMIS.

BioTyper and SARAMIS have been most commonly used to identify bacteria at the genus and species levels. BioTyper uses a score-based classification system for bacterial identification, while SARAMIS uses a percentage-based method. These algorithms are described more fully later in this chapter. In addition, studies have suggested that BioTyper has similar (Mather et al. 2014) or superior (Chen et al. 2013) abilities to SARAMIS with regard to identification at the genus and species levels. With regard to the strain-level characterization, BioTyper has been suggested to be useful in characterizing bacteria at the strain level, but typically with the assistance of additional software, such as ClinProTools (Bruker Daltonics) and R. Ayyadurai et al. (2010) reported that, with assistance of ClinProTools, three biotypes of *Yersinia pestis* strains could be differentiated. Nakano et al. (2014) used three ClinProTools models to differentiate vanA-positive *Enterococcus faecium* from vanA-negative *Enterococcus faecium* with MALDI and BioTyper. All three ClinProTools models yielded > 90% recognition capability . Karger et al. (2011) used BioTyper with the R-package caMassClass to filter out some peaks, and this data-reduction strategy enhanced categorization of strains of Shiga toxin-producing *E.coli* . Strain-level applications of SARAMIS have not been frequently reported in the literature.

Assessing Strain-Level Performance

Objectives

Strain-level characterization often entails one or more three distinct objectives: (1) strain categorization, (2) strain differentiation, and (3) strain identification (Fig. 6.4). These three objectives often require different levels of taxonomic resolution. Strain categorization involves grouping bacterial strains that share a particular trait, such as their origin (Dubois et al. 2010; Siegrist et al. 2007), antibiotic resistance (Shah et al. 2011; Wolters et al. 2011), pathogenicity (Stephan et al. 2011), and/or as different subspecies/biotypes (Ayyadurai et al. 2010; Lundquist et al. 2005; Zautner et al. 2013). Strain categorization does not typically involve discriminating single strains. In contrast, strain differentiation requires distinguishing single strains and thus higher taxonomic resolution. Many studies have reported that bacterial strains, of both medical and environmental relevance, can be differentiated based on the presence and/or absence of one or more strain-specific biomarker peaks (Donohue et al. 2006; Everley et al. 2008; Ghyselinck et al. 2011; Majcherczyk et al. 2006; Ruelle et al. 2004; Zautner et al. 2013) or by cluster analysis (Balážová et al. 2014b; Holland et al. 2014a). Among these three objectives, strain identification requires the highest taxonomic resolution. When comparing with a reference library, strains often cannot be confidently identified based on the presence of only one or a few

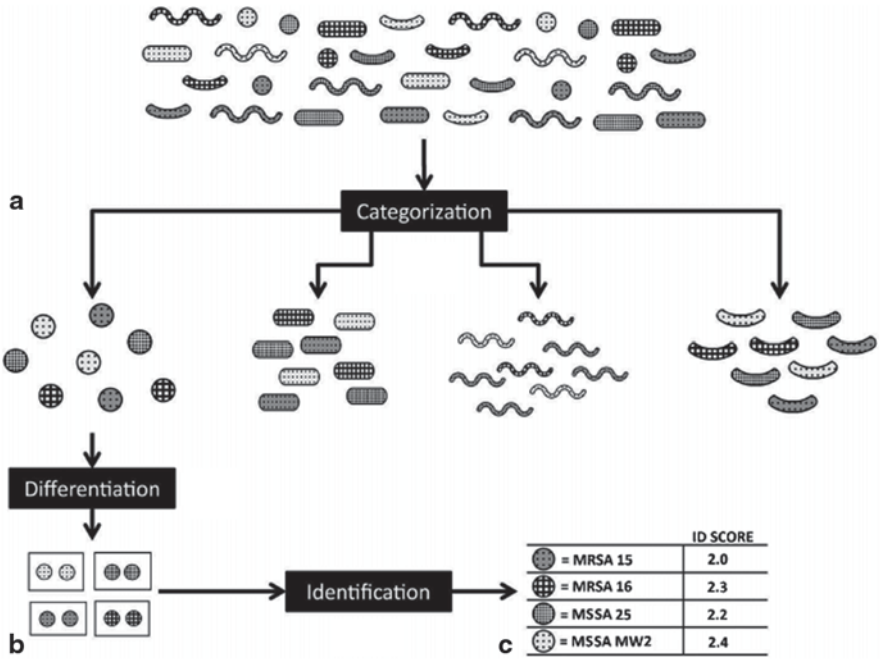


Fig. 6.4 Objectives of strain-level applications of MALDI-TOF MS-enabled characterization of bacteria have included categorization (**a**), differentiation (**b**), and identification (**c**). The requisite level of taxonomic resolution tends to increase as one progresses from efforts to categorizing strains to identifying individual strains. (Adapted from Sandrin et al. 2013, copyright John Wiley & Sons)

biomarker peaks. Analysis of the entire spectrum (Holland et al. 2014b) with rigorous analytical tools, such as those in R, ClinProTools, and BioTyper software (Nakano et al. 2014), is often required to obtain reliable strain identification.

Reproducibility

Closely related strains of bacteria yield highly similar mass spectra (Fig. 6.5). To reliably characterize strains, the reproducibility of replicated mass spectra of the same strain must be quantified before conducting further analysis. Here, reproducibility refers to how similar replicate spectra of the same strain are to one another based on comparing peak presence/absence and/or peak intensity. Reproducibility (similarity) between replicates of the same strain must exceed the similarity of mass profiles of closely related bacterial strains. Several previous studies have examined reproducibility based primarily on visual inspection of spectra (Arnold and Reilly 1998; Jackson et al. 2005), while more recently, studies have quantified reproducibility more rigorously using software packages described previously. The coefficient of

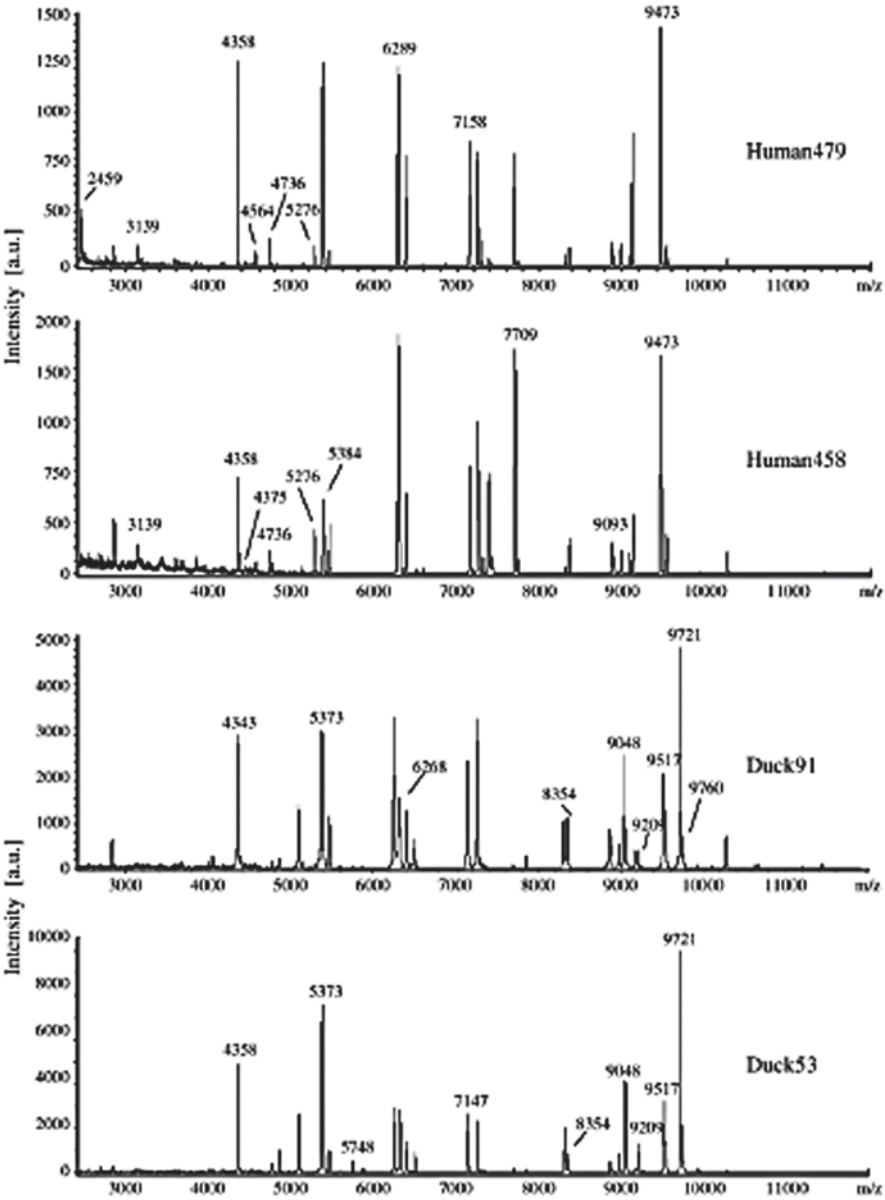


Fig. 6.5 Bacteria of the same species tend to produce similar MALDI-TOF spectra. Four environmental isolates of *Escherichia coli* yielded MALDI spectra that were particularly similar within a single environmental source (i.e., duck or human origin), but adequately distinct across different sources to allow characterization of isolates based upon source. (Adapted from Siegrist et al. 2007, with permission from Elsevier)

variation of each mass signal of replicate spectra has been used to quantify the reproducibility of replicate spectra (Freiwald and Sauer 2009). Chen et al. (2008) combined analysis of variance and principal component analysis (ANOVA–PCA) to quantify the reproducibility of replicate spectra . Toh-Boyo et al. (2012) reported using ANOVA, PCA, and multivariate ANOVA (MANOVA) to assess reproducibility . The curve-based Pearson correlation coefficient has also been used to measure reproducibility (Giebel et al. 2010; Schumaker et al. 2012; Zhang et al. 2014a). Binary coefficients (which do not include peak intensity measurements), including the Dice similarity coefficient (Ziegler et al. 2012) and Jaccard index (Erler et al. 2014), have also been used. As detailed below, both experimental and data analysis factors have been reported to affect the reproducibility of replicate spectra.

Group Separation/Performance Metrics

Score

BioTyper (Bruker Daltonics, Billerica, MA, USA) and SARAMIS (bioMérieux, SA, Marcyl'Etoile, France) are the most commonly used software and database packages for bacterial identification at the genus and species levels. Both use a score/percentage-based method to assess performance. BioTyper compares patterns of peaks of unknowns with reference spectra based on peak position, peak intensity, and peak frequencies. A log score is obtained for each comparison using a manufacturer-defined algorithm. The manufacturer proposes that a score ≥ 2 indicates species identification; a score between 1.7 and 1.9 indicates genus identification, and a score < 1.7 indicates no identification. Similar to BioTyper, SARAMIS uses a confidence percentage for genus and species identifications. Genus-level identification is acceptable when the confidence percentage ranges from 98% to 90%. Species identification is acceptable when the confidence percentage is $> 98\%$. Identification is not acceptable when the confidence percentage is $< 90\%$ (Chen et al. 2013). With regard to the Sepsityper™ kit (Bruker Daltonics; Billerica, MA, USA) to directly identify bacteria from positive blood cultures, the score cutoff is 1.8 for identification at species level and 1.6 for that at the genus level (Nonnemann et al. 2013).

Reliable identification of aerobic and anaerobic bacteria at the genus and species levels using BioTyper scores has been reported in many studies (Balada-Llasat et al. 2013; Coltella et al. 2013; Ferreira et al. 2010a; Hsueh et al. 2014; Ikryannikova et al. 2013; Lacroix et al. 2014; Mather et al. 2014; Sanchez-Juanes et al. 2013; Schulthess et al. 2014). The correct identification rate at the genus and species levels ranges from 100% (Ikryannikova et al. 2013; Sanchez-Juanes et al. 2013) down to 72% (Hsueh et al. 2014). Furthermore, some studies have suggested that reducing the species cutoff, for example, from 2.0 to 1.7 (Mather et al. 2014; Pereyre et al. 2013; Schulthess et al. 2014), and from 1.8 to 1.5 for positive blood cultures using Sepsityper™ kit (Nonnemann et al. 2013), can increase species identifica-

tion rates for some species. However, unreliable and/or no identification for some bacterial species, such as *Nocardia* spp. (except for *N. nova* and *N. otitidiscaviarum*), *Tsukamurella* spp., *Gordonia* spp., and *Avibacterium spp*. (except for *A. paragallinarum*), using BioTyper scores has also been reported (Alispahic et al. 2014; Hsueh et al. 2014).

Jackknife/Bootstrapping/Threshold Cutoffs

The jackknife test has also been used to evaluate the performance of MALDI. The principle of the jackknife test is to take out one entry from a group (e.g., genus, species, and/or strains) and to identify this entry against different groups (Johnson and Wichern 2007). The procedure is repeated for all entries in this group, and the percentage of correct group identifications is used to assess the accuracy of MALDI for characterization of the group. The jackknife test has been used to evaluate the accuracy of bacteria identification at the species (De Bruyne et al. 2011) and strain levels (Goldstein et al. 2013). Besides the jackknife statistic, bootstrapping has also been used to estimate the performance of classification models of isolates. For example, AlMasoud et al. (2014) used PCA and support vector machines (SVM) to supervise peak classification of a peak table matrix containing 243 unique peaks for characterizing *Bacillus* spp. The SVM models were validated by using a bootstrap replacement procedure coupled with cross-validation for selection of model parameters. Classification accuracies at the *Bacillus* species level reached about 90 % using the validated SVM models.

Culture Conditions

A variety of media, including solid agar (e.g., Grosse-Herrenthey et al. 2008; Pennanec et al. 2010) and liquid broth (e.g., Wensing et al. 2010), have been used to support the growth of bacteria to be characterized by MALDI. Mass spectra of bacteria consist of peaks mainly derived from ribosomal proteins and other abundant proteins (Ryzhov and Fenselau 2001; Teramoto et al. 2007b). Ribosomal proteins are highly conserved and are not expected to be affected by culture conditions (Arnold and Reilly 1999). Several studies have also shown that a core set of species-specific peaks are constantly observed regardless of the media used (De Bruyne et al. 2011; Grosse-Herrenthey et al. 2008; Hsu and Burnham 2014; Kern et al. 2013; Lartigue et al. 2009; Sauer et al. 2008); however, culture conditions can influence the expression pattern of other proteins (Valentine et al. 2005). Thus, media effects have been widely reported (Horneffer et al. 2004; Moura et al. 2008; Ruelle et al. 2004; Walker et al. 2002), though some studies have suggested that the effects are subtle and do not affect the overall ability of MALDI to characterize bacteria (Bernardo et al. 2002; Dieckmann et al. 2008; Kern et al. 2013; Vargha et al. 2006).

Taken together, these results indicate that medium effects at the strain level may be more pronounced and significant than at higher taxonomic ranks.

Medium Type

Several studies have shown that medium components can affect the rate of successful identification of many bacteria. For example, Anderson et al. (2012) reported that the correct identification rate at the genus level for bacteria grown on different types of agars varied for *Pseudomonas, Staphylococcus,* and enteric isolates. The effect was most pronounced for *Staphylococcus* isolates, varying from 75% on colistin-nalidixic acid agar to 95% on blood agar and mannitol salt agar (Anderson et al. 2012). Variations have been reported with other *Staphylococci* using different media (Rajakaruna et al. 2009; Walker et al. 2002). Similarly, results from one study suggested that changing the medium could improve differentiation between closely related members of the family *Enterobacteriaceae* (Keys et al. 2004). Ford and Burnham (2013) grew 24 enteric Gram-negative bacteria (EGNB) and 25 non-glucose-fermenting/fastidious Gram-negative bacteria (NFGNB) on different types of agars. Results showed that the rate of successful identification on EGNB at the species level was approximately 20% less than on other types of agars. For NFGNB, rates of correct identification at the species level varied from less than 60% on OFPBL agar to 90% on sheep blood agar, and one misidentification was observed for bacteria grown on MacConkey agar (Ford and Burnham 2013).

With regard to subspecies and strain-level characterization, a few studies have reported that medium components do not affect the strain-level resolution (Bernardo et al. 2002; Dieckmann et al. 2008; Vargha et al. 2006). However, other studies have reported that medium composition significantly affects strain-level performance. Šedo et al. (2013) anaerobically cultivated strains from six *Lactobacillus* species in two kinds of liquid media, De Man-Rogosa-Sharpe broth and anaerobe basal broth, and on blood agar plates. No obvious effect of medium type on species-level resolution was observed, but some closely related strains could be distinguished only with a specific cultivation medium. Wieme et al. (2014a) studied effects of five different culture media on the differentiation of 25 strains of acetic acid bacteria, including *Acetobacter, Gluconacetobacter,* and *Gluconobacter* strains, at the species and strain levels. For each strain grown on a particular medium type, a single composite mass spectrum was obtained. Results showed that culture media did not affect species-level differentiation, but strongly affected the number of shared strain-specific peak classes in the composite mass spectra of the same strain grown on different media, in particular when the culture media did not sustain optimal growth. Balážová et al. (2014) tested the effects of four culture media on the discriminatory power of MALDI to characterize 10 strains belonging to *Mycobacterium phlei* and *Mycobacterium smegmatis* (Balážová et al. 2014a). The successful identification rate calculated from Biotyper score cutoffs was similar for all the four media at the genus level, but twofold higher for one preferable medium over the others at the species level (Balážová et al. 2014a).

Medium type has also affected strain-level differentiation and identification. The influence seems specific for particular species and strains. For example, Šedo et al. (2013) examined the influence of growth conditions on strain differentiation within the *Lactobacillus acidophilus* group (17 strains representing six different species from the *Lactobacillus acidophilus* group). Results showed that two *Lactobacillus acidophilus* strains could be distinguished after cultivation on blood agar, but could not be distinguished when grown on other kinds of media, while other *Lactobacillus acidophilus* stains could be distinguished regardless of medium types. Balážová et al. (2014) reported that, generally, M7H9 medium generated a higher correct identification rate than Herrold's egg yolk medium (HEYM) medium for characterizing *Mycobacterium* strains. The effect was even more significant for *Mycobacterium phlei* strains. For example, 89% of *Mycobacterium phlei* strains were correctly identified using M7H9 medium, while only 50% of *Mycobacterium phlei* strains were correctly identified using HEYM medium. With regard to *Mycobacterium smegmatis* strains, HEYM medium showed a lightly better identification result (60%) than M7H9 medium (52%; Balážová et al. 2014a). These studies clearly showed that medium type has the potential to affect MALDI-TOF MS profiles, but the effects may be bacterium specific. To maximize taxonomic resolution, especially at the strain level, the potential effects should be thoroughly investigated. Databases containing multiple reference strains grown on different culture media may need to be established.

Medium Form (Broth/Agar)

Besides the components in the medium, medium form (agar or broth) has also been investigated with regard to potential effects on the taxonomic resolution of MALDI-TOF MS. At the species level, Lotz et al. (2010) cultivated 311 *Mycobacterium* strains both on agar plate and in liquid broth. Using an intact cell preparation method, correct identification rates were 97% for solid media and only 77% for liquid media. The low identification rate for liquid media was suggested to be because of the failures of spectrum acquisition (Lotz et al. 2010). Balada-Llasat et al. (2013) cultured 178 mycobacterial isolates using both solid and liquid media. Using a protein extraction-based sample preparation method, 93.8% of the isolates were identified correctly at the species level with both forms of media, suggesting that medium forms may not affect MALDI resolution at the species level. At the strain level, though, Goldstein et al. (2013) cultured MRSA and MSSA on both solid agar plates and broth media. Results showed that liquid media generated higher jackknife values when differentiating MRSA from MSSA, suggesting that culture in liquid media enhances the discriminatory power of MALDI. The higher discriminatory power may be because of the more homogeneous populations of cells synchronized in their growth phase in the broth cultures, whereas on an agar plate, colonies consist of older, senescent cells in the center and newer, more actively growing cells at the perimeter (Sandrin et al. 2013) .

Sample Preparation

As comprehensively reviewed by Šedo et al. (2011), a myriad of diverse sample preparation techniques have been used to profile bacteria with MALDI. These diverse techniques can be classified into two types: intact cell-based and protein extract-based methods. Intact cell-based methods involve deposition of cells or cell suspensions onto the MALDI target, while protein extract-based methods involve deposition of cell extracts onto the MALDI target. While intact cell-based methods do not involve intentional extraction of cell materials, the chemicals added may still cause cell degradation. Intact-cell based methods are simpler and more rapid than cell protein extract-based methods, because they do not require additional steps to break cells and extract proteins (Sauer et al. 2008). Sample preparation methods have been suggested to affect the taxonomic resolution of MALDI-based approaches to bacterial characterization (Šedo et al. 2011). For example, Zampieri et al. (2013) used both intact cell-based and protein extraction-based methods to identify 11 bacteria isolated from bovine semen at the genus and species levels. The intact cell-based method resulted in correct identification of nine bacteria at the genus and species levels, while a protein extraction-based method afforded correct identification of all 11 bacteria at the genus and species levels (Zampieri et al. 2013). Therefore, many studies have explored multiple preparation methods to maximize the taxonomic resolution of MALDI profiling (e.g., Dieckmann et al. 2008; Ruelle et al. 2004; Williams et al. 2003). Interestingly, commercially available platforms have been applied using different sample preparation approaches. For example, Lohmann et al. (2013) used a cell extract-based sample preparation approach with Bruker's BioTyper, but an intact cell-based approach with bioMérieux's SARAMIS.

Intact Cells

Generally, cells have been directly deposited onto MALDI targets in two ways. Some studies have described placement of cell suspensions/colonies directly on the MALDI targets and subsequently overlaid matrix solutions which usually contain TFA and ACN (Carbonnelle et al. 2007; Christner et al. 2014; Han et al. 2014; Jackson et al. 2005; Walker et al. 2002). Other studies have reported mixing cell suspensions with matrix solutions prior to deposition onto the MALDI target (AlMasoud et al. 2014; Arnold and Reilly 1998; Dickinson et al. 2004; Donohue et al. 2006; Helmel et al. 2014; Moura et al. 2008; Ryzhov et al. 2000; Welham et al. 1998). Both "overlaid" and "premixed" approaches of intact cell-based sample preparation have afforded strain-level resolution (Kuehl et al. 2011; Zautner et al. 2013); however, only a few studies have directly compared the effects of these two deposition approaches on the taxonomic resolution of MALDI profiling of bacteria. Jackson et al. (2005) compared the performance of MALDI to profile MRSA at the

stain level by directly depositing MRSA colonies from agar plates with directly depositing MRSA cell suspensions onto the target plate. Results indicated that directly deposited colonies yielded higher quality spectra and higher reproducibility than deposited cell suspensions (Jackson et al. 2005). Kuehl et al. (2011) applied three techniques, including premixed, overlaid (ethanol added as an organic modifier), and a sandwich method with cells placed between two layers of matrix, for applying samples to the MALDI target. Results showed that the sandwich method generated the highest quality mass spectra when characterizing *Enterococcus faecalis* (Kuehl et al. 2011). Toh-Boyo et al. (2012) compared the reproducibility of the mass spectra resulting from laser sampling at different regions within a single target well (intrasample) and between target wells (intersample) using both overlaid and premixed pipet-based deposition methods with three different matrices . The authors observed that the crystalline morphology of the sample on the target greatly influenced intrasample reproducibility (Fig. 6.6a). Samples deposited using the pipet-based premixed method yielded less variability between spots for a single sample than the pipet-based overlay method (Fig. 6.6b). More recently, AlMasoud et al. (2014) reported similar results showing that a premixed deposition method worked best for typing *Bacillus* and *Brevibacillus* species over the other three deposition methods, including (1) overlaying matrix onto protein extracts, (2) overlaying protein extracts onto matrix, and (3) a sandwich method in which protein extracts were situated between two layers of matrix .

Protein Extracts

Similar to intact cell-based methods, several protocols have been described to extract proteins from cells. The most commonly used protein extraction method is an ethanol-formic extraction, in which a crude protein extraction is performed in a microcentrifuge tube (e.g., Freiwald and Sauer 2009). In addition, plate-based formic acid extraction has also been described (e.g., Schulthess et al. 2014).

Several studies have suggested that protein extraction-based methods afford higher taxonomic resolution than intact cell-based methods. For example, Schulthess et al. (2014) compared three sample preparation methods for identification of 190 Gram-positive rods including 64 species from 21 genera. Results showed that species-level identification rates were higher with a plate-based formic acid extraction and an ethanol-formic acid extraction than with an intact cell-based direct smear method. Rates of correct identification were 68.4 % with plate-based formic acid extraction, 71.6 % with ethanol-formic acid extraction, and 63.7 % with a direct smear method.

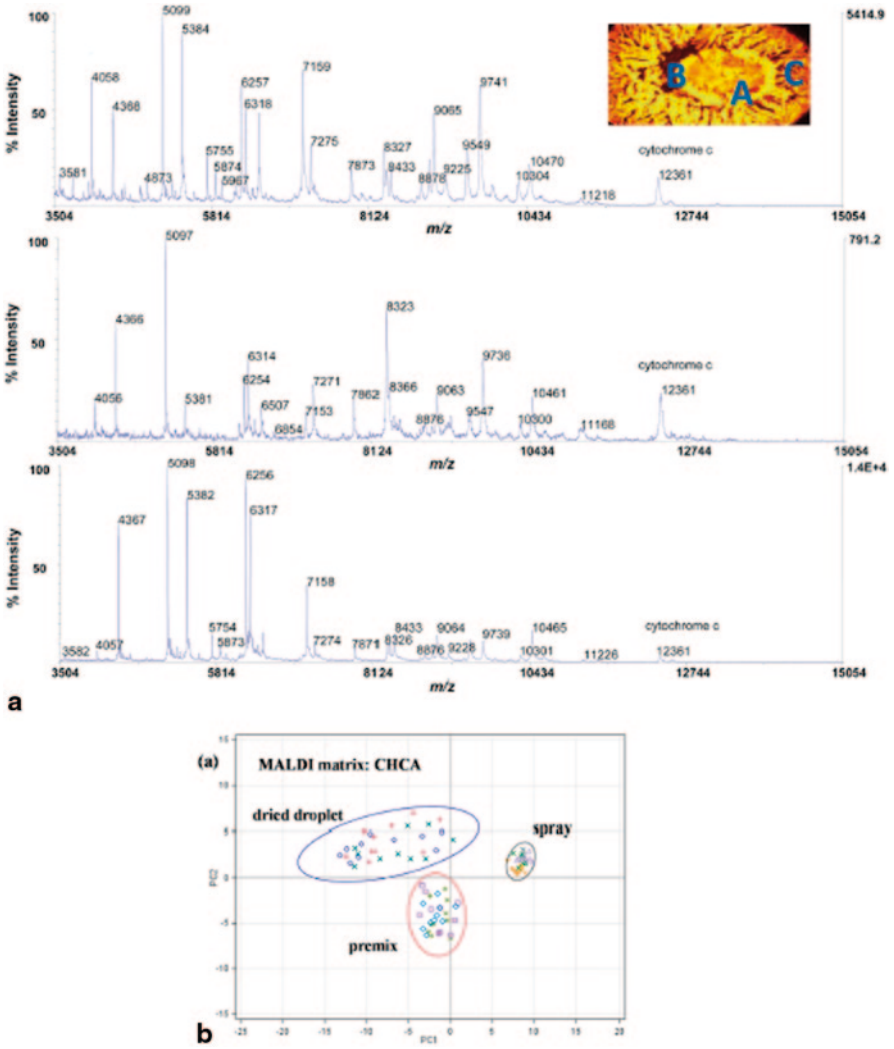


Fig. 6.6 MALDI-TOF spectra from a single sample in a single MALDI target well yielded distinct spectra (**a**). Similarly, the mode of application of sample to the MALDI target plate affected spectra (**b**). These findings underscore the need to (i) ensure sample preparation techniques maximize sample homogeneity on the target plate and (ii) ensure standardized sample deposition protocols are followed when using library-based MALDI-TOF-based approaches at the strain level. (Adapted from Toh-Boyo et al. 2012, copyright American Chemical Society)

Pretreatment to Enhance Taxonomic Resolution

Several pretreatments have been used to increase the discriminatory power of library-based MALDI fingerprint approaches. Horneffer et al. (2004) used wet-heat treatment to extract additional analytes that facilitated strain-level resolution of *B.*

subtilis and *B. cereus*. Enzymatic pretreatment has also been used to facilitate more complete extraction of cell contents. Two of the most commonly used enzymes are trypsin (Balážová et al. 2014b; Krasny et al. 2014; Krishnamurthy et al. 1996) and lysozyme (Giebel et al. 2008; Vargha et al. 2006). Balážová et al. (2014) used trypsin to digest cells of *S. aureus, Staphylococcus haemolyticus,* and *B. subtilis* and applied a 2-min microwave irradiation after digestion. Strain-level differentiation was achieved for *S. aureus* and *S. haemolyticus*, and improved for *B. subtilis* based on ecotypes. An increase in the number of strain-specific peaks was also observed when using this microwave-assisted tryptic digestion sample preparation method (Balážová et al. 2014b). Abdelhamid et al. (2014) used a preconcentration technique, ultrasound-enhanced surfactant-assisted dispersive liquid–liquid microextraction (UESA-DLLME) technique, with *Pseudomonas aeruginosa* and *S. aureus*. This method improved the number of biomarker peaks and identification. Nanoparticles, for example, zinc oxide nanoparticles modified with polymethyl methacrylate, have also been synthesized for extracting bacteria from aqueous samples, which enhanced the sensitivity and quality of MALDI-MS spectra for characterizing bacteria such as *S. aureus* and *P. aeruginosa* (Gedda et al. 2014). Modification of the surface of the MALDI target has also been applied to enhance characterization of bacteria. For example, Hasan et al. (2014) and Gopal et al. (2013) demonstrated that using titanium chips as MALDI target and with appropriate surface pretreatments (using heat treatment at different temperatures), the chips could selectively capture either *P. aeruginosa* or *S. aureus*, leading to an improvement in spectrum quality for these two bacteria. Such modifications might be applied in the future to enhance strain-level characterization.

Data Acquisition

Modern MALDI TOF mass spectrometers can be operated with software that facilitates completely the automatic acquisition of spectra. User-defined parameters influence laser power, peak evaluation strategies, mass spectra accumulation, and laser movement on each sample. While convenient and supportive of high throughput applications, automation has been shown to affect performance. In particular, the mode of data acquisition (i.e., automated (Eddabra et al. 2012; Khot et al. 2012) or manual (Khot et al. 2012; Schumaker et al. 2012) data acquisition) may affect the taxonomic resolution of MALDI-TOF MS profiling technique. Schumaker et al. (2012) rigorously examined the effects of data acquisition modes on spectrum quality and reproducibility. Results suggested that manual data acquisition yielded more reproducible and higher quality mass spectra. Similar results were reported by Balážová et al. (2014).

Though manual data acquisition yielded more reproducible spectra, automated data acquisition is still desirable when there is a large quantity of analyses to perform, especially in clinical labs. Recent work in our lab showed that automated data acquisition can be optimized to yield spectra with reproducibility comparable to

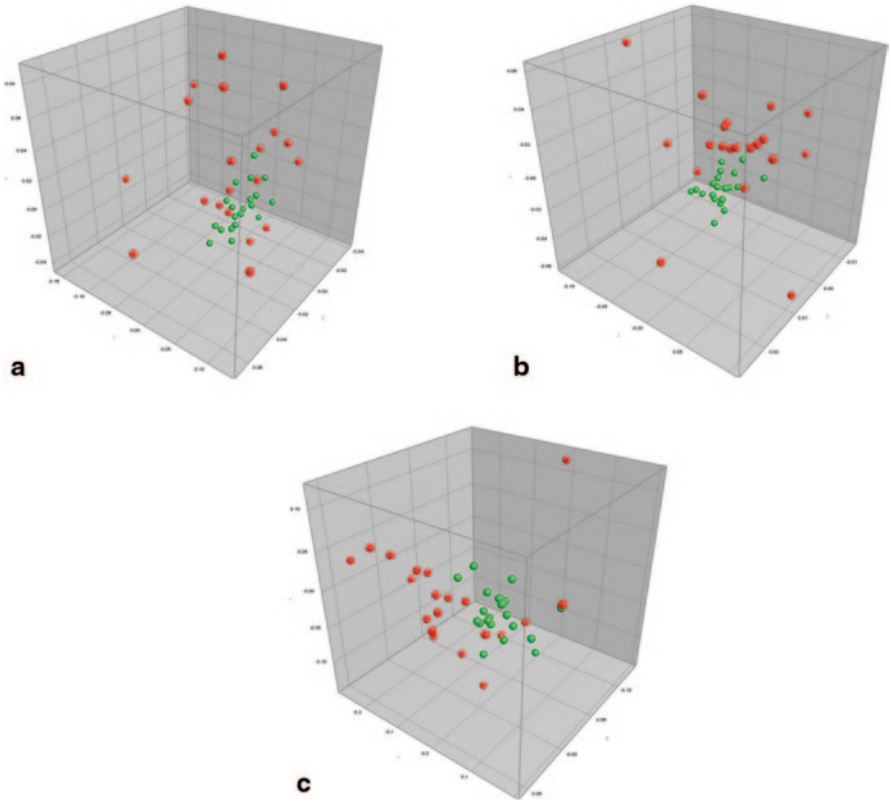


Fig. 6.7 A designed experiments approach to optimization of automated acquisition of MALDI spectra increased the reproducibility of spectra obtained from *Pseudomonas aeruginosa* (**a**), *Klebsiella pneumoniae* (**b**), and *Serratia marcescens* (**c**). (Adapted from Zhang et al. 2014a, creative commons attribution PLOS)

those obtained manually using a statistical design of experiments approach (Zhang et al. 2014a). Results showed that the reproducibility of replicate *P. aeruginosa* spectra increased from 90% to 97% by optimizing the automated data acquisition conditions. Similar results were reported for *Klebsiella pneumonia* (94% before optimization vs. 98% after optimization) and *Serratia marcescens* (85% before optimization vs. 94% after optimization; Zhang et al. 2014a) (Fig. 6.7).

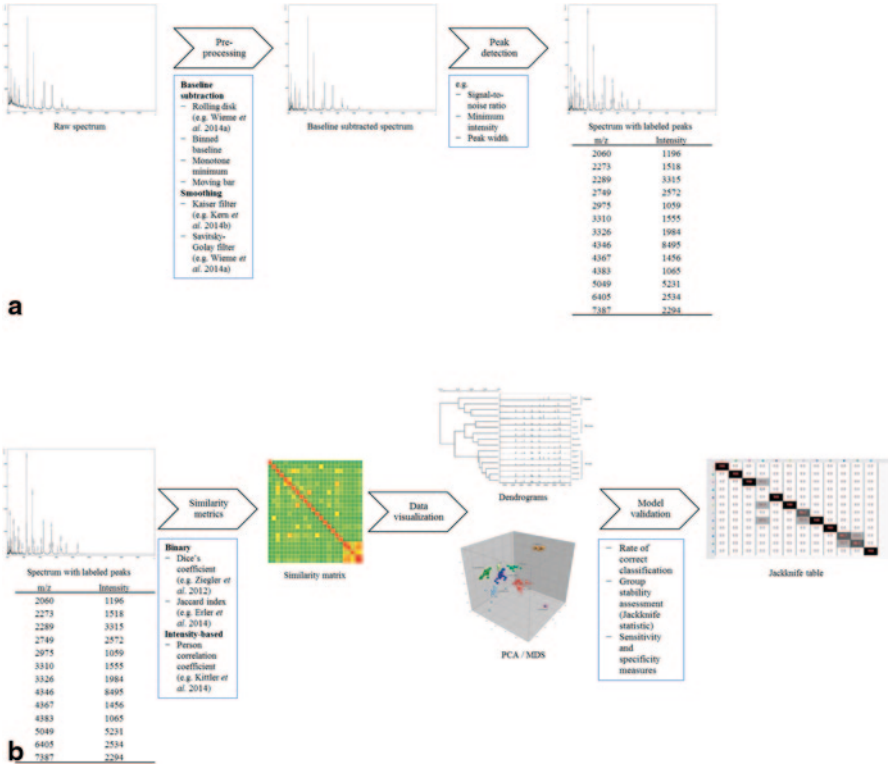


Fig. 6.8 Raw MALDI spectra are typically subjected to multiple data preprocessing (**a**) as well as postprocessing and model validation (**b**) steps. No standard workflow has been widely adopted, in spite of the fact that modifications to many steps may affect performance of the method, particularly when applied at the strain level

Data Analysis

Generally, the workflow to analyze bacteria using MALDI-TOF MS includes preprocessing, postprocessing, and model validation (Fig. 6.8). The objective of preprocessing is to reduce noise in the spectra. A variety of software has been used to preprocess raw mass spectra, such as FlexAnalysis (Bruker Daltonik, GmbH, Germany; Schrodl et al. 2012), BioNumerics (Applied Maths; Kittler et al. 2014; Wieme et al. 2014a), and DataExplorer software (Applied Biosystems, Foster City, CA, USA; Vanlaere et al. 2008).

Three common steps in the preprocessing procedures are baseline subtraction, smoothing, and peak detection. For each step, algorithms as well as parameter values associated with each algorithm can be varied. For example, baseline subtraction is used to flatten the varying base profile of a spectrum. When using BioNumerics to preprocess data, baseline subtraction can be conducted using a rolling disk algorithm (Wieme et al. 2014a) with a user-specified radius of the disk. Other baseline subtraction algorithms include monotone minimum, moving bar, and binning.

Smoothing is used to average data points with neighboring points as in a time-series of data to further reduce noise. For smoothing, Savitsky–Golay filters (Wieme et al. 2014a) and Kaiser filters (Kern et al. 2014b) have been used. Peak detection is used to separate real peaks from false peaks representing noise. A user-defined signal-to-noise ratio is usually applied for software to automatically pick peaks. Smaller signal-to-noise ratios may support higher taxonomic resolution when using peak-picking software which requires the user to specify a minimum signal-to-noise ratio such as in BioNumerics (*Applied Maths*; personal communication).

Postprocessing procedures include calculation of similarity coefficients to establish a set of classes (clusters). In the literature, similarity coefficients of spectra have been calculated in mainly two ways: using binary- and intensity-based measurements. Binary-based analysis of similarity considers only the presence/absence of peaks without considering their intensities. Such measurements can be achieved by calculating Jaccard index (Erler et al. 2014) or the Dice coefficient (Hazen et al. 2009). In contrast, intensity-based coefficients of similarity consider both the presence/absence of peaks and the peak intensity. The Pearson correlation coefficient (Kittler et al. 2014) has been used for intensity-based analysis. With regard to the dendrogram type, the unweighted pair group method with arithmetic mean (UPGMA), average linkage method (Quintela-Baluja et al. 2013), and single-link agglomerative algorithm (Andres-Barrao et al. 2013) have been used in MALDI-fingerprint studies. Though various algorithms have been applied to pre- and postprocessing of mass spectra, few studies have directly compared the effects of these algorithms on taxonomic resolution. Model validation is used to further evaluate MALDI performance on group (e.g., genus, species, and strain) separation. For example, jackknife analysis which reports the percentage of correct and false identifications has been used to quantify the stability of groups of fingerprints of MSSA and MRSA (Goldstein et al. 2013). Receiver operating characteristic (ROC) curves, which illustrates the performance of a binary classification test, were also used in the literature to evaluate *E. coli* isolate classification by spectrum similarity (Christner et al. 2014). In addition, sensitivity (true positive rate) and specificity (true negative rate) measures, which are also used for validation of the performance of binary classifications, have been reported (De Bruyne et al. 2011).

Summary

MALDI-TOF MS, shown to be a highly effective tool to characterize bacteria at the MALDI BioTyper CA System, was recently (November 2013) approved by the US FDA for the identification of Gram-negative bacterial colonies cultured from human specimens. Besides applications in clinical laboratories for identification of Gram-negative bacteria, studies have also shown that MALDI-TOF MS fingerprint-based methods can successfully characterize Gram-positive bacteria, bacteria isolated from various environments, and, in some cases, characterize bacteria at the subspecies and strain levels. Rapid innovation and advances in this area increase

the likelihood that strain-level applications will receive similar regulatory approval in the future.

Strain-level characterization using MALDI-TOF MS has included three objectives: strain categorization, strain differentiation, and strain identification. The taxonomic resolution reported for each objective has varied considerably. Generally, strain identification requires the highest taxonomic resolution, while strain categorization requires relatively low resolution. Factors, such as culture media, sample preparation method, data acquisition, and data analysis, have been shown to affect the limits of taxonomic resolution achieved. Many efforts have been made to increase taxonomic resolution. Different sample preparation methods and sample deposition have been compared with regard to their effects on taxonomic resolution. Treatments of samples, such as using enzymes to help break cell walls or using microwave radiation to help extract proteins, have also been employed. In addition, several novel approaches have been developed to increase the taxonomic resolution of MALDI-TOF MS fingerprint-based methods, including stable isotope-dependent methods that are described in more detail elsewhere in this book.

Overall, MALDI-TOF profiling of bacteria has shown remarkable promise at the genus, species, and strain levels for various bacteria; however, the limits of taxonomic resolution of this technique may impede its broader implementation. Additional efforts to maximize the taxonomic resolution of this method by optimizing experimental conditions—from culture condition through data analysis—are warranted.

Acknowledgments

The authors thank Beau Grothendick for his assistance in preparing the manuscript. Funding for this work was provided by the New College of Interdisciplinary Arts and Sciences at Arizona State University.

References

Abdelhamid HN, Bhaisare ML, Wu HF. Ceria nanocubic-ultrasonication assisted dispersive liquid-liquid microextraction coupled with matrix assisted laser desorption/ionization mass spectrometry for pathogenic bacteria analysis. Talanta. 2014;120:208–17.

Alispahic M, Hummel K, Jandreski-Cvetkovic D, Nobauer K, Razzazi-Fazeli E, Hess M, Hess C. Species-specific identification and differentiation of *Arcobacter, Helicobacter* and *Campylobacter* by full-spectral matrix-associated laser desorption/ionization time of flight mass spectrometry analysis. J Med Microbiol. 2010;59(3):295–301.

Alispahic M, Christensen H, Bisgaard M, Hess M, Hess C. MALDI-TOF mass spectrometry confirms difficulties in separating species of the *Avibacterium* genus. Avian Pathol. 2014;43(3):258–63.

AlMasoud N, Xu Y, Nicolaou N, Goodacre R. Optimization of matrix assisted desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) for the characterization of *Bacillus* and *Brevibacillus* species. Anal Chim Acta. 2014;840:49–57.

Anderson NW, Buchan BW, Riebe KM, Parsons LN, Gnacinski S, Ledeboer NA. Effects of solid-medium type on routine identification of bacterial isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50(3):1008–13.

Andres-Barrao C, Benagli C, Chappuis M, Perez RO, Tonolla M, Barja F. Rapid identification of acetic acid bacteria using MALDI-TOF mass spectrometry fingerprinting. Syst Appl Microbiol. 2013;36(2):75–81.

Anhalt JP, Fenselau C. Identification of bacteria using mass-spectrometry. Anal Chem. 1975;47(2):219–25.

Arnold RJ, Reilly JP. Fingerprint matching of *E-coli* strains with matrix-assisted laser desorption ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. Rapid Commun Mass Sp. 1998;12(10):630–6.

Arnold RJ, Reilly JP. Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. Anal Biochem. 1999;269(1):105–12.

Ayyadurai S, Flaudrops C, Raoult D, Drancourt M. Rapid identification and typing of *Yersinia pestis* and other *Yersinia* species by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. Bmc Microbiol. 2010;10:285.

Bain R, Cronk R, Hossain R, Bonjour S, Onda K, Wright J, Yang H, Slaymaker T, Hunter P, Pruss-Ustun A, Bartram J. Global assessment of exposure to faecal contamination through drinking water based on a systematic review. Trop Med Int Health. 2014;19(8):917–27.

Balada-Llasat JM, Kamboj K, Pancholi P. Identification of *Mycobacteria* from solid and liquid media by matrix-assisted laser desorption ionization-time of flight mass spectrometry in the clinical laboratory. J Clin Microbiol. 2013;51(9):2875–9.

Balážová T, Makovcova J, Šedo O, Slany M, Faldyna M, Zdrahal Z. The influence of culture conditions on the identification of *Mycobacterium* species by MALDI-TOF MS profiling. Fems Microbiol Lett. 2014a;353(1):77–84.

Balážová T, Sedo O, Stefanic P, Mandic-Mulec I, Vos M, Zdrahal Z. Improvement in *Staphylococcus* and *Bacillus* strain differentiation by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling by using microwave-assisted enzymatic digestion. Rapid Commun Mass Sp. 2014b;28(17):1855–61.

Bernardo K, Pakulat N, Macht M, Krut O, Seifert H, Fleer S, Hunger F, Kronke M. Identification and discrimination of *Staphylococcus aureus* strains using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. Proteomics. 2002;2(6):747–53.

Bohme K, Fernandez-No IC, Barros-Velazquez J, Gallardo JM, Canas B, Calo-Mata P. Rapid species identification of seafood spoilage and pathogenic Gram-positive bacteria by MALDI-TOF mass fingerprinting. Electrophoresis. 2011a;32(21):2951–65.

Bohme K, Fernandez-No IC, Gallardo JM, Canas B, Calo-Mata P. Safety assessment of fresh and processed seafood products by MALDI-TOF mass fingerprinting. Food Bioprocess Tech. 2011b;4(6):907–18.

Bohme K, Fernandez-No IC, Pazos M, Gallardo JM, Barros-Velazquez J, Canas B, Calo-Mata P. Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing versus MALDI-TOF MS fingerprinting. Electrophoresis. 2013;34(6):877–87.

Cain TC, Lubman DM, Weber WJ. Differentiation of bacteria using protein profiles from matrix-assisted laser-desorption ionization time-of-flight mass-spectrometry. Rapid Commun Mass Sp. 1994;8(12):1026–30.

Carbonnelle E, Beretti JL, Cottyn S, Quesne G, Berche P, Nassif X, Ferroni A. Rapid identification of *Staphylococci* isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization—time of flight mass spectrometry. J Clin Microbiol. 2007;45(7):2156–61.

Chen P, Lu Y, Harrington PB. Biomarker profiling and reproducibility study of MALDI-MS measurements of *Escherichia coli* by analysis of variance-principal component analysis. Anal Chem. 2008;80(5):1474–81.

Chen JHK, Ho PL, Kwan GSW, She KKK, Siu GKH, Cheng VCC, Yuen KY, Yam WC. Direct bacterial identification in positive blood cultures by use of two commercial matrix-as-

sisted laser desorption ionization-time of flight mass spectrometry systems. J Clin Microbiol. 2013;51(6):1733–9.
Chiu TC. Recent advances in bacteria identification by matrix-assisted laser desorption/ionization mass spectrometry using nanomaterials as affinity probes. Int J Mol Sci. 2014;15(5):7266–80.
Christner M, Trusch M, Rohde H, Kwiatkowski M, Schluter H, Wolters M, Aepfelbacher M, Hentschke M. Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of shiga-toxigenic *Escherichia coli*. Plos ONE. 2014;9(7).
Coltella L, Mancinelli L, Onori M, Lucignano B, Menichella D, Sorge R, Raponi M, Mancini R, Russo C. Advancement in the routine identification of anaerobic bacteria by MALDI-TOF mass spectrometry. Eur J Clin Microbiol. 2013;32(9):1183–92.
Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. Fems Microbiol Rev. 2012;36(2):380–407.
Dallagassa CB, Huergo LF, Stets MI, Pedrosa FO, Souza EM, Cruz LM, Assis FEA, Wolf S, Volanski W, Picheth G, Pigatto-Denardi CP, Farah SMSS, Fadel-Picheth CMT. Matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis of *Escherichia coli* categories. Genet Mol Res. 2014;13(1):716–22.
De Bruyne K, Slabbinck B, Waegeman W, Vauterin P, De Baets B, Vandamme P. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. Syst Appl Microbiol. 2011;34(1):20–9.
Demirev PA, Fenselau C. Mass spectrometry for rapid characterization of microorganisms. Annu Rev Anal Chem. 2008;1:71–93.
Demirev PA, Antoine MD, Feldman AB, Kowalski P, Lin LS. Top-down proteomics for microorganism identification in mixtures. Abstr Pap Am Chem. S. 2005;230:U373–U373.
Dickinson DN, Duc MT L, Haskins WE, Gornushkin I, Winefordner JD, Powell DH, Venkateswaran K. Species differentiation of a diverse suite of *Bacillus* spores by mass spectrometry-based protein profiling. Appl Environ Microb. 2004;70(1):475–82.
Dieckmann R, Helmuth R, Erhard M, Malorny B. Rapid classification and identification of *Salmonellae* at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microb. 2008;74(24):7767–78.
Donohue MJ, Smallwood AW, Pfaller S, Rodgers M, Shoemaker JA. The development of a matrix-assisted laser desorption/ionization mass spectrometry-based method for the protein fingerprinting and identification of *Aeromonas* species using whole cells. J Microbiol Meth. 2006;65(3):380–9.
Dubois D, Leyssene D, Chacornac JP, Kostrzewa M, Schmit PO, Talon R, Bonnet R, Delmas J. Identification of a variety of *Staphylococcus* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2010;48(3):941–5.
Eddabra R, Prevost G, Scheftel JM. Rapid discrimination of environmental *Vibrio* by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Microbiol Res. 2012;167(4):226–30.
Eells SJ, McKinnell JA, Wang AA, Green NL, Whang D, O'Hara P, Brown ML, Miller LG. A comparison of clinical outcomes between healthcare-associated infections due to community-associated methicillin-resistant *Staphylococcus aureus* strains and healthcare-associated methicillin-resistant *S. aureus* strains. Epidemiol Infect. 2013;141(10):2140–48.
Epstein L, Hunter JC, Arwady MA, Tsai V, Stein L, Gribogiannis M, Frias M, Guh AY, Laufer AS, Black S, Pacilli M, Moulton-Meissner H, Rasheed JK, Avillan JJ, Kitchel B, Limbago BM, MacCannell D, Lonsway D, Noble-Wang J, Conway J, Conover C, Vernon M, Kallen AJ. New Delhi metallo-beta-lactamase-producing carbapenem-resistant *Escherichia coli* associated with exposure to duodenoscopes. Jama-J Am Med Assoc. 2014;312(14):1447–55.
Erler R, Wichels A, Heinemeyer EA, Hauk G, Hippelein M, Reyes NT, Gerdts G. VibrioBase: A MALDI-TOF MS database for fast identification of *Vibrio* spp. that are potentially pathogenic in humans. Syst Appl Microbiol. 2014;38:16–25.
Everley RA, Mott TM, Wyatt SA, Toney DM, Croley TR. Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species. J Am Soc Mass Spectrom. 2008;19(11):1621–28.

Fagerquist CK, Garbus BR, Miller WG, Williams KE, Yee E, Bates AH, Boyle S, Harden LA, Cooley MB, Mandrell RE. Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. Anal Chem. 2010;82(7):2717–25.

Fenselau CC. Rapid characterization of microorganisms by mass spectrometry—what can be learned and how? J Am Soc Mass Spectr. 2013;24(8):1161–6.

Fenselau C, Russell S, Swatkoski S, Edwards N. Proteomic strategies for rapid characterization of micro-organisms. Eur J Mass Spectrom. 2007;13(1):35–9.

Fernandez-No IC, Bohme K, Diaz-Bao M, Cepeda A, Barros-Velazquez J, Calo-Mata P. Characterisation and profiling of *Bacillus subtilis*, *Bacillus cereus* and *Bacillus licheniformis* by MALDI-TOF mass fingerprinting. Food Microbiol. 2013;33(2):235–42.

Ferreira L, Castano SV, Sanchez-Juanes F, Gonzalez-Cabrero S, Menegotto F, Orduna-Domingo A, Gonzalez-Buitrago JM, Munoz-Bellido JL. Identification of *Brucella* by MALDI-TOF mass spectrometry. Fast and reliable identification from agar plates and blood cultures. Plos ONE. 2010a;5 (12).

Ferreira L, Sanchez-Juanes F, Gonzalez-Avila M, Cembrero-Fucinos D, Herrero-Hernandez A, Gonzalez-Buitrago JM, Munoz-Bellido JL. Direct identification of urinary tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2010b;48(6):2110–5.

Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ. *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. J Food Protect. 2014;77(1):150–70.

Ford BA, Burnham CAD. Optimization of routine identification of clinically relevant gram-negative bacteria by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry and the Bruker Biotyper. J Clin Microbiol. 2013;51(5):1412–20.

Freiwald A, Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry. Nat Protoc. 2009;4(5):732–42.

Fujinami Y, Kikkawa HS, Kurosaki Y, Sakurada K, Yoshino M, Yasuda J. Rapid discrimination of *Legionella* by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Microbiol Res. 2011;166(2):77–86.

Gao BL, Gupta RS. Microbial systematics in the post-genomics era. Anton Leeuw Int J G. 2012;101(1):45–54.

Ge MY, Li B, Wang L, Tao ZY, Mao SF, Wang YL, Xie GL, Sun GC. Differentiation in MALDI-TOF MS and FTIR spectra between two pathovars of *Xanthomonas oryzae*. Spectrochim Acta A. 2014;133:730–4.

Gedda G, Abdelhamid HN, Khan MS, Wu HF. ZnO nanoparticle-modified polymethyl methacrylate-assisted dispersive liquid-liquid microextraction coupled with MALDI-MS for rapid pathogenic bacteria analysis. Rsc Adv. 2014;4(86):45973–83.

Gentile RC, Shukla S, Shah M, Ritterband DC, Engelbert M, Davis A, Hu DN. Microbiological spectrum and antibiotic sensitivity in endophthalmitis a 25-year review. Ophthalmology. 2014;121(8):1634–42.

Ghyselinck J, Van Hoorde K, Hoste B, Heylen K, De Vos P. Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. J Microbiol Methods. 2011;86(3):327–36.

Giebel RA, Fredenberg W, Sandrin TR. Characterization of environmental isolates of *Enterococcus* spp. by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Water Res. 2008;42(4–5):931–40.

Giebel R, Worden C, Rust SM, Kleinheinz GT, Robbins M, Sandrin TR. Microbial fingerprinting using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS): applications and challenges. Adv Appl Microbiol. 2010;71:149–84.

Goldstein JE, Zhang L, Borror CM, Rago JV, Sandrin TR. Culture conditions and sample preparation methods affect spectrum quality and reproducibility during profiling of *Staphylococcus aureus* with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Lett Appl Microbiol. 2013;57(2):144–50.

Goncalves JL, Tomazi T, Barreiro JR, Braga PAD, Ferreira CR, Araujo JP, Eberlin MN, dos Santos MV. Identification of *Corynebacterium* spp. isolated from bovine intramammary infections by

matrix-assisted laser desorption ionization-time of flight mass spectrometry. Vet Microbiol. 2014;173(1–2):147–51.

Gopal J, Hasan N, Wu HF. Fabrication of titanium based MALDI bacterial chips for rapid, sensitive and direct analysis of pathogenic bacteria. Biosens Bioelectron. 2013;39(1):57–63.

Goto K, Yamamoto M, Asahara M, Tamura T, Matsumura M, Hayashimoto N, Makimura K. Rapid identification of *Mycoplasma pulmonis* isolated from laboratory mice and rats using matrix-assisted laser desorption ionization time-of-flight mass spectrometry. J Vet Med Sci. 2012;74(8):1083–6.

Grosse-Herrenthey A, Maier T, Gessler F, Schaumann R, Bohnel H, Kostrzewa M, Kruger M. Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Anaerobe. 2008;14(4):242–9.

Grundmann H, Livermore DM, Giske CG, Canton R, Rossolini GM, Campos J, Vatopoulos A, Gniadkowski M, Toth A, Pfeifer Y, Jarlier V, Carmeli Y, Grp CW. Carbapenem-non-susceptible *Enterobacteriaceae* in Europe: conclusions from a meeting of national experts. Eurosurveillance. 2010;15(46):22–34.

Han SK, Hong Y, Kwak HL, Kim ES, Kim MJ, Shrivastav A, Oh MH, Kim HY. Identification of lactic acid bacteria in pork meat and pork meat products using SDS-PAGE, 16S rRNA gene sequencing and MALDI-TOF mass spectrometry. J Food Safety. 2014;34(3):224–32.

Hasan N, Gopal J, Wu HF. Surface pretreatment effects on titanium chips for the adhesion of pathogenic bacteria in the MALDI-TOF MS. Appl Surf Sci. 2014;314:52–63.

Hazen TH, Martinez RJ, Chen Y, Lafon PC, Garrett NM, Parsons MB, Bopp CA, Sullards MC, Sobecky PA. Rapid identification of *Vibrio parahaemolyticus* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microb. 2009;75(21):6745–56.

Helmel M, Marchetti-Deschmann M, Allmaier G. Improved sample preparation for intact cell mass spectrometry (biotyping) of mycelium samples taken from a batch fermentation process of *Penicillium chrysogenum*. Rapid Commun Mass Sp. 2014;28(8):957–64.

Hettick JM, Kashon ML, Slaven JE, Ma Y, Simpson JP, Siegel PD, Mazurek GN, Weissman DN. Discrimination of intact *Mycobacteria* at the strain level: a combined MALDI-TOF MS and biostatistical analysis. Proteomics. 2006;6(24):6416–25.

Holland RD, Wilkes JG, Rafii F, Sutherland JB, Persons CC, Voorhees KJ, Lay JO. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Commun Mass Sp. 1996;10(10):1227–32.

Holland RD, Wilkes JG, Cooper WM, Alusta P, Williams A, Pearce B, Beaudoin M, Buzatu D. Thymol treatment of bacteria prior to matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis aids in identifying certain bacteria at the subspecies level. Rapid Commun Mass Sp: RCM. 2014a;28 (23):2617–2626.

Holland TL, Arnold C, Fowler VG. Clinical management of *Staphylococcus aureus* bacteremia: a review. Jama-J Am Med Assoc. 2014b;312(13):1330–41.

Horneffer V, Haverkamp J, Janssen HG, Notz R. MALDI-TOF-MS analysis of bacterial spores: wet heat-treatment as a new releasing technique for biomarkers and the influence of different experimental parameters and microbiological handling. J Am Soc Mass Spectr. 2004;15(10):1444–54.

Hsu YMS, Burnham CAD. MALDI-TOF MS identification of anaerobic bacteria: assessment of pre-analytical variables and specimen preparation techniques. Diagn Micr Infec Dis. 2014;79(2):144–8.

Hsueh PR, Kuo LC, Chang TC, Lee TF, Teng SH, Chuang YC, Teng LJ, Sheng WH. Evaluation of the Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system for identification of blood isolates of *Acinetobacter* species. J Clin Microbiol. 2014;52(8):3095–100.

Ikryannikova LN, Filimonova AV, Malakhova MV, Savinova T, Filimonova O, Ilina EN, Dubovickaya VA, Sidorenko SV, Govorun VM. Discrimination between *Streptococcus pneumoniae*

and *Streptococcus mitis* based on sorting of their MALDI mass spectra. Clin Microbiol Infec. 2013;19(11):1066–71.

Ilina EN, Borovskaya AD, Serebryakova MV, Chelysheva VV, Momynaliev KT, Maier T, Kostrzewa M, Govorun VM. Application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for the study of *Helicobacter pylori*. Rapid Commun Mass Spectrom. 2010;24(3):328–34.

Jackson KA, Edwards-Jones V, Sutton CW, Fox AJ. Optimisation of intact cell MALDI method for fingerprinting of methicillin-resistant *Staphylococcus aureus*. J Microbiol Meth. 2005;62(3):273–84.

Johnson RA, Wichem DW. Applied multivariate statistical analysis. 6th edn. New Jersey: Pearson Prentice Hall; 2007

Karas M, Bachmann D, Bahr U, Hillenkamp F. Matrix-assisted ultraviolet-laser desorption of nonvolatile compounds. Int J Mass Spectrom. 1987;78:53–68.

Karger A, Ziller M, Bettin B, Mintel B, Schares S, Geue L. Determination of serotypes of shiga toxin-producing *Escherichia coli* isolates by intact cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microb. 2011;77(3):896–905.

Kern CC, Usbeck JC, Vogel RF, Behr J. Optimization of matrix-assisted-laser-desorption-ionization-time-of-flight mass spectrometry for the identification of bacterial contaminants in beverages. J Microbiol Meth. 2013;93(3):185–91.

Kern CC, Vogel RF, Behr J. Differentiation of *Lactobacillus brevis* strains using matrix-assisted-laser-desorption-ionization-time-of-flight mass spectrometry with respect to their beer spoilage potential. Food Microbiol. 2014a;40:18–24.

Kern CC, Vogel RF, Behr J. Identification and differentiation of brewery isolates of *Pectinatus* sp by matrix-assisted-laser-desorption-ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Eur Food Res Technol. 2014b;238(5):875–80.

Keys CJ, Dare DJ, Sutton H, Wells G, Lunt M, McKenna T, McDowall M, Shah HN. Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. Infect Genet Evol. 2004;4(3):221–42.

Khot PD, Couturier MR, Wilson A, Croft A, Fisher MA. Optimization of matrix-assisted-laser-desorption-ionization-time-of-flight mass spectrometry analysis for bacterial identification. J Clin Microbiol. 2012;50(12):3845–52.

Klein S, Zimmermann S, Kohler C, Mischnik A, Alle W, Bode KA. Integration of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in blood culture diagnostics: a fast and effective approach. J Med Microbiol. 2012;61(3):323–31.

Kopcakova A, Stramova Z, Kvasnova S, Godany A, Perhacova Z, Pristas P. Need for database extension for reliable identification of bacteria from extreme environments using MALDI TOF mass spectrometry. Chem Pap. 2014;68(11):1435–42.

Krasny L, Rohlova E, Ruzickova H, Jiri SA, Hynek R, Hochel I. Differentiation of *Cronobacter* spp. by tryptic digestion of the cell suspension followed by MALDI-TOF MS analysis. J Microbiol Meth. 2014;98:105–13.

Krishnamurthy T, Ross PL, Rajamani U. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Sp. 1996;10(8):883–8.

Kuehl B, Marten SM, Bischoff Y, Brenner-Weiss G, Obst U. MALDI-TOF mass spectrometry-multivariate data analysis as a tool for classification of reactivation and non-culturable states of bacteria. Anal Bioanal Chem. 2011;401(5):1593–600.

Lacroix C, Gicquel A, Sendid B, Meyer J, Accoceberry I, Francois N, Morio F, Desoubeaux G, Chandenier J, Kauffmann-Lacroix C, Hennequin C, Guitard J, Nassif X, Bougnoux ME. Evaluation of two matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) systems for the identification of *Candida* species. Clin Microbiol Infec. 2014;20(2):153–8.

Lartigue MF, Hery-Arnaud G, Haguenoer E, Domelier AS, Schmit PO, van der Mee-Marquet N, Lanotte P, Mereghetti L, Kostrzewa M, Quentin R. Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by matrix-assisted-laser-desorption-ionization-time-of-flight mass spectrometry. J Clin Microbiol. 2009;47(7):2284–7.

Lasch P, Fleige C, Stammler M, Layer F, Nubel U, Witte W, Werner G. Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of *Enterococcus faecium* and *Staphylococcus aureus* isolates. J Microbiol Meth. 2014;100:58–69.

Lazcka O, Del Campo FJ, Munoz FX. Pathogen detection: a perspective of traditional methods and biosensors. Biosens Bioelectron. 2007;22(7):1205–17.

Li WJ, Raoult D, Fournier PE. Bacterial strain typing in the genomic era. Fems Microbiol Rev. 2009;33(5):892–916.

Livermore DM. Fourteen years in resistance. Int J Antimicrob Ag. 2012;39(4):283–94.

Lohmann C, Sabou M, Moussaoui W, Prevost G, Delarbre JM, Candolfi E, Gravet A, Letscher-Bru V. Comparison between the Biflex III-Biotyper and the Axima-SARAMIS systems for yeast identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2013;51(4):1231–6.

Lotz A, Ferroni A, Beretti JL, Dauphin B, Carbonnelle E, Guet-Revillet H, Veziris N, Heym B, Jarlier V, Gaillard JL, Pierre-Audigier C, Frapy E, Berche P, Nassif X, Bille E. Rapid identification of mycobacterial whole cells in solid and liquid culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2010;48(12):4481–6.

Lundquist M, Caspersen MB, Wikstrom P, Forsman M. Discrimination of *Francisella tularensis* subspecies using surface enhanced laser desorption ionization mass spectrometry and multivariate data analysis. Fems Microbiol Lett. 2005;243(1):303–10.

Majcherczyk PA, McKenna T, Moreillon P, Vaudaux P. The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*. Fems Microbiol Lett. 2006;255(2):233–9.

Martin-Loeches I, Diaz E, Valles J. Risks for multidrug-resistant pathogens in the ICU. Curr Opin Crit Care. 2014;20(5):516–24.

Martinez RM, Bauerle ER, Fang FC, Butler-Wu SM. Evaluation of three rapid diagnostic methods for direct identification of microorganisms in positive blood cultures. J Clin Microbiol. 2014;52(7):2521–9.

Masarikova M, Mrackova M, Sedlinska M. Application of matrix-assisted laser desorption ionization time-of-flight mass spectrometry in identification of stallion semen bacterial contamination. J Equine Vet Sci. 2014;34(6):833–6.

Mather CA, Rivera SF, Butler-Wu SM. Comparison of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of *Mycobacteria* using simplified protein extraction protocols. J Clin Microbiol. 2014;52(1):130–8.

Morgan DJ, Kaye KS, Diekema DJ. Reconsidering Isolation Precautions for Endemic Methicillin-Resistant *Staphylococcus aureus* and Vancomycin-Resistant *Enterococcus*. Jama-J Am Med Assoc. 2014;312(14):1395–6.

Moura H, Woolfitt AR, Carvalho MG, Pavlopoulos A, Teixeira LM, Satten GA, Barr JR. MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. Fems Immunol Med Mic. 2008;53(3):333–42.

Murray PR. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: usefulness for taxonomy and epidemiology. Clin Microbiol Infec. 2010;16(11):1626–30.

Nagy E. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: a new possibility for the identification and typing of anaerobic bacteria. Future Microbiol. 2014;9(2):217–33.

Nagy E, Becker S, Soki J, Urban E, Kostrzewa M. Differentiation of division I (cfiA-negative) and division II (cfiA-positive) *Bacteroides fragilis* strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Med Microbiol. 2011;60(11):1584–90.

Nagy E, Urban E, Becker S, Kostrzewa M, Voros A, Hunyadkurti J, Nagy I. MALDI-TOF MS fingerprinting facilitates rapid discrimination of phylotypes I, II and III and *Propionibacterium acnes*. Anaerobe. 2013;20:20–6.

Nakano S, Matsumura Y, Kato K, Yunoki T, Hotta G, Noguchi T, Yamamoto M, Nagao M, Ito Y, Takakura S, Ichiyama S. Differentiation of vanA-positive *Enterococcus faecium* from vanA-

negative *E. faecium* by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Int J Antimicrob Ag. 2014;44(3):256–9.

Niederman MS. Treatment options for nosocomial pneumonia due to MRSA. J Infection. 2009;59:S25–31.

Niyompanich S, Jaresitthikunchai J, Srisanga K, Roytrakul S, Tungpradabkul S. Source-identifying biomarker ions between environmental and clinical *Burkholderia pseudomallei* using whole-cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Plos ONE. 2014;9(6).

Nonnemann B, Tvede M, Bjarnsholt T. Identification of pathogenic microorganisms directly from positive blood vials by matrix-assisted laser desorption/ionization time of flight mass spectrometry. Apmis. 2013;121(9):871–7.

Nucera DM, Lomonaco S, Costa A, Morra P, Grassi MA. Diagnostic performance of rep-PCR as a rapid subtyping method for *Listeria monocytogenes*. Food Anal Method. 2013;6(3):868–71.

Pavlovic M, Huber I, Konrad R, Busch U. Application of MALDI-TOF MS for the identification of food borne bacteria. Open Microbiol J. 2013;7:135–41.

Pennanec X, Dufour A, Haras D, Rehel K. A quick and easy method to identify bacteria by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun Mass Sp. 2010;24(3):384–92.

Pereyre S, Tardy F, Renaudin H, Cauvin E, Machado LDN, Tricot A, Benoit F, Treilles M, Bebear C. Identification and subtyping of clinically relevant human and ruminant *Mycoplasmas* by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2013;51(10):3314–23.

Perez-Boto D, Lopez-Portoles JA, Simon C, Echeita MA. Application of restriction fragment length polymorphism-polymerase chain reaction-flaA and resistotype to identify potential undiagnosed outbreaks of *Campylobacteriosis* in Spain. Enferm Infec Micr Cl. 2014;32(7):428–33.

Pribil PA, Patton E, Black G, Doroshenko V, Fenselau C. Rapid characterization of *Bacillus* spores targeting species-unique peptides produced with an atmospheric pressure matrix-assisted laser desorption/ionization source. J Mass Spectrom. 2005;40(4):464–74.

Quintela-Baluja M, Bohme K, Fernandez-No IC, Morandi S, Alnakip ME, Caamano-Antelo S, Barros-Velazquez J, Calo-Mata P. Characterization of different food-isolated *Enterococcus* strains by MALDI-TOF mass fingerprinting. Electrophoresis. 2013;34(15):2240–50.

Rajakaruna L, Hallas G, Molenaar L, Dare D, Sutton H, Encheva V, Culak R, Innes I, Ball G, Sefton AM, Eydmann M, Kearns AM, Shah HN. High throughput identification of clinical isolates of *Staphylococcus aureus* using MALDI-TOF-MS of intact cells. Infect Genet Evol. 2009;9(4):507–13.

Regecova I, Pipova M, Jevinova P, Maruskova K, Kmet V, Popelka P. Species identification and antimicrobial resistance of coagulase-negative *Staphylococci* isolated from the meat of sea fish. J Food Sci. 2014;79(5):M898–902.

Rodriguez-Sanchez B, Sanchez-Carrillo C, Ruiz A, Marin M, Cercenado E, Rodriguez-Creixems M, Bouza E. Direct identification of pathogens from positive blood cultures using matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Clin Microbiol Infec. 2014;20(7):O421–7.

Rossello GAM, Rodriguez MPG, Leonardo ROD, Domingo AO, Perez MAB. Procedure for microbial identification based on matrix-assisted laser desorption/ionization-time of flight mass spectrometry from screening-positive urine samples. Apmis. 2014;122(9):790–5.

Ruelle V, EL Moualij B, Zorzi W, Ledent P, De Pauw E. Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Sp. 2004;18(18):2013–9.

Russell SC, Edwards N, Fenselau C. Detection of plasmid insertion in *Escherichia coli* by MALDI-TOF mass spectrometry. Anal Chem. 2007;79(14):5399–406.

Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal Chem. 2001;73(4):746–50.

Ryzhov V, Hathout Y, Fenselau C. Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl Environ Microb. 2000;66(9):3828–34.

Salaun S, Kervarec N, Potin P, Haras D, Piotto M, La Barre S. Whole-cell spectroscopy is a convenient tool to assist molecular identification of cultivatable marine bacteria and to investigate their adaptive metabolism. Talanta. 2010;80(5):1758–70.

Samb-Ba B, Mazenot C, Gassama-Sow A, Dubourg G, Richet H, Hugon P, Lagier JC, Raoult D, Fenollar F. MALDI-TOF identification of the human gut microbiome in people with and without diarrhea in Senegal. Plos ONE. 2014;9(5).

Sanchez-Juanes F, Ferreira L, de la Vega PA, Valverde A, Barrios ML, Rivas R, Mateos PF, Martinez-Molina E, Gonzalez-Buitrago JM, Trujillo ME, Velazquez E. MALDI-TOF mass spectrometry as a tool for differentiation of *Bradyrhizobium* species: application to the identification of *Lupinus* nodulating strains. Syst Appl Microbiol. 2013;36(8):565–71.

Sandrin TR, Goldstein JE, Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom Rev. 2013;32(3):188–217.

Sauer S, Freiwald A, Maier T, Kube M, Reinhardt R, Kostrzewa M, Geider K. Classification and identification of bacteria by mass spectrometry and computational analysis. Plos ONE. 2008;3(7).

Sauer S, Kliem M. Mass spectrometry tools for the classification and identification of bacteria. Nat Rev Microbiol. 2010;8(1):74–82.

Schafer MO, Genersch E, Funfhaus A, Poppinga L, Formella N, Bettin B, Karger A. Rapid identification of differentially virulent genotypes of *Paenibacillus larvae*, the causative organism of American foulbrood of honey bees, by whole cell MALDI-TOF mass spectrometry. Vet Microbiol. 2014;170(3–4):291–7.

Schrodl W, Heydel T, Schwartze VU, Hoffmann K, Grosse-Herrenthey A, Walther G, Alastruey-Izquierdo A, Rodriguez-Tudela JL, Olias P, Jacobsen ID, de Hoog GS, Voigt K. Direct analysis and identification of pathogenic *Lichtheimia* species by matrix-assisted laser desorption ionization-time of flight analyzer-mediated mass spectrometry. J Clin Microbiol. 2012;50(2):419–27.

Schrottner P, Rudolph WW, Eing BR, Bertram S, Gunzer F. Comparison of VITEK2, MALDI-TOF MS, and 16S rDNA sequencing for identification of *Myroides odoratus* and *Myroides odoratimimus*. Diagn Micr Infec Dis. 2014;79(2):155–9.

Schulthess B, Bloemberg GV, Zbinden R, Bottger EC, Hombach M. Evaluation of the Bruker MALDI Biotyper for identification of gram-positive rods: development of a diagnostic algorithm for the clinical laboratory. J Clin Microbiol. 2014;52(4):1089–97.

Schumaker S, Borror CM, Sandrin TR. Automating data acquisition affects mass spectrum quality and reproducibility during bacterial profiling using an intact cell sample preparation method with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Sp. 2012;26(3):243–53.

Šedo O, Sedlacek I, Zdrahal Z. Sample preparation methods for MALDI-MS profiling of bacteria. Mass Spectrom Rev. 2011;30(3):417–34.

Šedo O, Vavrova A, Vad'urova M, Tvrzova L, Zdrahal Z. The influence of growth conditions on strain differentiation within the *Lactobacillus acidophilus* group using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling. Rapid Commun Mass Sp. 2013;27(24):2729–36.

Setlhare G, Malebo N, Shale K, Lues R. Identification of airborne microbiota in selected areas in a health-care setting in South Africa. Bmc Microbiol. 2014;14:100

Shah HN, Rajakaruna L, Ball G, Misra R, Al-Shahib A, Fang M, Gharbia SE. Tracing the transition of methicillin resistance in sub-populations of *Staphylococcus aureus*, using SELDI-TOF mass spectrometry and artificial neural network analysis. Syst Appl Microbiol. 2011;34(1):81–6.

Shan XY, Zhang J, Zhou HJ, Zhu BQ, Xu L, Shao ZJ, Jiang BF. Genetic diversity of *Neisseria meningitidis* serogroup C ST-4821 in China based on multiple-locus variable number tandem repeat analysis. Plos ONE. 2014;9(11)e111866.

Siegrist TJ, Anderson PD, Huen WH, Kleinheinz GT, McDermott CM, Sandrin TR. Discrimination and characterization of environmental strains of *Escherichia coli* by matrix-assisted laser

desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). J Microbiol Meth. 2007;68(3):554–62.

Spanu V, Scarano C, Cossu F, Pala C, Spanu C, De Santis EPL. Antibiotic resistance traits and molecular subtyping of *Staphylococcus aureus* isolated from raw sheep milk cheese. J Food Sci. 2014;79(10):M2066–71.

Sparbier K, Lange C, Jung J, Wieser A, Schubert S, Kostrzewa M. MALDI Biotyper-based rapid resistance detection by stable-isotope labeling. J Clin Microbiol. 2013;51(11):3741–8.

Staley JT. The bacterial species dilemma and the genomic-phylogenetic species concept. Philos T R Soc B. 2006;361(1475):1899–909.

Stephan R, Cernela N, Ziegler D, Pfluger V, Tonolla M, Ravasi D, Fredriksson-Ahomaa M, Hachler H. Rapid species specific identification and subtyping of *Yersinia enterocolitica* by MALDI-TOF Mass spectrometry. J Microbiol Meth. 2011;87(2):150–3.

Stets MI, Pinto AS, Huergo LF, de Souza EM, Guimaraes VF, Alves AC, Steffens MBR, Monteiro RA, Pedrosa FD, Cruz LM. Rapid identification of bacterial isolates from wheat roots by high resolution whole cell MALDI-TOF MS analysis. J Biotechnol. 2013;165(3–4):167–74.

Sun LW, Teramoto K, Sato H, Torimura M, Tao H, Shintani T. Characterization of ribosomal proteins as biomarkers for matrix-assisted laser desorption/ionization mass spectral identification of *Lactobacillus plantarum*. Rapid Commun Mass Sp. 2006;20(24):3789–98.

Teramoto K, Sato H, Sun L, Torimura M, Tao H. A simple intact protein analysis by MALDI-MS for characterization of ribosomal proteins of two genome-sequenced lactic acid bacteria and verification of their amino acid sequences. J Proteome Res. 2007a;6(10):3899–907.

Teramoto K, Sato H, Sun L, Torimura M, Tao H, Yoshikawa H, Hotta Y, Hosoda A, Tamura H. Phylogenetic classification of *Pseudomonas putida* strains by MALDI-MS using ribosomal subunit proteins as biomarkers. Anal Chem. 2007b;79(22):8712–9.

Thevenon F, Regier N, Benagli C, Tonolla M, Adatte T, Wildi W, Pote J. Characterization of fecal indicator bacteria in sediments cores from the largest freshwater lake of Western Europe (Lake Geneva, Switzerland). Ecotox Environ Safe. 2012;78:50–6.

Toh-Boyo GM, Wulff SS, Basile F. Comparison of sample preparation methods and evaluation of intra- and intersample reproducibility in bacteria MALDI-MS profiling. Anal Chem. 2012;84(22):9971–80.

Trafny EA, Lewandowski R, Stepinska M, Kaliszewski M. Biological threat detection in the air and on the surface: how to define the risk. Arch Immunol Ther Ex. 2014;62(4):253–61.

Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. Appl Environ Microb. 2005;71(1):58–64.

Vanlaere E, LiPuma JJ, Baldwin A, Henry D, De Brandt E, Mahenthiralingam E, Speert D, Dowson C, Vandamme P. *Burkholderia latens* sp nov., *Burkholderia diffusa* sp nov., *Burkholderia arboris* sp nov., *Burkholderia seminalis* sp nov and *Burkholderia metallica* sp nov., novel species within the *Burkholderia cepacia* complex. Int J Syst Evol Micr. 2008;58:1580–90.

Vargha M, Takats Z, Konopka A, Nakatsu CH. Optimization of MALDI-TOF MS for strain level differentiation of *Arthrobacter* isolates. J Microbiol Meth. 2006;66(3):399–409.

Walker J, Fox AJ, Edwards-Jones V, Gordon DB. Intact cell mass spectrometry (ICMS) used to type methicillin-resistant *Staphylococcus aureus*: media effects and inter-laboratory reproducibility. J Microbiol Meth. 2002;48(2–3):117–26.

Wang WP, Xi HY, Huang M, Wang J, Fan M, Chen Y, Shao HF, Li XJ. Performance of mass spectrometric identification of bacteria and yeasts routinely isolated in a clinical microbiology laboratory using MALDI-TOF MS. J Thorac Dis. 2014;6(5):524–33.

Wang XH, Zhang G, Fan YY, Yang X, Sui WJ, Lu XX. Direct identification of bacteria causing urinary tract infections by combining matrix-assisted laser desorption ionization-time of flight mass spectrometry with UF-1000i urine flow cytometry. J Microbiol Meth. 2013;92(3):231–5.

Welham KJ, Domin MA, Scannell DE, Cohen E, Ashton DS. The characterization of micro-organisms by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Sp. 1998;12(4):176–80.

Welker M, Moore ERB. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst Appl Microbiol. 2011;34(1):2–11.

Wenning M, Breitenwieser F, Konrad R, Huber I, Busch U, Scherer S. Identification and differentiation of food-related bacteria: a comparison of FTIR spectroscopy and MALDI-TOF mass spectrometry. J Microbiol Meth. 2014;103:44–52.

Wensing A, Zimmermann S, Geider K. Identification of the corn pathogen *Pantoea stewartii* by mass spectrometry of whole-cell extracts and its detection with novel PCR primers. Appl Environ Microb. 2010;76(18):6248–56.

Werno AM, Christner M, Anderson TP, Murdoch DR. Differentiation of *Streptococcus pneumoniae* from nonpneumococcal *Streptococci* of the *Streptococcus mitis* Group by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50(9):2863–7.

Wessels E, Schelfaut JJG, Bernards AT, Claas ECJ. Evaluation of several biochemical and molecular techniques for identification of *Streptococcus pneumoniae* and *Streptococcus pseudopneumoniae* and their detection in respiratory samples. J Clin Microbiol. 2012;50(4):1171–7.

Wieme AD, Spitaels F, Aerts M, De Bruyne K, Van Landschoot A, Vandamme P. Effects of growth medium on matrix-assisted laser desorption-ionization time of flight mass spectra: a case study of acetic acid bacteria. Appl Environ Microb. 2014a;80(4):1528–38.

Wieme AD, Spitaels F, Aerts M, De Bruyne K, Van Landschoot A, Vandamme P. Identification of beer-spoilage bacteria using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Int J Food Microbiol. 2014b;185:41–50.

Williams MG, Palandra J, Shobe EM. Rapid determination of rat plasma uridine levels by HPLC-ESI-MS utilizing the Captiva ™ filter plates for sample preparation. Biomed Chromatogr. 2003;17(4):215–8.

Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, Aepfelbacher M, Christner M. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. Int J Med Microbiol. 2011;301(1):64–8.

Wynne C, Fenselau C, Demirev PA, Edwards N. Top-down identification of protein biomarkers in bacteria with unsequenced genomes. Anal Chem. 2009;81(23):9633–42.

Xiao JH, Liu Y, Wang MG, Jiang CH, You XX, Zhu CM. Detection of *Mycoplasma pneumoniae* P1 subtype variations by denaturing gradient gel electrophoresis. Diagn Micr Infec Dis. 2014;78(1):24–8.

Xiao D, Tao XX, Wang P, Liu GD, Gong YN, Zhang HF, Wang HB, Zhang JZ. Rapid and high-throughput identification of recombinant bacteria with mass spectrometry assay. Biomed Environ Sci. 2014a;27(4):250–8.

Xiao D, Ye CY, Zhang HF, Kan B, Lu JX, Xu JG, Jiang XG, Zhao F, You YH, Yan XM, Wang DC, Hu Y, Zhang MJ, Zhang JZ. The construction and evaluation of reference spectra for the identification of human pathogenic microorganisms by MALDI-TOF MS. Plos ONE. 2014b;9(9):e106312.

Xiao D, Zhang HF, He LH, Peng XH, Wang YD, Xue GH, Su P, Zhang JZ. High natural variability bacteria identification and typing: *Helicobacter pylori* analysis based on peptide mass fingerprinting. J Proteomics. 2014c;98:112–22.

Zampieri D, Santos VG, Braga PAC, Ferreira CR, Ballottin D, Tasic L, Basso AC, Sanches BV, Pontes JHF, da Silva BP, Garboggini FF, Eberlin MN, Tata A. Microorganisms in cryopreserved semen and culture media used in the in vitro production (IVP) of bovine embryos identified by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS). Theriogenology. 2013;80(4):337–45.

Zarate MS, Romano V, Nievas J, Smayevsky J. Utility of MALDI-TOF MS for the identification of anaerobic bacteria. Rev Argent Microbiol. 2014;46(2):98–102.

Zautner AE, Masanta WO, Tareen AM, Weig M, Lugert R, Gross U, Bader O. Discrimination of multilocus sequence typing-based *Campylobacter jejuni* subgroups by MALDI-TOF mass spectrometry. Bmc Microbiol. 2013;13:247.

Zeller-Peronnet V, Brockmann E, Pavlovic M, Timke M, Busch U, Huber I. Potential and limitations of MALDI-TOF MS for discrimination within the species *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides*. J Verbrauch Lebensm. 2013;8(3):205–14.

Zhang L, Borror CM, Sandrin TR. A designed experiments approach to optimization of automated data acquisition during characterization of bacteria with MALDI-TOF mass spectrometry. Plos ONE. 2014a;9(3):e92720.

Zhang YF, Liu YY, Ma QW, Song YQ, Zhang Q, Wang XY, Chen F. Identification of *Lactobacillus* from the saliva of adult patients with caries using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Plos ONE. 2014b;9(8).

Ziegler D, Mariotti A, Pfluger V, Saad M, Vogel G, Tonolla M, Perret X. In situ identification of plant-invasive bacteria with MALDI-TOF mass spectrometry. Plos ONE. 2012;7(5):e37189.

Part II
Subspecies Discrimination

Chapter 7
Modulation of the Discriminatory Power of MALDI-TOF MS Profiling for Distinguishing Between Closely Related Bacterial Strains

Ondrej Šedo and Zbyněk Zdráhal

Introduction

Sample preparation protocols that are recommended by the two main providers of commercially available systems (Biotyper from Bruker and VITEK-MS from Biomérieux) rely on alpha-cyano-4-hydroxycinnamic acid (CHCA) as a universal matrix for all types of bacteria, yeasts, and fungi (Maier and Kostrzewa 2007; Bizzini and Greub 2010; Martiny et al. 2012). Several different sample preparation protocols for matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) profiling of bacteria have been proposed (reviewed by Šedo et al. 2011). Many of these provide even better mass resolution, signal-to-noise ratio, more peaks in the high-mass range, and other improved mass spectral features compared with the manufacturer-recommended preparation protocols using CHCA; however, robustness of the standard method and the possibility of automatic mass spectral acquisition are two fundamental aspects of method applicability in routine clinical laboratory practice. These positive features mostly stem from the formation of a highly homogeneous layer of small crystals of CHCA. As a consequence, other factors affecting mass spectral quality, such as growth medium, amount of cells mounted on the MALDI target, unmanageable degrees of instability or differences in performance between MALDI-TOF MS systems, have only a minor influence and do not need to be considered when identification to the bacterial species level is required. An acceptable number of species-specific signals eclipse the low number of peaks or their changed relative intensities that result from these factors (Mazzeo et al. 2005; Valentine et al. 2005; Grosse-Herrenthey et al. 2008; Moura et al. 2008; Ilina et al. 2010).

In contrast, at the level of subspecies or individual strains, accurate discrimination is usually dependent on a significantly lower number of specific signals

Z. Zdráhal (✉) · O. Šedo
Research Group Proteomics, CEITEC—Central European Institute of Technology, Masaryk University, Kamenice 5, 62500 Brno, Czech Rep.
e-mail: zdrahal@sci.muni.cz

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_7

(Krishnamurthy et al. 1996; Lynn et al. 1999; Fedele et al. 1999). This fact increases demands on the reproducibility of sample preparation methods and analytical procedures. At the subspecies and strain levels, users have encountered limits in the discriminatory power of MALDI-TOF MS profiling more often than at the genus/species level. In our laboratory, we have already faced various issues where MALDI-TOF MS profiling analysis, conducted under a common sample preparation protocol (Freiwald and Sauer 2009), yielded insufficient output; possibilities of modulating the discriminatory power of the method were therefore examined. In the first two cases highlighted in this chapter, the influence of cultivation conditions on differentiation between closely related strains of the *Lactobacillus acidophilus* group and selected *Mycobacterium* spp. was assessed. The third and the fourth studies were aimed at developing variants of the standard sample preparation protocol, which enabled substantial modulation of the discriminatory power of the method, as demonstrated by distinguishing between closely related species of the *Acinetobacter calcoaceticus–Acinetobacter baumannii* complex (ACB complex), *Staphylococcus aureus* strains, and *Bacillus subtilis* ecotypes.

Differentiation Between Strains of *L. acidophilus* Group Grown under Different Cultivation Conditions

Background

Although cultivation conditions have a profound influence on the bacterial proteome, MALDI-TOF MS bacterial fingerprints remain, in most cases, largely unaffected by changes in the cultivation medium, time, or temperature (Wunschel et al. 2005; Parisi et al. 2008; Sauer et al. 2008). This is due to the fact that when compounds in a mixture are subjected to MALDI-TOF MS analysis, mutual suppression of ionization leads to the detection of only a limited set of components. These usually are not only the most abundant, but also the basic components and compounds with relatively low molecular weights. In the case of the whole bacterial proteome, these criteria are mostly fulfilled by ribosomal and DNA-binding proteins that represent “house-keeping” cellular components present in bacterial cells regardless of growth stage or cultivation conditions.

The Biotyper system (Bruker Daltonics) reflects this fact by omitting requirements for specific cultivation conditions in protocols for successful bacterial identification to the species level (Maier and Kostrzewa 2007). In the VITEK MS identification system (Biomérieux), the database overcomes mass spectral variability due to cultivation conditions by implementing “superspectra,” i.e., consensus spectra obtained from individual strains cultivated under different conditions, while the Biotyper database is constructed on the basis of repeated analyses of one culture. The influence of cultivation conditions on MALDI-TOF mass spectral quality has already been examined in numerous studies. These demonstrated that bacterial cells at different growth stages or even cells showing different morphologies exhibited

practically identical MALDI-TOF MS fingerprints (Lee et al. 2003; Wunschel et al. 2005). Although the repeatable detection of a core set of signals sufficient for correct bacterial species assignment under different conditions has been reported in several other papers, there have also been published studies that draw different conclusions. Identification success rates to the species level were affected by cultivation conditions particularly for staphylococci (Anderson et al. 2012), mycobacteria (Lotz et al. 2010; Balážová et al. 2014a), and bacilli (Ryzhov et al. 2000; Horneffer et al. 2004). For strain differentiation, comprehensive studies on cultivation conditions have not yet been conducted. For this reason, we carried out a series of analyses of 17 strains from the *L. acidophilus* group, grown under different cultivation conditions (Šedo et al. 2013b) . Apart from their close relationships, these strains were selected due to their different probiotic properties and uses in the dairy industry (Heller 2001; Monteagudo-Mera et al. 2011).

Protocol

Seventeen selected strains, representing six different species, from the *L. acidophilus* group (*L. acidophilus* CCM 4833^T and BCCM 8151; *L. amylovorus* CCM 4380^T, CCM 4381, and CCM 4382; *L. crispatus* CCM 7010^T, CCM 7776, and CCM 7777; *L. delbrueckii* ssp. *bulgaricus* CCM 4289, CCM 4290, and CCM 7190^T; *L. helveticus* CCM 3806, CCM 4287, and 7193^T; *L. jensenii* CCM 7560^T, CCM 7653, and CCM 7778) were involved in this study. All strains were cultivated in triplicate under standard cultivation conditions (at 37 °C De Man-Rogosa-Sharpe/MRS/broth for 24 h). In addition, two other cultivation media (M132, blood agar), two different cultivation temperatures (22 and 30 °C), and two longer cultivation times (2 days, 7 days) were tested on all 17 strains (see Table 7.1). In total, 327 samples were prepared (due to limited or no growth of some of the strains at low temperatures, mostly those of *L. crispatus,* some of the scheduled samples were not available). A standard extraction protocol adapted from Freiwald and Sauer (2009) was followed (inactivation by 70 % EtOH, formic acid/acetonitrile extraction and overlaying of the dried bacterial extract with CHCA solution), while each sample was deposited on

Table 7.1 Seven combinations of cultivation conditions tested on 17 strains

Medium	Temperature (°C)	Growth time (days)
MRS	37	1
MRS	30	1
MRS	22	1
MRS	37	2
MRS	37	7
M132	37	1
Blood agar	37	1

MRS De Man-Rogosa-Sharpe

three wells of the stainless steel MALDI target. Five independent MALDI-TOF mass spectra, comprising 1000 laser shots each, were acquired from each of the three wells (15,000 shots per sample in total). Within an individual well, the laser was directed automatically according to a predefined lattice raster. To assess mutual similarities between MALDI-TOF MS fingerprints, four different data evaluation methods that are routinely accessed by MALDI Biotyper and Bruker Daltonics software users were tested: visual inspection of strain-specific signals, cluster analysis (carried out within the Biotyper software by using the Pearson's product moment coefficient as a measure of similarity and the unweighted pair group average linked method (UPGMA) as a grouping method), Biotyper scoring, and Biotyper sub-typing algorithm.

Results

Initially, identification at the species level using the commercially available database was accomplished for all cultivation variants. Practically all strains were assigned to the correct species, regardless of the cultivation medium and temperature used. A score indicating species identification (>2.000) was achieved for 80 % of isolates. Even if the threshold score was not reached, the highest identification score always belonged to the correct species for 14 % of isolates. The remaining samples (6 %) yielded negative identification outputs, which were usually related to acquisition failures resulting from low quality mass spectral data from 7-day cultures where cells were difficult to separate from the growth medium.

These results demonstrated the robustness of discrimination between closely related species of the *L. acidophilus* group. Proteome changes related to cultivation conditions (Hussain et al. 2009) did not affect proteins detected in MALDI-TOF MS "whole cell" fingerprints of these lactobacilli to an extent that might have influenced their identification scores. However, at the level of strains belonging to the same species, the situation was found to be different. Six of the 17 tested strains were found to lack strain-specific peaks in their mass spectra, regardless of the culture conditions. Identification of strains grown under diverse cultivation conditions using the database of reference spectra created under standard conditions yielded correct identification results for 73 % of samples; an identification success rate of greater than 90 % was only achieved for seven of the strains. Interestingly, employment of the subtyping algorithm (incorporated in the Biotyper software), which adjusts the statistical weight of signals according to strain specificity, did not significantly improve the identification success rate at the strain level.

The clearest demonstration of the influence of cultivation conditions on the quality of MALDI-TOF MS fingerprints was obtained by cluster analysis. As all strains were cultivated in triplicate, the possibility of their discrimination could be derived by grouping replicate samples in branches of the dendrograms. Analyses of three *L. jensenii* strains are shown in Fig. 7.1, where *L. jensenii* strain CCM 7560^T (Fig. 7.1a) was clearly distinguishable from the remaining two strains regardless of the cultivation conditions. However, a mixed cluster analysis of samples of *L. jensenii* strains CCM 7653 and CCM 7778 indicated that their discrimination was not a straightfor-

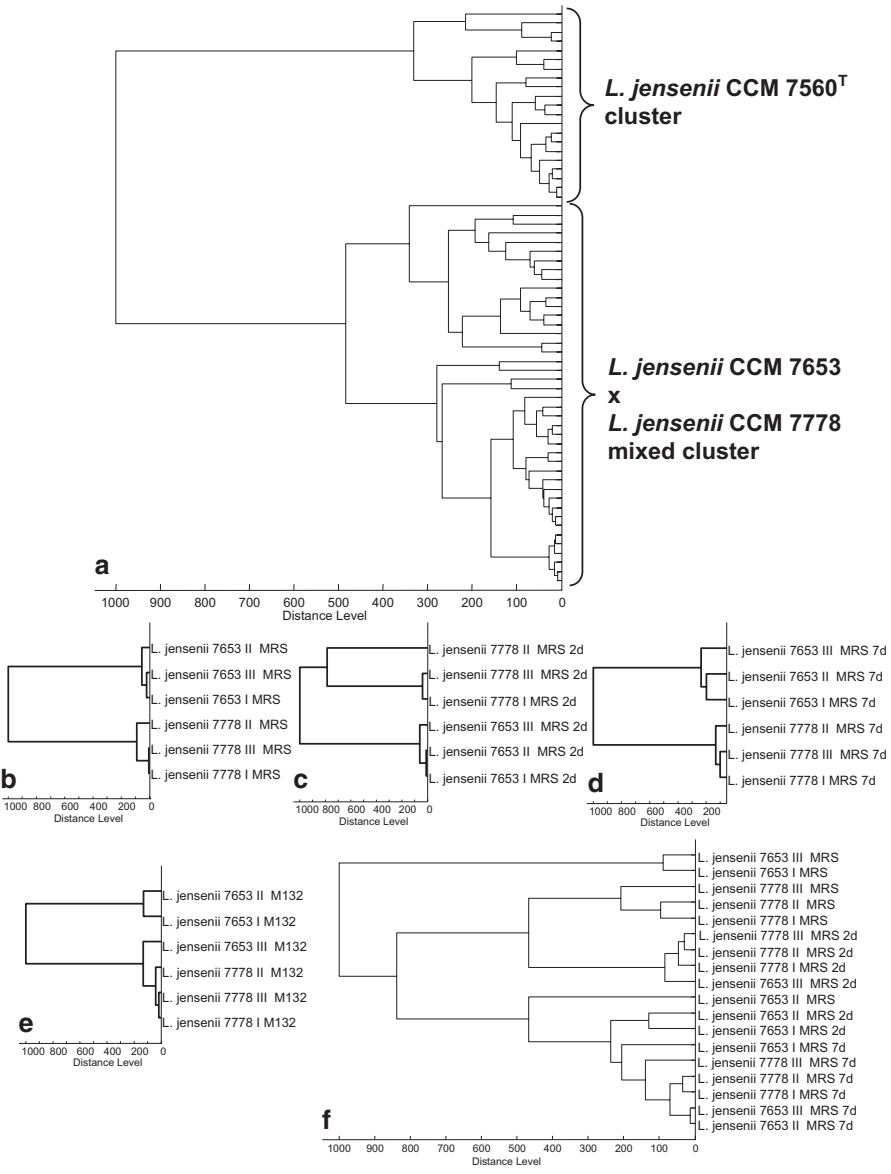


Fig. 7.1 Dendrogram of all analyses of three *L. jensenii* strains grown under various conditions (**a**). Dendrograms demonstrating the ability to distinguish between strains *L. jensenii* CCM 7653 and CCM 7778 grown **b** at 37 °C on MRS medium for 1 day, **c** at 37 °C on MRS medium for 2 days, **d** at 37 °C on MRS medium for 7 days, **e** on M132 medium for 1 day, and **f** at 37 °C on MRS for various growth times. Three independent cultivations and analyses (marked as "I," "II", and "III") were performed

ward task. Figure 7.1b, c, d demonstrates that when cultivation time was the same for all samples, and MRS agar was used as a cultivation medium, discrimination between these two strains was possible. However, when cultivation times were different, the strains remained indistinguishable (Fig. 7.1f). Similarly, when a different cultivation medium (M132) was used, the strains remained indistinguishable (Fig. 7.1e).

Summary

These experiments showed that proper control over cultivation conditions might increase the probability of successful strain differentiation. The selection of cultivation medium, temperature, and cultivation time may serve as factors worth optimizing for particular strains of interest, since their variation modulates the discriminatory power of the method at the strain level.

The Influence of Cultivation Conditions on Distinguishing Between Strains of Selected *Mycobacterium* spp.

Background

Although for most known bacteria, a common sample preparation protocol (Freiwald and Sauer 2009) routinely yields MALDI-TOF MS fingerprints, mycobacteria have been shown to be different. Because of specific properties of the mycobacterial cell wall, which is stabilized by peptidoglycans esterified with mycolic acids, both the inactivation of cells using 70 % ethanol, or protein extraction by the matrix solution or a formic acid/acetonitrile mixture, do not lead to satisfactory results. For that reason, both main providers of the commercially available systems have introduced alternative sample preparation protocols recommended for MALDI-TOF MS profiling of mycobacteria; these involve cell inactivation and disruption by heat and micro-bead treatment. Nevertheless, even after following these recommendations, we have not achieved positive identification outputs, and a large proportion of samples did not yield MALDI-TOF mass spectral signals at all. For that reason, we conducted a comprehensive study on selected rapidly growing *Mycobacterium* spp. to establish a reliable protocol for their analysis and discrimination, including differentiation between strains belonging to the same *Mycobacterium* spp. (Balážová et al. 2014a).

Protocol

The protocol for mycobacterial sample preparation, as recommended by Bruker Daltonics (Timke and Kostrzewa 2011), involves cell inactivation by heat treat-

ment (30 min at 95 °C and cell disruption by zirconia/silica microbeads), and this was followed. Initial selection of cultivation conditions was carried out using two strains (*Mycobacterium phlei* DSM 43239 and *M. smegmatis* CAPM 5040) grown in triplicate in four different cultivation media: Middlebrook 7H9 broth (M7H9) with OADC medium enrichment (oleic acid–albumin–dextrose–catalase), Middlebrook 7H10 (M7H10) with OADC medium enrichment, Herrold's egg yolk medium (HEYM), and Lesslie (L) medium (Stonebrink with methyl violet). The strains were grown over five different cultivation times (3, 5, 7, 10, and 12 days). Ultimately, only a limited range of cultivation conditions (HEYM and M7H9 media with cultivation times of 3, 5, and 7 days) were applied for three strains of *M. phlei* (CCM 5639, DSM 43214, and DSM 43239) and seven strains of *M. smegmatis* (CCM 2300, CCM 4622, CCM 1693, DSM 43059, DSM 43286, DSM 43756, and CAPM 5040). Each sample was deposited on three wells of the stainless steel MALDI target. Five independent MALDI-TOF mass spectra, comprising 1000 laser shots each, were acquired from each of the three wells (15,000 shots per sample in total). Within an individual well, the laser was directed automatically according to a predefined lattice raster. The species identification was accomplished by using default settings of the Biotyper software and strain discrimination was carried out within the Biotyper software by using the Pearson's product moment coefficient as a measure of similarity and the UPGMA as a grouping method.

Results

In the case of the two initially selected mycobacterial strains, signal quality appropriate for automatic mass spectral acquisition was obtained for only 30 % of samples. In the remaining cases, some signals were obtained by manual laser navigation over the sample spots, although this approach is not appropriate for clinical laboratories that are dependent on validated automatic data acquisition. The limited mass spectral quality and striking differences between spectra obtained after cultivation on different media (for comparison see Fig. 7.2) strongly influenced the identification success rate. When the mass spectra were compared with the commercially available extension of the Biotyper database (this database version contains reference spectra for mycobacteria prepared according to their recommended protocol), the identification success rate varied notably, as demonstrated in Table 7.2. Interestingly, the best conditions for automatic mass spectral acquisition significantly differed from those optimized for identification at the species level; this makes it difficult to select one medium overall for the best analytical output. The influence of cultivation time (3–7 days) on acquisition and identification success rates was not so prominent, as the influence of cultivation medium. However, similar to our previous study with lactobacilli (Šedo et al. 2013b), strains analyzed after long cultivation periods (10 and 12 days) yielded lower quality MALDI-TOF mass spectral data because complete separation of bacterial cells from the cultivation medium was not achievable caused by firm attachment of the cells to the medium.

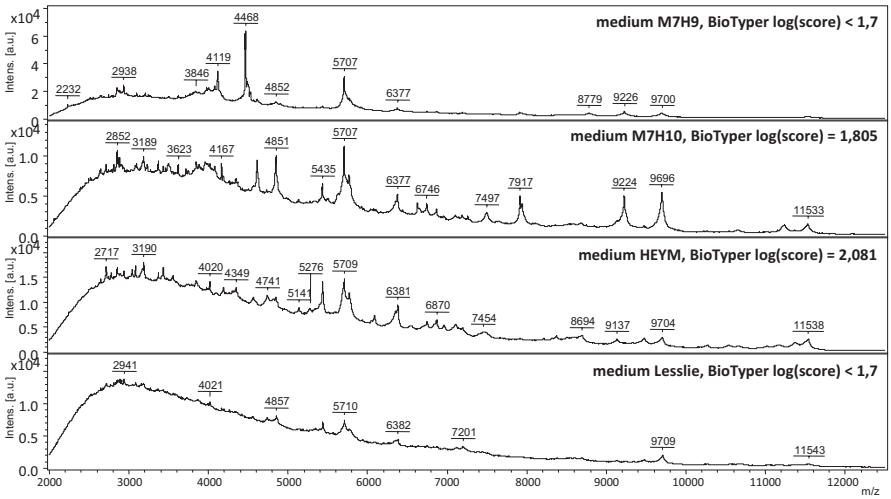


Fig. 7.2 Comparison between MALDI-TOF mass spectra and Biotyper log (scores) of *Mycobacterium phlei* DSM 43239 cultivated for 10 days on different media

Table 7.2 The influence of culture media on the possibility of automatic mass spectra acquisition and on the identification to genus/species level (the success rate in %; sum of results from triplicate samples of *M. phlei* DSM 43239 and *M. smegmatis* CAPM 5040 cultivated for 3, 5, 7, 10, and 12 days)

Medium	Automatic acquisition (%)	Identification at genus level (%)	Identification at species level (%)
M7H9	70	47	17
M7H10	30	33	7
L	0	33	13
HEYM	20	60	47

HEYM Herrold's egg yolk medium

Cultivation conditions also influenced the potential to distinguish between individual strains of the same species. The dendrograms in Fig. 7.3 demonstrate that although *M. phlei* strains cultivated on M7H9 did not yield spectra enabling their discrimination, after being cultured on HEYM medium, the strains were distinguishable by cluster analysis. This fact was derived from grouping of three triplicates of DSM 43239 samples prepared at monthly intervals. On the other hand, *M. smegmatis* strains remained indistinguishable regardless of the cultivation medium (not shown).

Summary

The selection of a proper cultivation medium can serve as a tool to increase the discriminatory power of the method for mycobacteria at the strain level. Optimization of the sample preparation protocol for distinguishing between selected groups

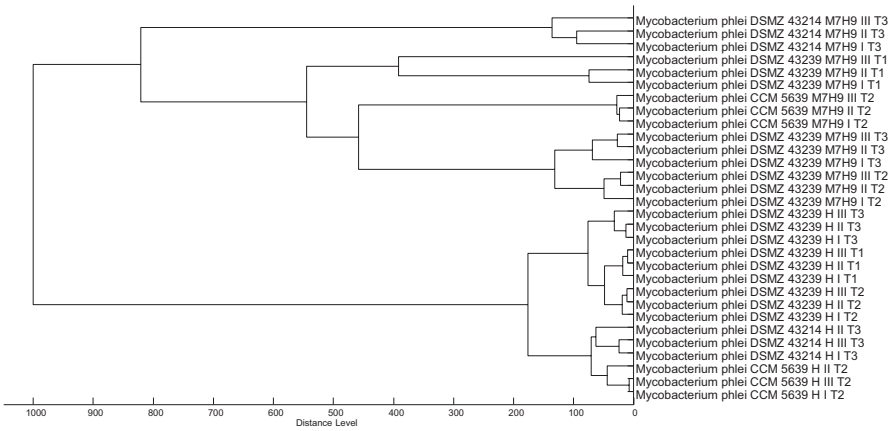


Fig. 7.3. Dendrogram calculated on the basis of cluster analysis of MALDI-TOF mass spectra of three *Mycobacterium phlei* strains analyzed after 5 days of growth. Repeated analyses of strains are marked as "I," "II," and "III" and cultivation terms as "T1," "T2," and "T3"

of strains of interest may be complicated by the need to identify conditions that are simultaneously suitable for automatic mass spectral acquisition and for identification at the genus/species level.

Modulation of the Discriminatory Power Using an Alternative Matrix Solution

Background

Apart from an overall satisfactory performance of the method, some studies also highlighted problematic groups of closely related species, the identification of which was confounded by ambiguous or incorrect results (Bizzini and Greub 2010; Carbonnelle et al. 2011). One of these groups, the ACB complex, contains species significantly differing in their clinical impact. Recent studies (Alvarez-Buylla et al. 2012; Espinal et al. 2012; Sousa et al. 2014) pointed out difficulties in discrimination, especially between *A. baumannii* and *A. nosocomialis*. For that reason, we aimed at developing an alternative sample preparation procedure that would modulate the discriminatory power of MALDI-TOF MS profiling to improve differentiation between strains of the ACB complex (Šedo et al. 2013a).

Protocol

A set of 105 well-characterized strains of the ACB complex (32 *A. baumannii* strains, 29 *A. nosocomialis* strains, 22 *A. pittii* strains, and 22 *A. calcoaceticus* strains) was

selected for the purpose of this study. These strains were prepared according to two sample preparation protocols differing in the composition of the matrix solution (a saturated solution of CHCA in water:acetonitrile:trifluoroacetic acid, 47.5:50:2.5, v/v vs. 12.5 mg. ml^{-1} ferulic acid in water:acetonitrile:formic acid, 50:33:17, v/v). For both protocols, all other steps were carried out as recommended by Bruker Daltonics for the standard sample preparation including inactivation by vortexing in 70% ethanol for 30 s and acetonitrile/formic acid extraction (Freiwald and Sauer 2009). Each sample was deposited on three wells of the stainless steel MALDI target. Five independent MALDI-TOF mass spectra, comprising 1000 laser shots each, were acquired from each of the three wells (15,000 shots per sample in total). In the case of sample prepared with CHCA, the laser was directed automatically according to a predefined lattice raster, whereas the signals were acquired manually in the case of sample prepared with ferulic acid. The species identification was accomplished by using default settings of the Biotyper software. Cluster analysis was carried out within the Biotyper software by using the Pearson's product moment coefficient as a measure of similarity and the UPGMA as a grouping method.

Results

Limits in discrimination between *A. baumannii* and *A. nosocomialis* stem from detection of an insufficient number of species-specific peaks. While strains of *A. baumannii* yielded one such signal at 5748 Da, *A. nosocomialis* strains provided no species-specific signals that would allow their discrimination from *A. baumannii*. As a result, some of the strains belonging to these two species were not separated by cluster analysis correctly (Fig. 7.4). Similarly, when applying the Biotyper scoring algorithm, 31% of the tested *A. nosocomialis* strains were identified incorrectly as *A. baumannii*. The main criterion for selecting a suitable alternative sample preparation method was the detection of the highest number of signals that were different from those obtained using the standard sample preparation protocol. This was expected to increase the probability of detection of species-specific signals. As described in several papers, changes in mass spectral quality can be achieved by changing the matrix compound, matrix/sample solution, sample deposition method, or by treating the bacterial cells with various physical methods (Šedo et al. 2011) . As the most effective way, we adapted a protocol derived from the work of Madonna et al. (2000). The increased acidity of the solvent substantially changed the ionization properties of ferulic acid, yielding 10 peaks in consensus with the standard sample preparation protocol and providing 62 new signals of *A. baumanii* NIPH 501^T proteins, especially in the higher mass range. As shown in Fig. 7.5, mass spectra obtained by the alternative sample preparation protocol contained seven signals allowing discrimination between *A. baumannii* and *A. nosocomialis* strains. Improvement in the discriminatory power was reflected by cluster analysis, grouping all strains of the ACB complex to species-specific branches. The scoring algorithm also yielded correct identification results against a simple database of the ACB complex type strains analyzed using the alternative sample preparation protocol.

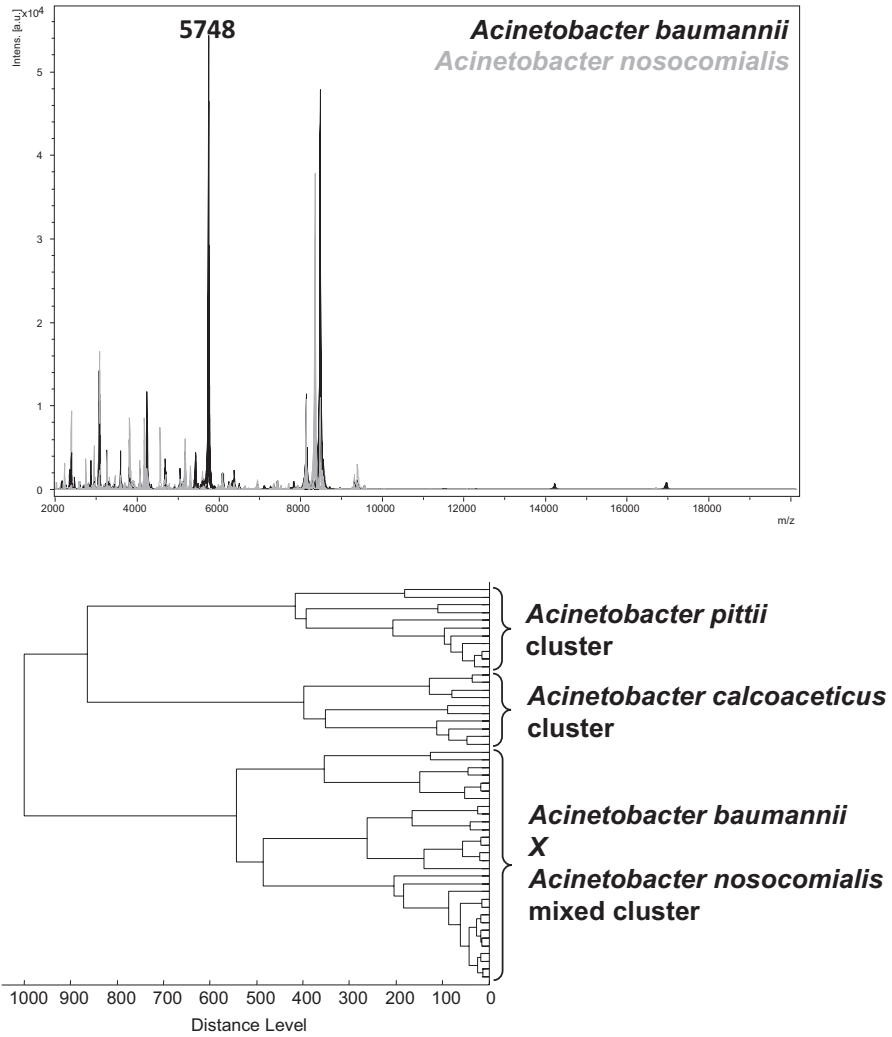


Fig. 7.4 Standard sample preparation protocol: MALDI-TOF mass spectra of *Acinetobacter baumannii* and *Acinetobacter nosocomialis* strains with one highlighted species-specific signal. Dendrograms constructed on the basis of the MALDI-TOF mass spectra of the strains of the ACB complex

Summary

The use of a highly acidified solution of ferulic acid as a MALDI matrix resulted in substantial modulation of the discriminatory power of MALDI-TOF MS profiling, which enabled successful discrimination of closely related species within the ACB complex. This alternative method can serve as a second method of choice for

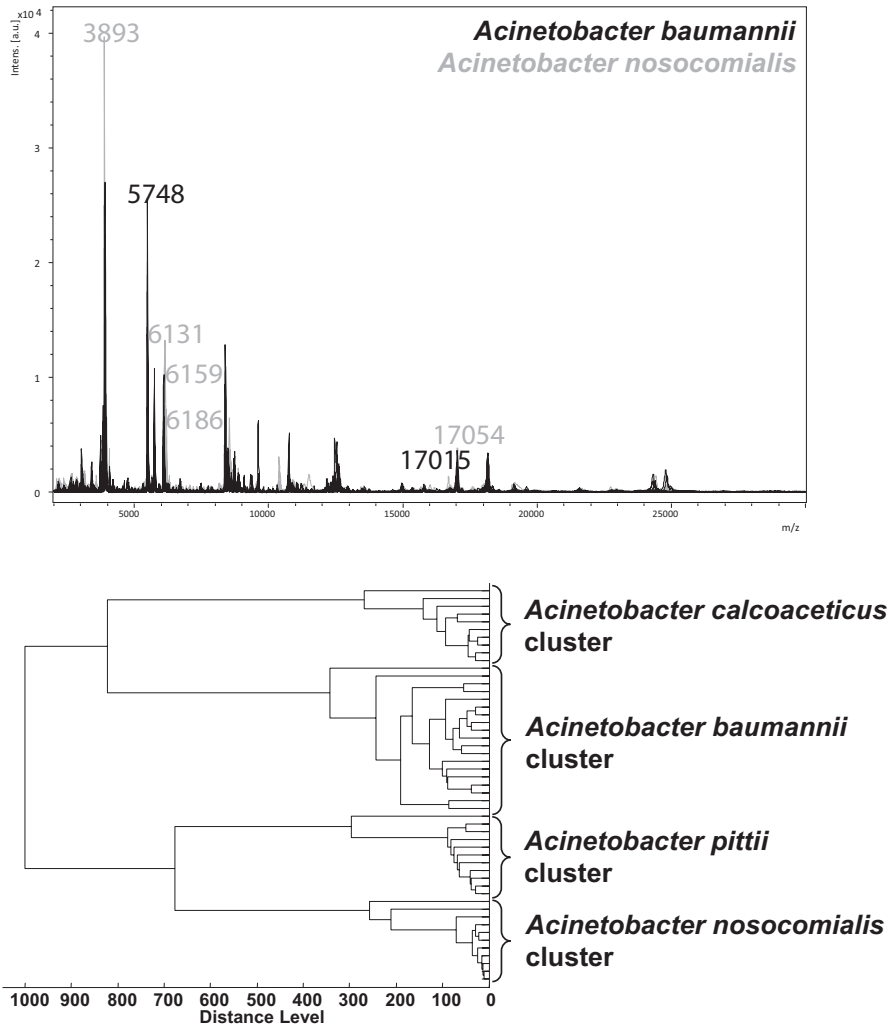


Fig. 7.5 Alternative sample preparation protocol: MALDI-TOF mass spectra of *Acinetobacter-baumannii* and *Acinetobacter nosocomialis* strains with highlighted species-specific signals. Dendrograms constructed on the basis of the MALDI-TOF mass spectra of the strains of the ACB complex

problematic cases where detection of an insufficient number of signals fails to distinguish between strains of interest. Limitations of the proposed alternative protocol include the failure to automatically acquire mass spectra. Due to non-homogeneous and irregular crystallization of ferulic acid, practically no valuable data were obtainable in the automatic acquisition mode. The need to construct an additional database of reference spectra is another item connected to the alternative protocol inevitably.

Modulation of the Discriminatory Power Using Microwave-Assisted Proteolysis

Background

Limitations in the discriminatory power of MALDI-TOF MS profiling in its standard arrangement demand the development of alternative approaches. Apart from modulation of the discriminatory power, new methodological variants should not change the routine workflow significantly in order to minimize efforts required for their introduction to laboratory practice. An interesting methodical variant was based on digestion of cellular material by trypsin, which would enable detection of different proteins from those obtained by the "whole-cell" approach (Warscheid and Fenselau 2003). However, the time required for proteolysis by trypsin significantly decreases analytical throughput. For this reason, we examined whether acceleration of this process by microwaves may be a simple and rapid methodological variant that may be useful in distinguishing between closely related bacterial strains (Balážová et al. 2014b).

Protocol

For an initial examination of the method, including repeated analyses at weekly intervals, and performing the reaction in three different commercially available microwave ovens, six *S. aureus* (SA) strains (CCM 885^T, CCM 1484, CCM 2022, CCM 2107, CCM 2351, and CCM 2352) were cultivated in triplicate. The method was finally tested to distinguish between 13 strains of *B. subtilis* ecotype PE10 and 15 strains of *B. subtilis* ecotype PE15. The bacterial cells were inactivated using 70% ethanol (1-day cultures were used to minimize the probability of sporulation of bacilli) and subsequently treated by microwave-assisted digestion with trypsin for 2 min. The MALDI-TOF MS analysis was conducted in the same way as in the standard sample preparation protocol (Freiwald and Sauer 2009) with the only change being an extension of the mass range of recorded spectra to lower *m/z* values. Each sample was deposited on the stainless steel MALDI target and three independent MALDI-TOF mass spectra, comprising 1000 laser shots each, were acquired from each of the sample. Within an individual well, the laser was directed automatically according to a predefined lattice raster. The cluster analysis was carried out within the Biotyper software by using the Pearson's product moment coefficient as a measure of similarity and the UPGMA as a grouping method.

Results

Although the MALDI-TOF MS profiles obtained using the standard method did not distinguish between three of the six *S. aureus* strains, microwave-assisted tryptic digestion resulted in clear separation of the strains, as determined by cluster analysis

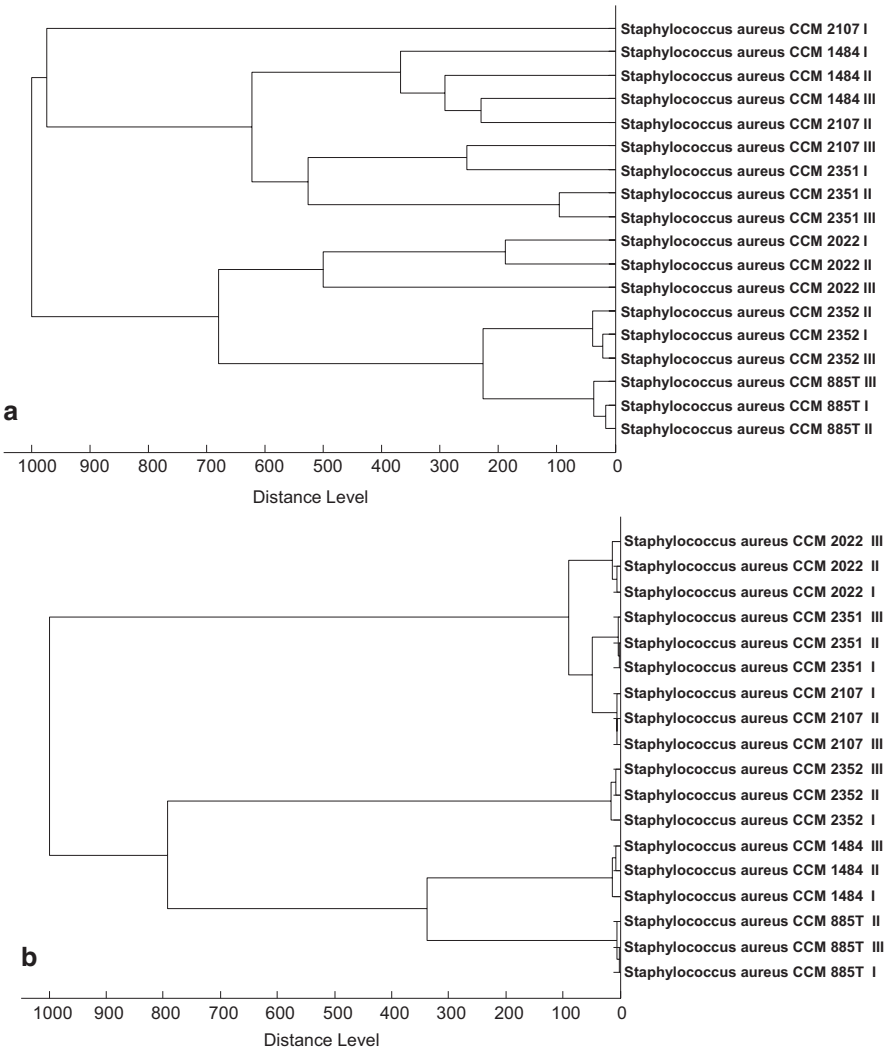


Fig. 7.6 Dendrograms calculated on the basis of cluster analysis of MALDI-TOF mass spectra of six *Staphylococcus aureus* strains analyzed in triplicate (marked as "I," "II," and "III") using **a** standard sample preparation protocol and **b** microwave-assisted digestion

(Fig. 7.6). The repeatability of the method was confirmed by two subsequent analyses conducted at weekly intervals and the robustness was verified by independent digestions in three different microwave ovens (data not shown). The discriminatory power, similarly to the previous example concerning acinetobacters, was strongly correlated with detection of specific signals. While the standard sample preparation method yielded only two strain-specific signals, seven strain-specific signals, accompanied by numerous repeatable differences in the relative intensity of other peaks, were obtained after digestion. Figure 7.7 compares discrimination between two *B. subtilis* ecotypes using the standard sample preparation protocol and the

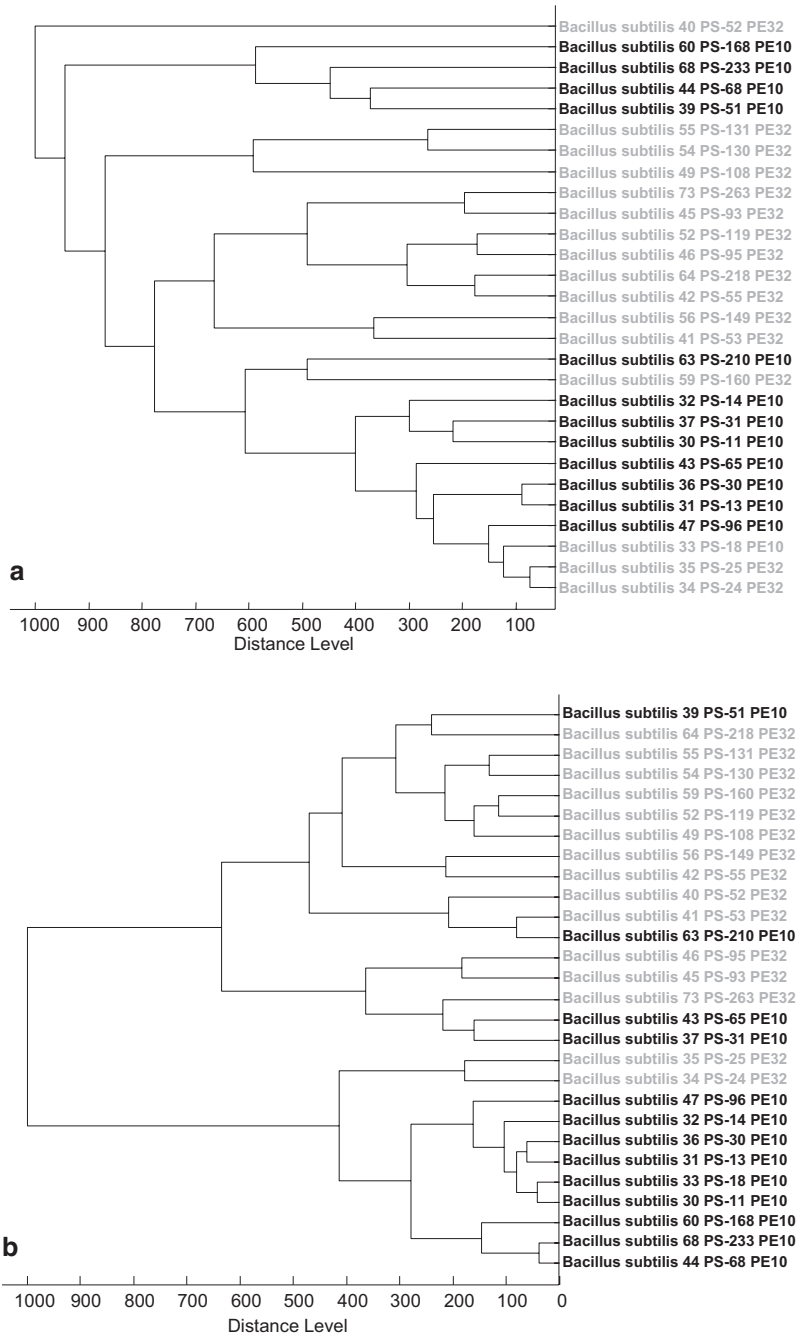


Fig. 7.7 Dendrograms calculated on the basis of cluster analysis of MALDI-TOF mass spectra of 28 *Bacillus subtilis* strains of two ecotypes (PE10 and PE32) using **a** standard sample preparation protocol and **b** microwave-assisted digestion

novel approach involving microwave-assisted trypsin digestion. While the standard protocol resulted in practically no clustering of the strains into any reasonable groups, clusters based on the MALDI-TOF mass spectra after digestion indicated the formation of two main clusters related to the ecotypes. Only six out of the 28 strains show incorrect positions in the dendrogram. The specificity of the method could even be demonstrated visually by detection of an ecotype-specific signal at 4087 Da, which was observed in 10 of ecotype PE10 strains and none of the ecotype PE32 strains. No ecotype-specific signals were found in the spectra of undigested cellular extracts obtained using the standard protocol.

Summary

Microwave-assisted tryptic digestion represents a simple methodological variant that is accessible for routine MALDI-TOF MS profiling users. As the analysis was carried out using CHCA as a matrix, automatic mass spectral acquisition, which was not applicable for the previous alternative protocol employing acidified ferulic acid, was possible in this case.

Conclusions

We have described approaches that modulate the discriminatory power of MALDI-TOF MS profiling in distinguishing between closely related bacterial strains. Cultivation conditions certainly play a prominent role in strain discrimination, and their proper selection and control may greatly contribute to the performance of the method. For cases concerning insufficient numbers of specific signals from strains of interest, two simple variants of the standard sample preparation protocol led to an improvement in the discriminatory power. None of the described methods could be considered as a universal tool for strain differentiation. However, these relatively simple methods could substantially improve the discriminatory power of MALDI MS profiling for particular sets of bacterial strains, enabling their use in a wide range of industrial and biotechnological applications where strain differentiation within defined sets is required.

Acknowledgments This work was supported by project CEITEC—Central European Institute of Technology (CZ.1.05/1.1.00/02.0068). All co-authors of our studies are gratefully acknowledged for fruitful cooperation.

References

Alvarez-Buylla A, Culebras E, Picazo JJ. Identification of Acinetobacter species: is Bruker biotyper MALDI-TOF mass spectrometry a good alternative to molecular techniques? Infect Genet Evol. 2012;12:345–9.

Anderson NW, Buchan BW, Riebe KM, Parsons LN, Gnacinski S, Ledeboer NA. Effects of solid-medium type on routine identification of bacterial isolates by use of matrix-assisted laser desorption ionization—time of flight mass spectrometry. J Clin Microbiol. 2012;50:1008–13.

Balážová T, Makovcová J, Šedo O, Slaný M, Faldyna M, Zdráhal Z. The influence of culture conditions on the identification of *Mycobacterium* species by MALDI-TOF MS profiling. FEMS Microbiol Lett. 2014a;353:77–84.

Balážová T, Šedo O, Štefanić P, Mandić-Mulec I, Vos M, Zdráhal Z. Improvement in *Staphylococcus* and *Bacillus* strain differentiation by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling by using microwave-assisted enzymatic digestion. Rapid Commun Mass Spectrom. 2014b;28:1855–61.

Bizzini A, Greub G. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. Clin Microbiol Infect. 2010;16:1614–9.

Carbonnelle E, Mesquita C, Bille E, Day N, Dauphin B, Beretti JL, Ferroni A, Gutmann L, Nassif X. MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. Clin Biochem. 2011;44:104–9.

Espinal P, Seifert H, Dijkshoorn L, Vila J, Roca I. Rapid and accurate identification of genomic species from the *Acinetobacter baumannii* (Ab) group by MALDI-TOF MS. Clin Microbiol Infect. 2012;18:1097–103.

Fedele L, Seraglia R, Battistotti B, Pinelli C, Traldi P. Qualitative characterization of bacterial strains employed in the production of yogurt by matrix-assisted laser desorption/ionization mass spectrometry. J Mass Spectrom. 1999;34:1385–8.

Freiwald A, Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry. Nat Protoc. 2009;4:732–42.

Grosse-Herrenthey A, Maier T, Gessler F, Schaumann R, Böhnel H, Kostrzewa M, Krüger M. Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization—time-of-flight mass spectrometry (MALDI–TOF MS). Anaerobe. 2008;14:242–9.

Heller KJ. Probiotic bacteria in fermented foods: product characteristics and starter organisms. Am J Clin Nutr. 2001;73:374S–9S.

Horneffer V, Haverkamp J, Janssen HG, Notz R. MALDI-TOF-MS analysis of bacterial spores: wet heat-treatment as a new releasing technique for biomarkers and the influence of different experimental parameters and microbiological handling. J Am Soc Mass Spectrom. 2004;15:1444–54.

Hussain MA, Knight MI, Britz ML. Proteomic analysis of lactose-starved *Lactobacillus casei* during stationary growth phase. J Appl Microbiol. 2009;106:764–73.

Ilina EN, Borovskaya AD, Serebryakova MV, Chelysheva VV, Momynaliev KT, Maier T, Kostrzewa M, Govorun NV. Application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for the study of *Helicobacter pylori*. Rapid Commun Mass Spectrom. 2010;24:328–34.

Krishnamurthy T, Ross PL, Rajamani U. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1996;10:883–8.

Lee H, Williams SKR, Wahl KL, Valentine NB. Analysis of whole bacterial cells by flow field-flow fractionation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Chem. 2003;75:2746–52.

Lotz A, Ferroni A, Beretti JL, Dauphin B, Carbonnelle E, Guet-Revillet H, Veziris N, Heym B, Jarlier V, Gaillard JL, Pierre-Audigier C, Frapy E, Berche P, Nassif X, Bille E. Rapid identification of mycobacterial whole cells in solid and liquid culture media by matrix-assisted laser desorption ionization—time of flight mass spectrometry. J Clin Microbiol. 2010;48:4481–6.

Lynn EC, Chung MC, Tsai WC, Han CC. Identification of Enterobacteriaceae bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. Rapid Commun Mass Spectrom. 1999;11:2022–7.

Madonna AJ, Basile F, Ferrer I, Meetani MA, Rees JC, Voorhes KJ. On-probe sample pretreatment for the detection of proteins above 15 kDa from whole cell bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2000;14:2220–9.

Maier T, Kostrzewa M. Fast and reliable MALDI-TOF MS-based microorganism identification. Chim Oggi-Chem Today. 2007;25:68–71.

Martiny A, Busson L, Wybo I, El Haj RA, Dediste A, Vandenberg O. Comparison of the microflex LT and vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol. 2012;50:1313–25.

Mazzeo MF, Sorrentino A, Gaita M, Cacace G, Di Stasio M, Facchiano A, Comi G, Malorni A, Siciliano RA. Matrix-assisted laser desorption ionization—time of flight mass spectrometry for the discrimination of food-borne microorganisms. Appl Environ Microbiol. 2005;72:1180–6.

Monteagudo-Mera A, Caro I, Rodríguez-Aparicio L, Rúa J, Ferrero MA, García-Armesto MR. Characterization of certain bacterial strains for potential use as starter/probiotic cultures in dairy products. J Food Prot. 2011;74:1379–86.

Moura H, Woolfitt AR, Carvalho MG, Pavlopoulos A, Teixeria LM, Satten GA, Barr JR. MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. FEMS Immunol Med Microbiol. 2008;53:333–42.

Parisi D, Magliulo M, Nanni P, Casale M, Forina M, Roda A. Analysis and classification of bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and a chemometric approach. Anal Bioanal Chem. 2008;391:2127–34.

Ryzhov V, Hathout Y, Fenselau C. Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl Environ Microbiol. 2000;66:3828–34.

Sauer S, Freiwald A, Maier T, Kube M, Reinhardt R, Kostrzewa M, Geider K. Classification and identification of bacteria by mass spectrometry and computational analysis. PLoS ONE. 2008;3:e2843.

Šedo O, Nemec A, Křížová L, Kačalová M, Zdráhal Z. Improvement of MALDI-TOF MS profiling for the differentiation of species within the *Acinetobacter calcoaceticus–Acinetobacter baumannii* complex. Syst Appl Microbiol. 2013a;36:572–8.

Šedo O, Sedláček I, Zdráhal Z. Sample preparation methods for MALDI-MS profiling of bacteria. Mass Spectrom Rev. 2011;30:417–34.

Šedo O, Vávrová A, Vaďurová M, Tvrzová L, Zdráhal Z. The influence of growth conditions on strain differentiation within the *Lactobacillus acidophilus* group using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling. Rapid Commun Mass Spectrom. 2013b;27:2729–36.

Sousa C, Botelho J, Silva L, Grosso F, Nemec A, Lopes J, Peixe L. MALDI-TOF MS and chemometric based identification of the *Acinetobacter calcoaceticus–Acinetobacter baumannii* complex species. Int J Med Microbiol. 2014;304:669–77.

Timke M, Kostrzewa M. "Inactivated mycobacteria bead preparation" method for identification of mycobacteria by MALDI-TOF MS. Int J Med Microbiol. 2011;301:1–1.

Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. Appl Environ Microbiol. 2005;71:58–64.

Warscheid B, Fenselau C. Characterization of *Bacillus* spore species and their mixtures using postsource decay with a curved-field reflectron. Anal Chem. 2003;75:5618–27.

Wunschel SC, Hill EA, McLean JS, Jarman K, Gorby YA, Valentine N, Wahl K. Effects of varied pH, growth rate and temperature using controlled fermentation and batch culture on matrix assisted laser desorption/ionization whole cell protein fingerprints. J Microbiol Met. 2005;16:456–62.

Chapter 8
Discriminatory Power of MALDI-TOF Mass Spectrometry for Phylogenetically Closely Related Microbial Strains

Peter Lasch, Daniela Jacob, Silke R. Klee and Guido Werner

Abbreviations

ACME	Arginine catabolic mobile element
ANN	Artificial neural network
BSL	Biosafety level
BURP	Based upon repeat pattern
CA	Community-Acquired
CC	Clonal complex
DHB	2,5-dihydroxybenzoic acid
EntA	Enterotoxin A
EntB	Enterotoxin B
HCCA	α-cyano-4-hydroxycinnamic acid
MALDI-TOF	Matrix-assisted laser desorption/ionization—time of flight
MLST	Multilocus sequence typing
MLVA	Multiple-locus variable number tandem repeat analysis
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSSA	Methicillin-sensitive *Staphylococcus aureus*
MS	Mass spectrometry
MSP	Main spectral projection
MW	Molecular weight

P. Lasch (✉)
Proteomics and Spectroscopy at the Centre for Biological Threats and Special Pathogens (ZBS6), Robert Koch-Institute, Seestraße 10, 13353 Berlin, Germany
e-mail: LaschP@rki.de

D. Jacob · S. R. Klee
ZBS 2, Highly Pathogenic Microorganisms, Robert Koch-Institute, Nordufer 20, 13353 Berlin, Germany

G. Werner
FG 13, Nosocomial Pathogens and Antibiotic Resistances, Robert Koch-Institute, Burgstraße 37, 38855 Wernigerode, Germany

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_8

PFGE	Pulsed-field gel electrophoresis
PSM	Phenol-soluble modulins
PVL	Panton-valentine leukocidin
SASP	Small, acid-soluble protein
SNR	Signal-to-noise ratio
TFA	Trifluoroacetic acid
TSST	Toxic shock syndrome toxin
VRE	Vancomycin-resistant *Enterococcus*
VREfm	Vancomycin-resistant Enterococcus *faecium*
VSE	Vancomycin-sensitive *Enterococcus*
VSEfm	Vancomycin-sensitive *Enterococcus faecium*
UHCA	Unsupervised hierarchical cluster analysis

Introduction

The past two decades have witnessed dramatic progress in the way microorganisms are identified. New spectrometry- and spectroscopy-based methods have been developed that allowed for rapid, reliable, and cost-effective differentiation, identification, and classification of microorganisms. For instance, techniques such as Raman and infrared (IR) spectroscopy (Helm et al. 1991; Schmitt et al. 1998; Goodacre et al. 1998; Maquelin et al. 2000; Rösch et al. 2005), pyrolysis MS (John and Catherine 1975; Goodacre 1994), and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Fenselau 1994; Claydon et al. 1996; Holland et al. 1996; Krishnamurthy et al. 1996; Cain et al. 1994) have considerably enhanced the reliability and accelerated the speed of microbial diagnostics. In particular, MALDI-TOF MS has been increasingly used to support microbial identification in clinical routine diagnostics (Maier et al. 2006; Dieckmann et al. 2008; Seng et al. 2009), in biodefense (Lasch et al. 2009, 2010; Elhanany et al. 2001), or in food safety applications (Wenning et al. 2014; Quintela-Baluja et al. 2013; Bohme et al. 2013). The MALDI-TOF MS methodology is based on the reproducible detection of protein mass patterns obtained from whole cells, cell lysates, or crude bacterial extracts (Lay 2001). Microbial MALDI-TOF mass spectra can be regarded as snapshots of the protein composition of the strains under study. Many of the mass spectral signals have been assigned as high-abundance proteins with housekeeping functions, such as basic ribosomal proteins or nucleic acid-binding proteins (Ryzhov and Fenselau 2001; Pineda et al. 2003; Dieckmann et al. 2008). These proteins are highly conserved and consistently expressed under nearly all growth conditions; they can be thus regarded as robust biomarker candidates of the organisms studied.

For microbial identification, the mass spectra can be analyzed in two principally different ways: database- or library-based methods compare mass peak tables from unknown bacterial strains by matching them against libraries with validated microbial reference spectra (Holland et al. 1996; Arnold and Reilly 1998). Today,

library-based methods are more frequently utilized; such methods do not require the precise knowledge of biomarker identities (Sandrin et al. 2013). On the contrary, bioinformatics-enabled methods which often rely on the information revealed by peak/protein identification and sequence database searches (Demirev et al. 2005), or alternatively on the comparison of the experimental mass peak patterns with protein masses predicted from microbial genomes (Demirev et al. 1999; Pineda et al. 2000), are more rarely used. It is however widely accepted that both approaches allow accurate typing of microorganisms at the genus and also at the species level. In addition, there is a growing body of evidence which indicates the potential of MALDI-TOF MS for microbial characterization below the species level. It is, however, not the goal of this mini-review to summarize, or to evaluate, the current state of knowledge in this field of research; for this purpose the reader is referred to the excellent review of Sandrin et al. (2013) and to other chapters of this textbook. Within the framework of this mini-review we want to outline especially the experiences our groups have gained when applying MALDI-TOF MS to characterize microorganisms below the species level. The strategy in our studies on subspecies differentiation consisted primarily in not to compromise—whenever possible—the simplicity and the speed advantage of the MALDI-TOF MS technique. Therefore, our activities did not involve complex technical changes or major modifications of the sample preparation methods that are known to increase the data richness of the mass spectral profiles. On the basis of existing and widely used standard protocols, it should rather be systematically investigated whether further improvements of the taxonomic resolution are achievable by (1) extension of the spectral database through a better, that is, more comprehensive, and more representative strain selection; (2) cautious optimization of existing sample preparation and data acquisition protocols; (3) rigorous standardization; and (iv) utilization of an optimized data analysis pipeline that involves advanced methods of spectral preprocessing, feature selection, and supervised methods of multivariate classification analysis. Based on the examples of three important pathogens, *Staphylococcus aureus*, *Enterococcus faecium*, and *Bacillus cereus* sensu lato (s.l.), we will illustrate our point of view regarding the perspectives and the limitations of the MALDI-TOF MS-based method for microbial typing.

Materials and Methods

Sample Preparation

It is generally accepted that cultivation conditions such as growth time, growth temperature, the composition of the culture medium, or the presence/absence of oxygen, or CO_2 may exert a noticeable influence on the protein expression of the microorganisms under study (Anderson et al. 2012; Balazova et al. 2014; Goldstein et al. 2013; Keys et al. 2004; Valentine et al. 2005; Sedo et al. 2013; Sandrin et al. 2013). Moreover, sample preparation and instrument-specific factors may also af-

fect the mass spectral profiles. Interestingly, however, variation of cultivation conditions and data acquisition parameters or the lack of standardized sample preparation protocols lead to changes of only a part of the mass signals: a large fraction of the signals present in microbial mass spectra remains more or less stable when cultivation conditions are changed and sample preparation or measurement protocols are varied (Valentine et al. 2005; Brandt 2006; Wieme et al. 2013). The molecular identities of these invariant mass peaks have been assigned by systematic investigations. It could be established that many of these consistently observed signals arise from a family of small basic proteins of the ribosomal subunits. Ribosomal proteins can be consistently observed in all bacterial life-forms which is key to the stability and robustness of the MALDI-TOF MS profiling technique. On the contrary, it has been often stated, that the variable part of the mass signals is not critical for microbial identification at the genus or the species level. Therefore, the methodology for identifying microbial species can be generally regarded as a relatively robust technology. Below the species level, however, this statement is no longer accurate. The accuracy of the MALDI-TOF MS methodology for identification and differentiation at the subspecies level is significantly affected if cultivation conditions, sample preparation, and measurement parameters are not rigorously standardized (Sauer et al. 2008). As this study was not primarily concerned with species identification, but with differentiation and identification at the subspecies level, it was thus vital to pay special attention on standardization and cautious optimization of the whole MALDI-TOF MS workflow.

Bacterial strains of *S. aureus* and *E. faecium* were prepared by growing each strain by two passages under aerobic conditions. Cells were cultured on Caso agar (Merck) for 24 h at 37°C. For microbial sample preparation, a modification of the trifluoroacetic acid (TFA) (Lasch et al. 2008) inactivation protocol for highly pathogenic microorganisms was used (for details, see Lasch et al. 2014) Microbial sample solutions were mixed 1:1 (v/v) with a concentrated HCCA solution (12 mg/mL) which was prepared by dissolving HCCA in TA2, a 2:1 (v/v) mixture of 100% acetonitrile and 0.3% TFA. From these mixtures a volume of 2 µl was spotted on steel sample targets before the measurements. For each microbial strain, biological replicate spectra were collected from ideally three individual microbial cultures.

Bacillus strains were grown under aerobic conditions on in-house prepared LB agar plates by two passages for 24 h at 37°C. Cells were then harvested and an equivalent of three full blue plastic loops (approx. 4 mg) was added to sterile water (see Lasch et al. 2009 for details). Further steps of microbial sample preparation were the same as those used for *S. aureus* and *E. faecium*.

MALDI-TOF Mass Spectrometry

Microbial mass spectra were acquired from samples prepared on stainless steel targets using an *Autoflex I* mass spectrometer from Bruker Daltonics. The mass spectrometer was equipped with a slightly defocused nitrogen laser ($\lambda = 337$ nm) and controlled by the FlexControl 3.0 data acquisition software from Bruker. Spec-

tra were acquired in the manual mode of operation. Pulse ion extraction time was varied between 200 and 300 ns while the sampling rate was set to 0.5 GHz. Mass spectral measurements were carried out in the linear measurement mode using an acceleration voltage of 20.00 (ion source 1) and 18.25 (ion source 2) kV. Lens voltage was 6.50 kV. Spectra were stored in the range between m/z 2000 and 20,000 whereas the majority of the mass signals was observed in the m/z mass range between 2000 and 12,000. For calibration we used Bruker's protein calibration standard I (measurements before 2011), or the *Escherichia coli* DSM 3871 reference strain. For calibration with *E. coli* the doubly charged 50S ribosomal protein L29 (RL29; m/z 3637.8; $[M+2H]^{2+}$), RL36 (m/z 4365.3, $[M+H]^{+}$), RS22 (m/z 5096.8, $[M+H]^{+}$), RL34 (m/z 5381.4, $[M+H]^{+}$), RL33meth (m/z 6255.6, $[M+H]^{+}$), RL32 (m/z 6316.2, $[M+H]^{+}$), RL30 (m/z 6411.6, $M+H]^{+}$), RL35 (m/z 7158.8, $[M+H]^{+}$), RL29 (m/z 7274.5, $[M+H]^{+}$), RL31 (m/z 7872.1, $[M+H]^{+}$), RS21 (m/z 8369.8, $[M+H]^{+}$), DNA-binding protein HU-beta (m/z 9226.6, $[M+H]^{+}$), and RS20 (m/z 9536.3, $[M+H]^{+}$) were used as external calibrants enabling a mass accuracy of approximately 300 ppm. For each MALDI-TOF mass spectrum at least 600 individual laser shots were co-added in order to achieve a high spectral signal-to-noise ratio (SNR).

Data Analysis

Microbial MALDI-TOF mass spectra are complex signals which can carry an enormous amount of information. Mass spectra of microorganisms may contain several dozens of mass peaks; a comprehensive analysis involves several distinct analysis steps: spectral preprocessing and peak detection followed by the classification/identification analysis procedure in the strict sense.

Preprocessing The main goals of spectral preprocessing can be summarized as follows: (1) improvement of the robustness and accuracy of subsequent classification analysis, (2) improved interpretability, (3) detection and removal of outliers and trends, and (4) reduction of the dimensionality of subsequent data-mining tasks. This step often involves the removal of irrelevant and/or redundant information by feature selection (Lasch 2012).

Within the context of subspecies identification, we found that a specifically designed preprocessing workflow with routine tests for spectral quality, smoothing, baseline correction, and intensity normalization are key to all subsequent analysis steps.

The quality of microbial mass spectra was assessed visually immediately after data acquisition with regard to the following criteria: first and foremost, the SNR and the presence of a sufficient number of mass peaks were evaluated. Further quality criteria were a flat spectral baseline and the absence of interfering, or confounding, mass peaks from plasticizers and other synthetic polymer additives. Outliers, that is, spectra failing to meet one or more of the quality requirements were not accepted for subsequent analyses. The remaining spectra were de-noised in the following by

applying a Savitzky–Golay smoothing filter with 17–25 smoothing points (Savitzky and Golay 1964). The next preprocessing step, baseline correction, is aimed at flattening the spectral baseline. Most baseline correction procedures have in common that an estimate of a background is subtracted from the unprocessed spectrum leading to a better interpretable signal (Shin et al. 2010; Williams et al. 2005; Lasch and Naumann 2011). Normalization as the final step of spectral preprocessing enables comparison of MS intensity values, which is considered essential for producing *so-called* pseudo-gel views (see below). When analyzing microbial MALDI-TOF mass spectra we found vector normalization (also called 2-norm) particularly well suited. The result of vector normalization is a spectrum in which the sum of the squared intensities over all m/z values is a constant.

Peak Detection Peak detection can be regarded as an integral part of the spectral preprocessing pipeline. Any peak detection procedure is a specific form of feature selection, during which mass spectra are reduced to peak tables. In this way a significant data reduction, by a factor of 100 or even more, is achieved, which considerably facilitates further analysis of the microbial mass spectra. The scientific literature contains a number of references to studies in which peak detection methods are suggested (Yang et al. 2009; Yu et al. 2006; Kempka et al. 2004; Renard et al. 2008). However, due to practical aspects, such as the availability of software implementations, these rather complex peak detection algorithms are utilized only rarely. In most of the studies dealing with microbial identification, peak detection is carried out using standard procedures implemented in commercial software products (e.g., Flex Analysis from Bruker Daltonics). In general we found that peak detection methods such as centroid, snap, or sum (MALDI BioTyper 3.0 User Manual 2012) usually performed reasonably well.

In addition to commercial software products we also took advantage of custom designed software (MicrobeMS; Lasch 2015) for the evaluation of our microbial mass spectra. MicrobeMS is Matlab-based and involves a specifically optimized peak detection routine. One of the key features of peak detection in MicrobeMS is a sigmoid intensity threshold function which was introduced to model the m/z dependence of the analytical sensitivity of MALDI-TOF MS. This threshold function defines intensity thresholds at each m/z value. In the MicrobeMS implementation, an intensity threshold at low m/z values is larger than at high m/z values. Another feature of the MicrobeMS peak detection routine allows to precisely define the number of resulting peaks per spectrum. This particular feature makes peak detection partially independent from the SNR which turned out to be extremely useful for subsequent classification analysis.

Data Visualization When dealing with large spectral databases which may sometime comprise hundreds of individual mass spectra, data inspection and data assessment cannot be done anymore by manual screening of all individual mass spectra. In such instances simulated pseudo-gel views are useful means for visual analyses. Pseudo-gel views allow comparing multiple groups of spectra and are obtained from preprocessed mass spectra by converting spectral intensities into color or gray scales. In a gel view representation color or gray intensities are plotted as a function of the m/z values and the spectra index.

Identification of Taxon-Specific Biomarkers Experiences from our work and those of others strongly suggest that efficient strategies for identification of taxon-specific biomarkers should ideally involve measurements of a representative number of strains per taxonomic unit (Lasch and Naumann 2011). This is particularly important for accurate identification at and below the species level. To differentiate microorganisms below the species level, MALDI-TOF MS characterization should be ideally carried out by repetitive measurements from an adequate number of independent microbial cultures per strain. Therefore, the current experimental workflow at the Robert Koch-Institute (RKI) encompasses measurements of four technical replicate spectra per each individual culture, that is, per biological replicate. Since we normally generate three independent cultures per strain subsequent statistical analyses can be carried out on the basis of at least 12 spectra per strain. In this way it is assured that the mass spectral database contains sufficient information with respect to repeatability and reproducibility. The expanded number of spectra serves as an improved basis for systematic statistical analyses to identify taxon-specific microbial biomarkers.

Systematic searches for taxon-specific biomarkers involved a series of independent (unpaired) *t*-tests. In these *t*-test series, null hypotheses H0 are verified which assume that the means of two normally distributed populations are equal. For testing these hypotheses we usually subdivided a given microbial database into two distinct categories, for example into “strain A” and “strain B.”. Then, the discriminative power of each spectral feature is assessed by an independent (unpaired) *t*-test. In this way, the number of t-tests to be performed equals the number of features in the data set after preprocessing and peak detection. Each *t*-test provided a *p* value which can be considered a statistical measure of the distinctness of the features found in categories A and B at the given m/z values. A *p* value of one confirms the null hypothesis (equal class means) while small *p* values cast doubt on the H0 hypothesis. The *p* values can be plotted in an inverted logarithmic scale as a function of the m/z values. These plots are easy to interpret: the higher the –log(*p*) value, the higher the discriminative potential of the spectral feature under investigation (Lasch and Naumann 2011). In the original scientific publications and within the context of the present mini-review, *p* value plots were extensively used to statistically assess the discriminative potential of mass spectral biomarker candidates.

Classification Analysis Identification and differentiation of microbial strains, species, and genera is commonly performed by pattern matching (Sauer and Kliem 2010). In MS-based microbial identification, pattern matching usually involves application of tailored methods that compare peak tables extracted from experimental mass spectra with reference peak tables of a commercial or dedicated database. A reference peak list, also called main spectral projection (MSP, BioTyper Bruker Daltonics), or SuperSpectrum™ (VITEK-MS, bioMérieux), is obtained from a representative number of spectra per taxonomic unit and contains not only intensity values and the m/z positions of relevant signals, but also the relative frequency of these mass peaks. As it is beyond the scope of this mini-review to cover all aspects of peak-matching algorithms implemented in the commercial software products we refer the readers to the respective software manuals (MALDI BioTyper 3.0 User

Manual 2012; VITEK® MS User Manual 2011). In the context of the present mini-review, we wish to focus, however, on the description of multilayer perceptron artificial neural networks (MLP-ANN). These types of networks are powerful tools to solve complex classification problems when "strong," that is, highly sensitive/specific biomarkers are absent and were found particularly useful for rapid, efficient, and reliable MS-based differentiation, identification, and classification of microorganisms (Lasch et al. 2009; 2010).

MLP-ANNs have been known since a long time as computational models to model complex relationships between inputs and outputs and to find patterns (Marques de Sa 2001). The MLP class of neural networks requires a set of training samples which are ideally representative for the given categories. To avoid overfitting and for assessing the robustness of ANN class assignment, the ANN analysis workflow includes procedures of training, internal validation, and (external) testing. Training and internal validation requires sets of spectra for which the class membership is known. During network training, the model performance is determined on the basis of errors between the obtained outputs and desired target values of both the training and the internal validation data sets. At the training stage ANN performance can be optimized by modifying the way of preprocessing, adding or eliminating spectral features, or changing the network's architecture. When training is completed the performance of the ANN classifier can be tested using an external validation data subset.

One of the constraints which limited broad application of neural networks in microbial MALDI-TOF MS is the usually large number of categories, that is, of taxonomic units (species, strains) to be identified. Furthermore, as training and validation of ANNs necessitates relatively large sample numbers, ANN model development for clinical microbiology often demands a lengthy and tedious training process. A popular approach to reduce time and efforts when designing neural network models for multiclass classification problems is the use of modular (hierarchical) ANN classification schemes (Auda and Kamel 1998; Udelhoven et al. 2000). Application of modular ANN classifiers involves compilation and linkage of small and flexible network modules specifically designed for specialized classification tasks. These individual ANNs can be trained and validated independently. The hierarchy of classifiers can be changed to perform additional or to adapt existing classification tasks. The primary advantage of modular ANN models lies in the fact that individual ANNs can be specifically optimized to accurately identify members of only a few—sometimes only two—classes (Udelhoven et al. 2000).

Staphylococcus Aureus

Population Structure and Typing of S. aureus (Clonal Complexes)

Staphylococcus aureus is a common colonizer of the interior nares of humans and animals and thus primarily a commensal bacterium. Due to their wide range of putative virulence factors, isolates of *S. aureus* can, under certain preconditions,

cause a variety of infections, from mild skin and soft tissue infections to severe and life-threatening invasive infections. A number of clinical progressions are directly linked to the expression of certain factors and/or toxins during different phases of an infection. The diagnostic recognition of these factors is an important measure in proposing infection progression and evaluating treatment options. Classically, these factors are assessed in specialized diagnostic laboratories by polymerase chain reaction (PCR) from microbial DNA or by specific enzyme-linked immunosorbent assay (ELISA)-based tests (for toxic shock syndrome toxin TSST, exfoliative toxins EntA/B, etc.). It is known that a number of these factors are accumulated in certain clonal lineages of *S. aureus* allowing a postulation of a distinct strain type from specific virulence gene patterns and vice versa (Robert et al. 2011).

Isolates of *S. aureus* have gained major medical importance as multidrug-/methicillin-resistant variants called MRSA. Methicillin-resistant *Staphylococcus aureus* (MRSA) derive from susceptible progenitors by acquisition of a mobile resistance gene cassette called SCCmec harboring a *mecA* gene complex (Hiramatsu et al. 2013). The *mecA* gene encodes an alternative penicillin-binding protein PBP2A which, when expressed, confers resistance to all beta-lactam antibiotics (except the new group V cephalosporins, ceftaroline and ceftobiprole (File et al. 2012)). Molecular typing methods such as multilocus sequence typing (MLST) and subsequent bioinformatics data analysis with software programs such as eBURST allow differentiation of MRSA/*S. aureus* isolates into clonal complexes (CC; Harmsen et al. 2003; Strommenger et al. 2008). Such analyses revealed that the population structure of MRSA is dominated by a small number of successful MRSA clones/CC which were identified all over the world (CC/ST22, CC/ST239, CC/ST8) (Grundmann et al. 2010). Molecular typing of MRSA/*S. aureus* by variability of a DNA sequence motif in a protein A gene sequence (*spa*) has turned into a frontline tool for typing and has substituted previously prominent techniques such as macrorestriction analysis in pulsed-field gel electrophoresis (PFGE) (Strommenger et al. 2008). The repeat region of the *spa* gene used for typing varies in the composition and the number of repeats. An assessment of these differences is used by a typing software called StaphType (Ridom, Münster, Germany) introducing a common nomenclature and allowing to deduce strain relatedness by a comparison of the DNA sequence and the repeat numbers (BURP—based on repeat pattern; Mellmann et al. 2007). Results of *spa* typing and MLST were mainly congruent allowing to deduce the MLST type or at least the CC from the corresponding *spa* type (Strommenger et al. 2008; Harmsen et al. 2003).

Over the past years MALDI-TOF MS has largely changed the process and workflow of bacterial diagnostics. This has raised the question to what extent the quality of routinely determined MALDI-TOF MS spectra allows a differentiation below the *S. aureus* species level and extraction of additional clinically relevant information on pathogenicity and resistance. To determine the species and assess the previously mentioned markers for pathogenicity, antibiotic resistance, and/or clonal relatedness by a single common approach such as MALDI-TOF MS, seemed a reasonable goal attainable by this technique.

MS Biomarkers for Strain or CC Allocation of S. aureus/MRSA

It is beyond doubt that isolates of the species *S. aureus* could be identified and differentiated from other, coagulase-negative staphylococci on the basis of their MALDI-TOF MS patterns (Moon et al. 2013; Schulthess et al. 2013) (cf. also Fig. 8.1). A few publications and presentations at conferences postulated to differentiate MRSA or *S. aureus* below the species level into clonal types and complexes (types and CC mainly based on MLST or *spa*) (Josten et al. 2013; Wolters et al. 2011). *S. aureus* isolates from different types or groups could be differentiated by single discriminating peaks (i.e., putative biomarkers), or by analyzing/comparing complete MALDI-TOF MS spectra.

In a previous study, our group investigated 59 diverse *S. aureus* strains, mainly MRSA, representing all major European hospital-associated strain types (clonal complexes: CC5, CC8, CC22, CC30, CC45, and CC398; Lasch et al. 2014). Sample preparation was carried out by the TFA sample preparation/inactivation protocol (Lasch et al. 2008). MALDI-TOF MS spectra were acquired from three independent replicates. MS data evaluation included manual peak inspection as well as unsupervised hierarchical cluster analysis (UHCA) and supervised ANN analysis (see Table 8.1 for the results of ANN classification). Furthermore, using statistical methods for MS biomarker identification we confirmed a number of m/z peaks described in previous publications but failed to identify biomarkers allowing a 100 % reliable differentiation of distinct lineages, CC, or *spa* types (Lasch et al. 2014). In one of the previous studies, Josten and coworkers described specifically identifying mass signals in spectra of distinct clonal lineages of *S. aureus* (Josten et al. 2013). Experimental data were convincing since the authors performed additional experiments to confirm the validity of the biomarkers such as by (1) analyzing mutants lacking the corresponding markers, (2) expression of antisense RNA that resulted in a knockdown of the gene of interest, and (3) tandem MALDI-TOF/TOF MS analyses. The sensitivity and specificity of the biomarkers proposed by Josten et al. was tested for certain clonal lineages of *S. aureus* on our strain collection (Lasch et al. 2014). It turned out that none of the suggested biomarkers allowed a 100 % accurate identification/discrimination of distinct clonal groups—a fact which was already admitted for most of the biomarkers by Josten et al. (2013). However, a few marker peaks revealed a comparably good sensitivity and/or specificity; for instance, the protein SA2420.1 at m/z 3891 turned out to be adequately sensitive and specific for clonal complex CC5 isolates under the conditions and the strain collections tested (see Fig. 8.2). Also for CC8 isolates, we reached an acceptable level of discrimination by the neural network approach (cf. Table 8.1 and Lasch et al. 2014). Boggs and coworkers also investigated CC8 isolates and postulated to identify isolates of the community-acquired (CA)-MRSA "USA300" subclone by a genetic algorithm model that was based on peak intensity values using the ClinProTools software provided by Bruker Daltonik (Bremen/Germany) (Boggs et al. 2012). USA300 is a prominent CA-MRSA clone appearing in the USA, Asia, and Europe which is—according to recent studies—the most prevalent ambulatory-

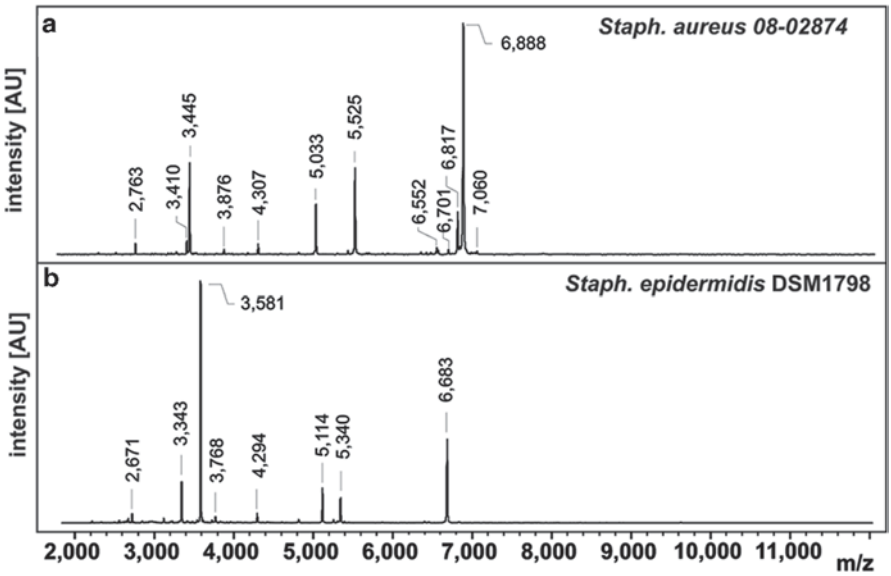


Fig. 8.1 **a** MALDI-TOF mass spectra of *S. aureus* 08–02874 (CC45, t004, hospital-acquired MRSA) and **b** the strain *S. epidermidis* DSM 1798 in the mass range of m/z 2000–12,000. Reproduced from Lasch et al. (2014) with permission

acquired pathogen in invasive skin and soft tissue infection in certain regions in the USA (Blanco et al. 2011; O'Hara et al. 2012; Nimmo 2012). USA300 contains, in addition to its methicillin resistance, some virulence properties, such as the ability to produce a Panton-Valentine Leukocidin (PVL) and enhanced levels of phenol-soluble modulins (PSM). This and the so-called arginine catabolic mobile element (ACME) which is involved in detoxifying harmful host-derived polyamines have considerably increased the CA-MRSA's pathogenicity and the strength of corre-

Table 8.1 Typing of *S. aureus* using MALDI-TOF mass spectra and a hierarchical artificial neural network (ANN): overview of the ANN classification results of the external test data set with spectra from *S. aureus*. The MALDI-TOF-MS-based classification was established by teaching a hierarchical ANN model with mass spectra of *S. aureus* from six clonal complexes. With the exception of CC8, the data indicated no reliable differentiation between spectra from the individual clonal complexes of *S. aureus*. Reproduced from Lasch et al. (2014) with permission

Clonal complex	Number of spectra	Number of correctly classified spectra	Number of wrongly or not classified spectra	True positive rate (%)
CC5	20	16	4	80
CC8	22	20	2	91
CC22	14	6	8	43
CC30	9	7	2	78
CC45	10	4	6	40
CC398	13	9	4	69

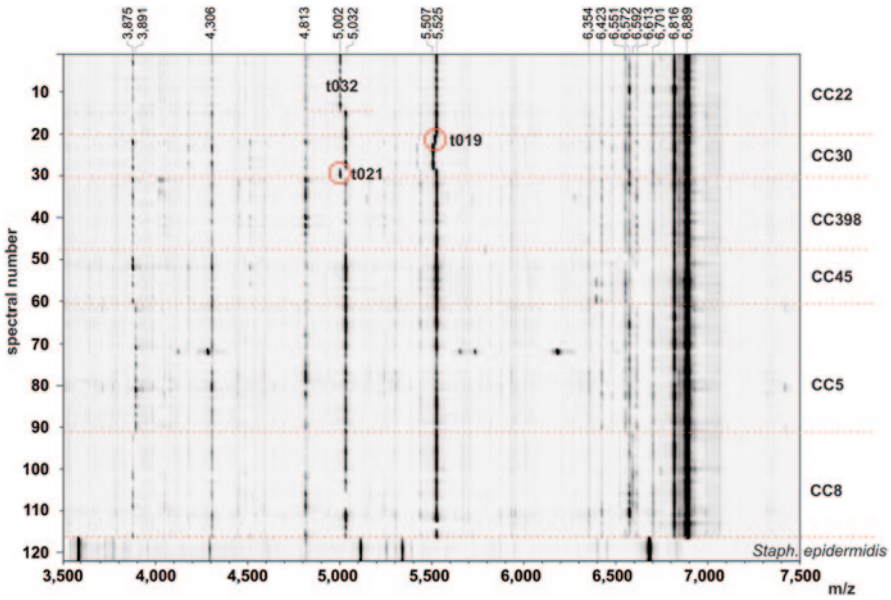


Fig. 8.2 Pseudo-gel view produced from mass spectra of 59 *S. aureus*, mainly MRSA (lines 1–116) and of *S. epidermidis* (lines 117–123) in the mass range of m/z 2000–12,000. Non-copy isolates from different clonal complexes CC22, CC30, CC398, CC45, CC5, and CC8 were included. MLST types were inferred from *spa* types; t0332, t019, and t021 depict corresponding spa types (see Lasch et al. 2014 for details). Reproduced from Lasch et al. (2014) with permission

sponding infections in otherwise immunocompetent patients (Otto 2013). Boggs and coworkers suggested in their article three mass peaks (m/z 5932, 6423, 6592) to be lineage-specific for USA300 and classified 197 of 224 tested *S. aureus* strains correctly (USA300 vs. non-USA300). Specificity was 89%, and sensitivity was 87%. However, the diversity of the non-USA300 group was not described in details; isolates were mainly typed by PFGE and grouped accordingly. Nevertheless, the m/z peak at 6592 was especially linked to our CC8 isolates and mainly absent in non-CC8 isolates (Lasch et al. 2014) (see also Fig. 8.2). Our own strain collection contained only a few CC8 strains and even less USA300 isolates to confirm or disprove this suggested lineage-specific peak also for the European USA300 variants.

MS Biomarkers for Predicting Virulence Genes and Methicillin Resistance of S. aureus/MRSA

Several studies hypothesized the prediction of distinct virulence markers from MALDI-TOF MS spectra such as TSST or PVL (Gagnaire et al. 2012; Bittar et al. 2009). Later studies challenged earlier attempts which aimed at identifying biomarker peaks for these distinct toxins and virulence markers. The suggested marker

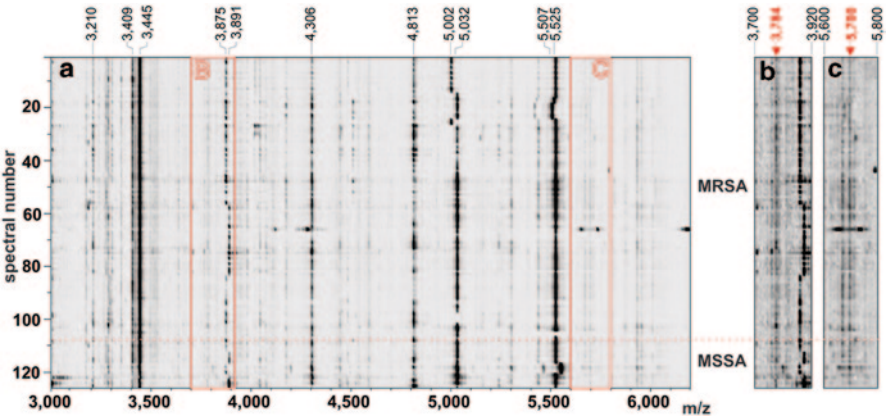


Fig. 8.3 **a** Pseudo-gel view produced from spectra of the RKI's *S. aureus* strain collection with 108 spectra of MRSA (lines 1–108) and 18 spectra of MSSA isolates (lines 109–126). **b** Detail of panel (**a**) in the m/z region 3700–3920 at an increased contrast. **c** Respective detail of (**a**) in the m/z region between 5600 and 5800. Low-intensity peaks at m/z 3784 and 5700 do not allow discrimination between MRSA and MSSA strains (see *arrows* in panels (**b**) and (**c**), see also text and Wang et al. (2013))

peaks for instance for TSST and PVL were not exclusively prevalent in all positive strains (tested by PCR and/or by specific antibodies) and absent in negative ones when tested on a larger, more comprehensive strain collection (Josten et al. 2013; Szabados et al. 2011; Dauwalder et al. 2010).

Data and studies describing the capability of MALDI-TOF MS to predict methicillin resistance in *S. aureus* are contradictory. A few publications postulated discriminating MRSA from MSSA isolates on the basis of supposed *mecA*/PBP2a peaks in MALDI-TOF MS spectra (Du et al. 2002; Wang et al. 2013). In our previous study, we also examined the question of whether methicillin resistance is associated with specific marker peaks absent in spectra of methicillin-susceptible strains (Lasch et al. 2014). We could not identify differences between our MRSA ($n=55$) and MSSA strains ($n=6$, low p values of the t-tests which did not support groupings, data not shown) and could not confirm marker peaks at m/z 3784 and 5700 postulated by other authors to be specific for MRSA (Wang et al. 2013)(see also Fig. 8.3). A recent analysis of an isogenic strain pair differing only in *mecA* but showing identical MALDI-TOF MS spectra also disproved the hypothesis that methicillin resistance in MRSA is identifiable by MALDI-TOF MS (Szabados et al. 2012). In recognizing these difficulties MALDI-TOF MS experts have already started using alternative approaches in measuring antibiotic resistance, for instance, using stable isotopes (Jung et al. 2014; Kostrzewa et al. 2013).

In general, size and composition of many strain collections varied widely in the studies investigating performance of MALDI-TOF MS for typing and strain characterization. In some of these studies, alternative or standard typing methods were lacking making it impossible to compare groups determined by MALDI-TOF MS spectra with groups based on MLST, *spa*, or PFGE typing. Despite major efforts

in standardizing cultivation, sample preparation, and analyses, the results of all the studies available so far suggest that sufficiently sensitive and specific biomarkers for *S. aureus*, clonal lineages, or types are an exception rather than a rule; however, in the case of potentially recent clonal types such as CC8/"USA300" an acceptable level of typing/discriminatory quality might be achievable.

Enterococcus faecium

Population Structure and Typing of E. faecium

Of the two medically important enterococcal species only *E. faecium* has shown a tendency toward increasing multidrug and vancomycin resistance. This and a growing number of infections with resistant strains of *E. faecium* have led to increased efforts to improve the diagnostics of *E. faecium*. The increasing number of infections with vancomycin-resistant *E. faecium* strains (VREfm) is primarily related to hospital-associated strains of *E. faecium* (Lester et al. 2008; Top et al. 2008; de Regt et al. 2012; Weisser et al. 2012): such strains are more often multiresistant and exhibit an enhanced spreading potential within the nosocomial setting (van Schaik and Willems 2010; Willems and van Schaik 2009; Willems et al. 2012). Ampicillin and high-level ciprofloxacin resistance are phenotypic markers for these hospital strain types (Galloway-Pena et al. 2009; Leavis et al. 2006; Lester et al. 2008; Werner et al. 2010; Zhang et al. 2012). In addition, hospital strains differ from colonizing variants by a specific core genome as assessed by MLST, previously designated as "clonal complex CC17," and an additional accessory genome of several 100 kb including a pathogenicity island, other genomic islands, prophages, and several plasmids encoding virulence genes, metabolic properties, and resistance determinants (de Been et al. 2013; Zhang et al. 2013; van Schaik and Willems 2010; van Schaik et al. 2010; Werner et al. 2011; Laverde Gomez et al. 2011; Freitas et al. 2010; Lebreton et al. 2013; Gilmore et al. 2013; Heikens et al. 2008). This is in contrast to *E. faecalis* where a differentiation between colonizing and hospital strain variants is much less determined by differences in the genome content (Kuch et al. 2012; Tedim et al. 2014; Ruiz-Garbajosa et al. 2006; Zischka et al. 2015; Mikalsen et al. 2015). One of the working hypotheses suggests that the mentioned genomic differences between commensal and hospital-associated *E. faecium* strains result in specific protein peak patterns, for instance, derived from expressed genes linked to the accessory genome in hospital-associated variants (or by differences in the core genome). To test this hypothesis, we investigated in a previous study (Lasch et al. 2014) a diversity of well-characterized *E. faecium* strains from different animal species, food products, and human sources. Our aim was to investigate the potential of MALDI-TOF MS for differentiating hospital-associated (previously designated "CC17") and commensal ("non-CC17") strains of *E. faecium*.

MS Biomarkers for Strain or CC Allocation of E. faecium and Prediction of Vancomycin Resistance

In our previous study, we analyzed 112 well-characterized *E. faecium* isolates including *vanA*- and *vanB*-type VREfm. The analysis of MALDI-TOF MS spectra of all 112 isolates involved examinations of pseudo-gel views. The pseudo-gel view of Fig. 8.4 displayed the mass spectra from enterococci in the mass range of m/z 4000–9500 from commensal ("non-CC17") and hospital strains of *E. faecium* ("CC17"). It demonstrated consistent and reproducible peak patterns for all isolates of *E. faecium*, but noticeable and evident differences between the two major groups

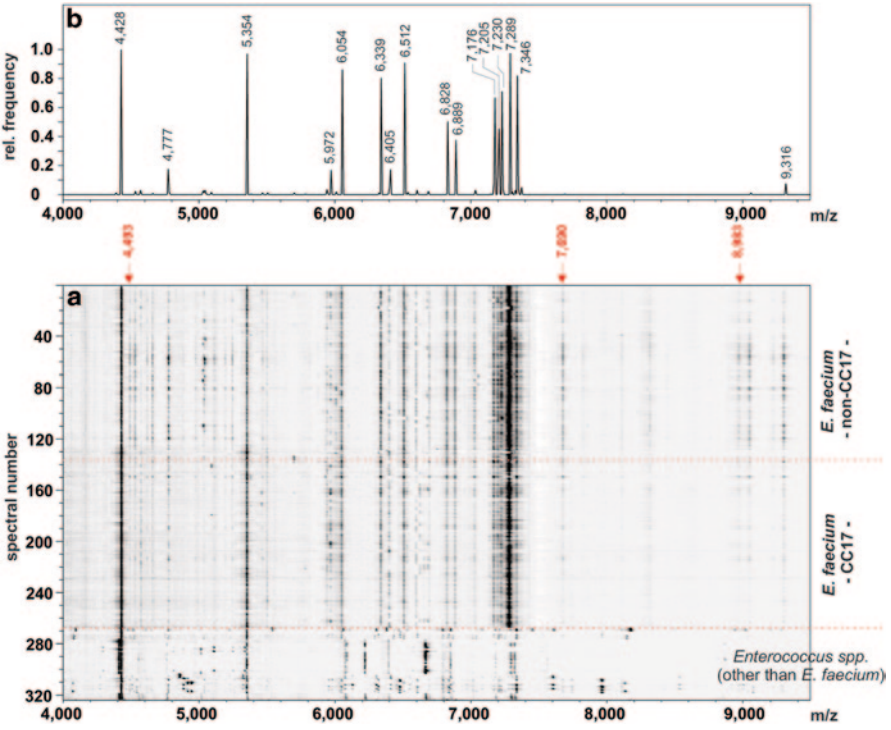


Fig. 8.4 **a** Gel view representation obtained from mass spectra of *E. faecium* (lines 1–266) and from spectra of *E. casseliflavus* (lines 267–269), *E. durans* (270–275), *E. faecalis* (276–302 and 318–322), *E. gallinarum* (303–305), and *E. hirae* (306–317). Spectral intensities from preprocessed MALDI-TOF mass spectra were gray scaled and plotted against the m/z values. **b** Peak frequency chart derived on the basis of peak tables with 15 peaks per spectrum obtained from 266 spectra of *E. faecium* isolates. The peak frequency curve was calculated by plotting $n+/n$ as a function of the mass/charge ratio m/z. The variable $n+$ denotes the number of spectra for which a peak was found at the respective m/z position. The symbol n indicates the total number of spectra ($n=266$). *Red arrows* denote selected biomarker peaks identified for commensal ("non-CC17") and hospital-associated ("CC17") strains of *E. faecium* (see, Lasch et al. 2014, for further details). Reproduced from Lasch et al. (2014) with permissions

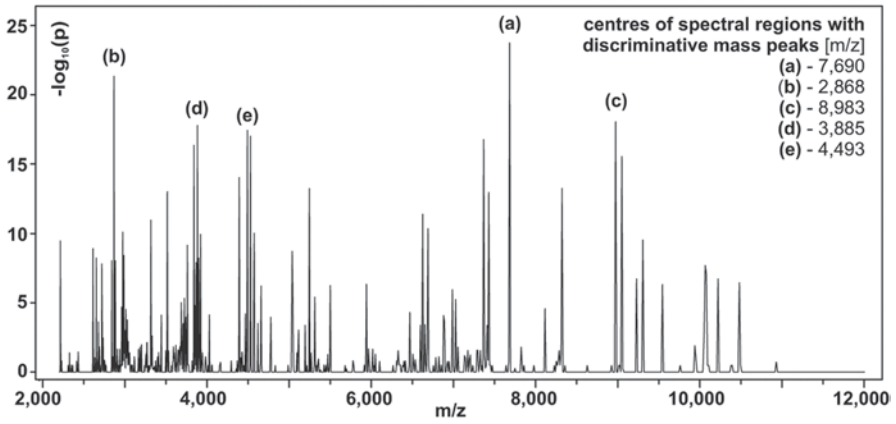


Fig. 8.5 Identification of the most discriminative mass spectral features between commensal ("non-CC17") and hospital-associated ("CC17") strains of *E. faecium*. For this purpose, univariate *t*-tests were carried out on the basis of altogether 266 mass spectra from which peak tables with 50 entries per mass spectrum were extracted. Independent *t*-tests were systematically carried out for each m/z region (width of 700 ppm). The *p* values of the *t*-tests were plotted against the centers of the m/z regions. Small *p* values cast doubt on the null-hypothesis of equal class means. Note the inverse logarithmic scaling. Reproduced from Lasch et al. (2014), with permission

of commensal and hospital strains were not found. The relative frequency of the m/z peaks in spectra of *E. faecium* is shown in Fig. 8.4b. This chart was derived from peak tables of *E. faecium* spectra which each contained the 15 most intense peaks of a given spectrum. Among these peaks, the following most prevalent peaks were identified at m/z 4428 (100% frequency), 7289 (97%), 5354 (97%), 3645 (95%), and 6512 (91%) (cf. Fig. 8.4b). Systematic statistical investigations for biomarkers with a discriminating potential between hospital-associated ("CC17") and commensal ("non-CC17") strains of *E. faecium* were carried out by a series of independent t-tests. These tests were carried out based on peak tables containing 50 peaks per spectrum (Lasch et al. 2014). The m/z dependence of the *p* values of the *t*-tests revealed putative biomarkers at m/z 7690, 2868, 8983, 3885, and 4493 (see Fig. 8.5). All these peaks have in common (1) a relatively low intensity (cf. Fig. 8.4, *red arrows*) and (2) an only limited discriminative potential as indicated by the low –log(*p*) values (Lasch et al. 2014).

Unsupervised hierarchical cluster analysis (UHCA) was carried out using all spectra of *E. faecium* and two additional spectra of preparations from *E. casseliflavus* and *E. faecalis*. Cluster analysis demonstrated notable differences between mass spectra of *E. faecium* and the other enterococcal species and, in addition to that, the presence of three clusters formed by spectra of *E. faecium* (Lasch et al. 2014). Among them, two large clusters were found dominated either by spectra from commensal "non-CC17," or by hospital-associated "CC17" strains, respectively. These results documented that identification of hospital-associated *E. faecium* by MALDI-TOF MS could, in principle, be achieved, but cluster differentiation required rigorous standardization and reached only a limited specificity. Further-

more, the data suggested that further sub-differentiation, such as into various MLST types or CC, is not feasible.

In two recent articles, authors aimed at discriminating between vancomycin-susceptible and vancomycin-resistant strains of *E. faecium* (VSEfm vs. VREfm) using MALDI-TOF MS spectral analyses and thus postulated to discriminate within the group of hospital-associated strains of *E. faecium* (Griffin et al. 2012; Nakano et al. 2014). Griffin and coworkers described MALDI-TOF MS as a suitable method for outbreak analysis of VRE isolates and for differentiating between *vanA* and *vanB* genotypes (Griffin et al. 2012). Comparisons between spectra of *vanA*-positive and *vanB*-positive isolates revealed a discriminating biomarker peak at m/z 5095 which was found in the examined *vanB*-positive, but not in *vanB*-negative strains. Another peak at m/z 6603 was claimed to be sensitive and specific for *vanB*-isolates of VRE and was thus suggested for discrimination from *vanA*-positive VREfm (Griffin et al. 2012). It must be noted, however, that MALDI-TOF MS cluster formation and differentiation was not compared to classical typing results from MLST or multiple-locus variable number tandem repeat analysis (MLVA). Furthermore, typing by PFGE was only performed for some of the isolates (but not shown). In the second study Nakano et al. investigated *vanA*-positive *E. faecium* and VSEfm strains from routine surveillance by MALDI-TOF MS and used three ClinProTools models for differentiation between the corresponding spectra yielding >90% of "recognition capability" (= sensitivity) and "cross-validation" (= specificity) (Nakano et al. 2014). The putative marker peaks described by Nakano et al. (m/z 3184, 5702, 7415, 7445, 12,662) were also different from Griffin et al.

In our recent study, we were unable to confirm the postulated genotype specificity of the peaks at m/z 5095 and m/z 6603: mass spectra available to our groups did not reveal evidences for the specificity of these biomarkers (see Fig. 8.6). While it seems that the peak at m/z 6603 is more frequently present in *vanB*-positive isolates, we have observed a large number of strains from VSEfm and *vanA*-positive strains which also exhibit this particular signal (see Fig. 8.6). Furthermore, the signal at m/z 5095 was found in only two out of 65 isolates: strain VREfm UW 1983 (*vanA*+) and VREfm UW 5662 (*vanB*+, see Fig. 8.6). We interpret the discrepancies between our experimental observations and published data as a result of a very different strain collection. Obviously, a broad and comprehensive strain collection in terms of human and animal sources, different geographical origin, etc., plays a decisive role (Lasch et al. 2014).

We also systematically compared MALDI-TOF mass spectra of two pairs of isogenic *E. faecium* strains which had received conjugative *vanA* plasmids (Werner et al. 2011) or conjugative *vanB* transposons. For pair 1 the recipient strain and all transconjugants ($n=9$) showed highly monomorphic MALDI patterns with slight but strain-specific differences independent from *vanA* or *vanB* gene acquisition (data not shown). These and similar findings for pair 2 further support our suggestion that *vanA*- and/or *vanB*-type resistance is not predictable from MALDI-TOF MS spectra. We postulate that differences identified within the corresponding spectra of VSEfm and VREfm strains described in previous studies were due to the observation of covariates in the strain selections examined, that is, strain-specific

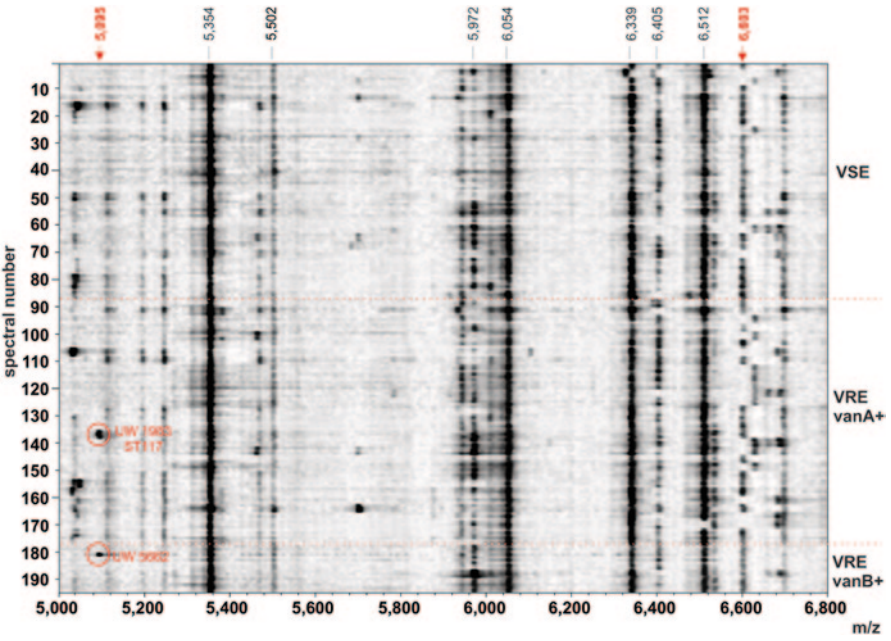


Fig. 8.6 Pseudo-gel view produced from 195 mass spectra of *E. faecium*, among them are 29 strains of VSEfm (lines 1–87) and 36 strains of VREfm (lines 88–195). Among the VREfm strains 30 were of the *vanA* (lines 88–177) and 6 of the *vanB* type (lines 178–195). The figure demonstrates the absence of *vanA* and *vanB* marker peaks postulated by Griffin et al. (2012) in the mass spectra of the RKI strain collection. *Arrows* at m/z 5095 and 6603 denote the position of the proposed marker peaks

marker peaks of the different VSEfm and VREfm strains which are independent of the *van* genotype. In addition, we would like to stress two more facts. Marker proteins encoded by the corresponding resistance genotypes would in general be too large to be directly detected by MALDI-TOF MS in the given m/z window. For example, the molecular mass of the VanA ligase protein is 40 kDa, whereas a typical microbial MALDI-TOF spectrum is recorded in the mass range between m/z 2000 and 20,000. Other proteins are membrane bound and are thus often hard to ionize (e.g., the VanS sensor protein). Furthermore, the resistance genotype is generally inducible by vancomycin exposure; otherwise, the resistance phenotype is expressed at a very basic level. To our knowledge, such inducing growth conditions were not considered in the studies carried out so far. It is conceivable that this mechanism allows differentiation between sensitive and resistant strains of certain microbial species.

In summary, we must conclude that the expectation to measure differences encoded by the core and accessory genome of commensal vs. hospital isolates of *E. faecium* by MALDI-TOF MS analyses using standard diagnostic protocols for sample preparation is not generally achievable. Recent experimental data suggest that individual strains of *E. faecium* might produce specific and stable MALDI-TOF

MS peak patterns which could, in principle, be used for strain typing. However, we have shown that the majority of *E. faecium* strains showed highly similar patterns indistinguishable by the standard MALDI-TOF MS spectral analysis pipeline for diagnostic purposes. According to our own experiences and the experimental data shown, vancomycin resistance prediction does not seem possible using the standard methodology of MALDI-TOF MS.

Differentiation of *Bacillus anthracis, Bacillus cereus* sensu stricto and *Bacillus thuringiensis*

The so-called *B. cereus* group occupies an important place within the genus *Bacillus*. Apart from *B. cereus*, after which this group has been named, the *B. cereus* group comprises five additional species: *B. anthracis*, *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides*, and *B. weihenstephanensis* (Vilas-Bôas et al. 2007). While *B. cereus* sensu stricto (s.s.) is ubiquitous in nature and clinically relevant because of its association with food poisoning (Rasko et al. 2005; Granum and Lund 1997), *B. thuringiensis* is known as a producer of a parasporal toxin which is effective against a variety of insect larvae (Vilas-Bôas et al. 2007; Schnepf et al. 1998; Hill et al. 2004). The monophyletic *B. anthracis* is the causative agent of anthrax, which can be lethal to humans and animals. *B. anthracis* is considered a potential biological weapon; in 2001 letters containing anthrax spores were mailed to US senators and news media offices causing illness in 22 persons and leading to 5 casualties (Bartlett et al. 2002; Kennedy 2001). It has been postulated that phenotype and pathogenicity factors of the latter three species are often plasmid encoded. For example, fully virulent *B. anthracis* strains harbor two plasmids, pXO1 and pXO2, which encode the toxin plasmid, pXO1 (Okinaka et al. 1999), and the capsule plasmid, pXO2 (Makino et al. 1989).

Several studies have suggested that *B. cereus* in the strict sense, *Bacillus cereus* s.s., *B. anthracis,* and *B. thuringiensis* represent one single species, designated *B. cereus* s.l. (Helgason et al. 2000; Bavykin et al. 2004; Daffonchio et al. 2000). Other authors do not support this opinion as they obtained sufficient genetic discrimination between these three classical species (Radnedge et al. 2003; Chang et al. 2003; Kolsto et al. 2009). There is still some controversy among microbiologists regarding this taxonomic question. Conclusive clarification is additionally complicated by the recent detection of mosaic strains. For example, Leendertz and coworkers described *Bacillus* isolates from a new west and central African habitat which were responsible for the death of wild great apes in Côte d'Ivoire and Cameroon (Leendertz et al. 2006; Leendertz et al. 2004). These strains share some genetic properties of *B. cereus* s.s. and *B. anthracis* (Klee et al. 2006; Leendertz et al. 2006) and have recently been designated as *B. cereus* biovar *anthracis* (Klee et al. 2010).

The differentiation and identification of *Bacillus* species by MALDI-TOF MS has been the subject of many studies, and there are multiple pieces of evidence that MALDI-TOF MS can be successfully employed to identify species such as *B.*

anthracis (Callahan et al. 2008; Castanha et al. 2006, 2007; Elhanany et al. 2001; Hathout et al. 1999, 2003; Ryzhov et al. 2000; Lasch et al. 2009), *B. subtilis,* or *B. licheniformis* (Fernandez-No et al. 2013). For example, Ryzhov et al. (2000) and Demirev et al. (2001) could identify members of the small, acid-soluble protein (SASP) family as mass spectral biomarker candidates which were later found sensitive and specific for *Bacillus* species such as *B. anthracis* (β-SASP @ m/z 6679) or *B. subtilis* (γ-SASP @ m/z 9137.5). SASPs are known as proteins present in endospores of various *Bacillus* species in very high concentrations, up to 20% of the total spore core protein content (Moeller et al. 2009). During the germination of spores SASPs serve as amino acid resources (Setlow 1988); furthermore, SASPs have been shown to confer resistance to UV radiation, heat, peroxides, and other harsh treatments (Moeller et al. 2009; Setlow 1988).

Today, MS-based identification works sufficiently well for most of the relevant *Bacillus* species. For example, the commercial BioTyper database with mass spectral data of clinically important microbial strains (Maier et al. 2006) (Bruker Daltonics) currently comprises entries from a total of 106 different species from the genus *Bacillus* (BioTyper database version v4.0.0.1; 4613–5627). We generally found that this high degree of coverage allows successful differentiation and identification of the majority of the known *Bacillus* species; however, sub-differentiation of *B. cereus* s.l. into the classical species of *B. anthracis*, *B. cereus* s.s., and *B. thuringiensis* still constitutes a significant challenge not only because of the close phylogenetic relationships but also because of the lack of spectral entries from highly pathogenic *B. anthracis* strains in Bruker's standard BioTyper database.

In a study published in 2009 which involved MALDI-TOF MS on 374 strains from *Bacillus* and related genera, including 102 strains of *B. anthracis* and 121 strains of *B. cereus* s.s., we were able to identify *B. cereus* group-specific as well as *B. anthracis*-specific biomarker candidates that allowed rapid, sensitive, and reliable identification (Lasch et al. 2009). For example, *B. cereus* group-specific signals were detected at m/z 5171, 5886, and 7368, whereas mass spectra of *B. anthracis* exhibited specifically identifying biomarkers at m/z 4606, 5413, and 6679. Even at the time of publication, it was found, however, that the concept of the existence of exclusive MS biomarkers for *B. anthracis* was not tenable: in our publication, we have pointed out that the statistically most significant marker of *B. anthracis*, the β-SASP peak at m/z 6679, was observable also in two strains of *B. cereus* s.s. (*B. cereus* B292 and B248; strain B248 is identical with *B. cereus* DSM 8438). Moreover, in further studies we detected even more strains, among them are *B. cereus* BW-B and *B. cereus* biovar *anthracis*, which also exhibited the postulated *B. anthracis*-specific β-SASP signal (see Fig. 8.7). Although it was discovered later that our findings on strain *B. cereus* B292 were not reproducible, probably because of mistaken identity (unpublished data, personal communication with W. Beyer and M. Dybwad), the presence of a β-SASP-signal at m/z 6679 in spectra of selected non-*B. anthracis* strains is backed by observations of other authors (Callahan et al. 2008, 2009; Dybwad et al. 2013; Chenau et al. 2013). For example, Dybwad et al. observed this signal when studying *B. cereus* DSM 8438, *B. thuringiensis* BGSC 4CC1, and *B. cereus* R3 (Dybwad et al. 2013). As an illustration, Fig. 8.8 shows the

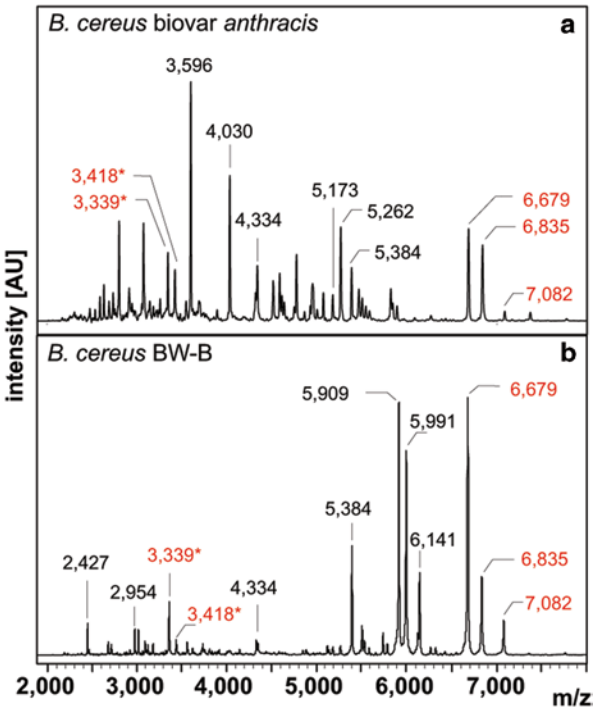


Fig. 8.7 MALDI-TOF mass spectra of *B. cereus* biovar *anthracis* (Klee et al. 2010) isolated from great apes that died from an anthrax-like disease *(upper panel)* and of the *B. cereus* strain BW-B *(lower panel)*. The preprocessed spectra of both strains exhibit SASP spore marker peaks (SASP: small acid-soluble proteins) at m/z 6679 (β-SASP), 6835 (α-SASP), and 7082 (α/β-SASP) indicated by red numbers. The β-SASP peak at m/z 6679 has been previously described as a potential biomarker candidate that is specific for *B. anthracis* due to the consistent observation of one or two amino acid substitutions in the β-SASP of *B. cereus* sensu stricto (s.s.) and *B. thuringiensis* (Castanha et al. 2006). The observation of a β-SASP at m/z 6679 in strains of *B. cereus* s.s. suggests that the specificity of this biomarker for *B. anthracis* is below 100 %. *AU* arbitrary units, * SASP peaks from doubly charged species

pseudo-gel view representation of the *B. cereus* s.l. database currently available to the authors of this study. The figure clearly demonstrates for most of the spectra the *B. cereus* group-specific signals at m/z 5171, 5886, and 7368. Furthermore, SASP spore marker peaks are discernible at m/z 6679, 6695, and 6711 (β-SASP), m/z 6835 (α-SASP) and at 7082 (α/β-SASP). The gel view illustrates also the presence of the *B. anthracis*-specific signal at m/z 6679 in spectra of selected strains of *B. cereus* s.s. such as *B. cereus* DSM 8438 and of *B. cereus* biovar *anthracis* (see arrows 1 and 2, respectively).

While SASP signals can be used as sensitive indicators for the presence of spores, the second important *B. anthracis* biomarker at m/z 5413 can be detected also when spores are absent (Lasch et al. 2009), cf. Fig. 8.8. Using nHPLC-MS/MS techniques, the signal at m/z 5413 has been assigned as a putative uncharacterized pro-

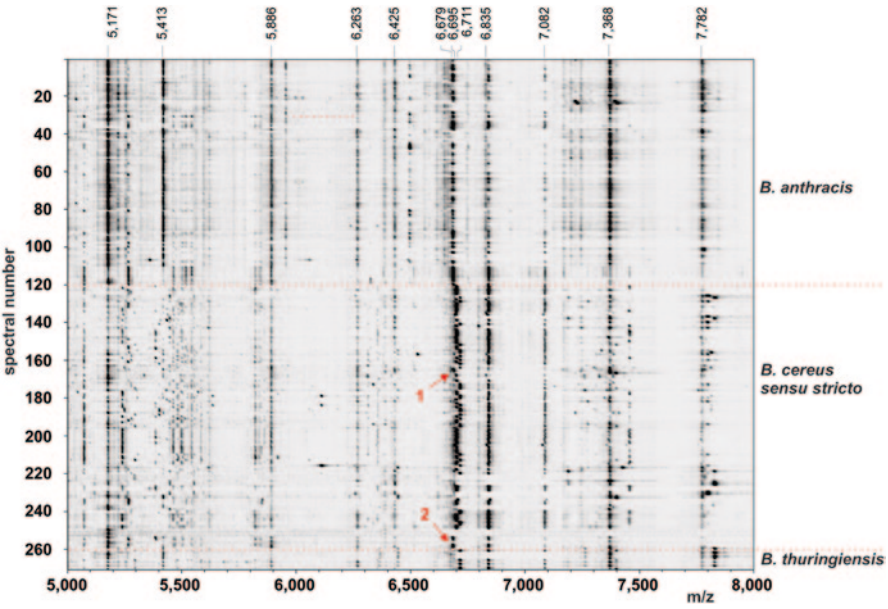


Fig. 8.8 Pseudo-gel view of mass spectra from *B. cereus* sensu lato (s.l.) in the diagnostically relevant mass range of m/z 5000–8000. Mass spectra of the three classical species *B. anthracis*, *B. cereus* sensu strict, (s.s.), and *B. thuringiensis* consistently exhibit marker peaks at m/z 5171, 5886, 6263, 6835, 7082, 7368, and 7782. The peak at m/z 6679 has been suggested as a biomarker specific for *B. anthracis*; however, there are a few strains of *B. cereus* in which the postulated biomarker was also observed (*arrows 1 and 2*, see *also* the text for details)

tein Q81U79 in the *B. anthracis* strain Ames Ancestor (cf. UniProtKB/Swiss-Prot database). Interestingly, a blast search of Q81U79 demonstrated 100% sequence identity with putatively uncharacterized proteins of at least 85 different strains of *B. anthracis*, *B. cereus* s.s., and *B. thuringiensis* (unpublished data). Moreover, experimental data acquired in the laboratory of the authors, and those of others, (Lasch et al. 2015) consistently revealed the presence of this specific biomarker in many strains of *B. cereus* s.s. and *B. thuringiensis*. Thus, experimental evidences as well as protein sequence data argue against a specific signal for *B. anthracis* at m/z 5413. On the other hand, as the gel view in Fig. 8.8 clearly suggests about the species specificity of this signal, we believe that further investigations are required to clarify the taxonomic significance of the mass peak at m/z 5413.

In summary, we have demonstrated that differentiation and unequivocal identification of taxonomically closely related species within *B. cereus* s.l. by MALDI-TOF MS constitutes a considerable challenge. Although data from a number of laboratories have raised reasonable doubts on the validity of originally postulated mass spectral biomarkers for *B. anthracis*, it has been demonstrated that advanced methods of multivariate pattern recognition such as neural network analyses (Lasch et al. 2009), or decision-tree techniques optimized on the basis of similarity-grouped reference libraries (Dybwad et al. 2013), represent appropriate data analysis tools

for reliable identification of *B. anthracis*. These findings indicate that MALDI-TOF MS can be used for reliable and statistically accurate (95 %) differentiation of the monophyletic species of *B. anthracis* from other members of *B. cereus* s.l. However, at this stage, it is not clear whether and how the two other species of *B. cereus* s.l., *B. cereus* s.s., and *B. thuringiensis* can be differentiated by MALDI-TOF MS. Our own preliminary investigations carried out on an only limited number of strains from *B. thuringiensis* (11 strains) suggested that both species are difficult to differentiate. However, further studies on a larger and a more representative strain collection are required to finally answer this question.

Summary and Outlook

In the context of the present mini-review, we have discussed the challenges and opportunities associated with the application of MALDI-TOF MS as a routine application in clinical/microbiological diagnostics for bacterial strain typing below the species level. Based on three examples of important bacterial pathogens, *S. aureus*, *E. faecium,* and *Bacillus cereus* s. l., we have explored the strengths and weaknesses of this evolving technique. It was found that the taxonomic resolution of MALDI-TOF MS is limited which often prevents reliable strain categorization or strain identification/typing. Only in selected cases we were able to detect mass spectral biomarkers and fingerprints that allowed for reliable microbial typing and strain differentiation. It was found that typing below the species level requires rigorous standardization, high spectral quality and the application of adequate sample processing methods, spectra acquisition procedures, and an optimized data analysis pipeline that includes multivariate fingerprinting approaches. However, these conditions cannot be expected in a routine setting. Therefore, future efforts to increase the taxonomic resolution should involve the application of alternative MS technologies, ideally in combination with sample preprocessing methods, such as separation techniques. It is hoped that this will open new avenues to improve the taxonomic resolution of MS-based technologies.

Acknowledgments The authors wish to thank B. Niederwöhrmeier (Bundeswehr Research Institute for Protective Technologies and NBC Protection in Munster, WIS), M. Ehling-Schulz (University of Veterinary Medicine Vienna), and W. Beyer (Institute for Environmental and Animal Hygiene and Veterinary Medicine at the University of Hohenheim) for providing *Bacillus* strains. The authors are furthermore grateful to T. Maier, M. Kostrzewa (Bruker Daltonik GmbH), T. Schwecke (RKI Berlin), and R. Dieckmann (Federal Institute for Risk Assessment, Department of Biological Safety) for fruitful discussions and support. Furthermore, the excellent technical assistance of P. Lochau, S. Becker, R. Heinrich, M. Stämmler, A. Brauer, A. Schneider, and S. Herfort (all RKI Berlin) is acknowledged.

References

Anderson NW, Buchan BW, Riebe KM, Parsons LN, Gnacinski S, Ledeboer NA. Effects of solid-medium type on routine identification of bacterial isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50(3):1008–13. doi:10.1128/JCM.05209-11.

Arnold RJ, Reilly JP. Fingerprint matching of *E. coli* strains with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. Rapid Commu Mass Spectrom. 1998;12(10):630–36.

Auda G, Kamel M. Modular neural network classifiers: a comparative study. J Intell Robot Syst. 1998;21(2):117–29. doi:10.1023/A:1007925203918.

Balazova T, Makovcova J, Sedo O, Slany M, Faldyna M, Zdrahal Z. The influence of culture conditions on the identification of *Mycobacterium* species by MALDI-TOF MS profiling. FEMS Microbiol Lett. 2014. doi:10.1111/1574-6968.12408.

Bartlett JG, Inglesby TV, Jr., Borio L. Management of anthrax. Clin Infect Dis. 2002;35(7):851–58.

Bavykin SG, Lysov YP, Zakhariev V, Kelly JJ, Jackman J, Stahl DA, Cherni A. Use of 16S rRNA, 23S rRNA, and gyrB gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. J Clin Microbiol. 2004;42(8) (37):11–30.

Bittar F, Ouchenane Z, Smati F, Raoult D, Rolain JM. MALDI-TOF-MS for rapid detection of staphylococcal Panton–Valentine leukocidin. Int J Antimicrob Agents. 2009;34 (5):467–70. doi:10.1016/j.ijantimicag.2009.03.017.

Blanco R, Tristan A, Ezpeleta G, Larsen AR, Bes M, Etienne J, Cisterna R, Laurent F. Molecular epidemiology of Panton-Valentine leukocidin-positive *Staphylococcus aureus* in Spain: emergence of the USA300 clone in an autochthonous population. J Clin Microbiol. 2011;49(1):433–36. doi:10.1128/JCM.02201-10.

Boggs SR, Cazares LH, Drake R. Characterization of a *Staphylococcus aureus* USA300 protein signature using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Med Microbiol. 2012;61(Pt 5):640–44. doi:10.1099/jmm.0.037978-0.

Bohme K, Fernandez-No IC, Pazos M, Gallardo JM, Barros-Velazquez J, Canas B, Calo-Mata P. Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing versus MALDI-TOF MS fingerprinting. Electrophoresis. 2013;34(6):877–87. doi:10.1002/elps.201200532.

Brandt A. Die Untersuchung von mikrobiellen Wachstumsbedingungen mittels MALDI-TOF-Massenspektrometrie. Diploma Thesis Technical University Berlin, Germany; 2006.

Cain TC, Lubman DM, Weber WJ, Vertes A. Differentiation of bacteria using protein profiles from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1994;8(12):1026–30. doi:10.1002/rcm.1290081224

Callahan C, Castanha ER, Fox KF, Fox A. The *Bacillus cereus* containing sub-branch most closely related to *Bacillus anthracis*, have single amino acid substitutions in small acid-soluble proteins, while remaining sub-branches are more variable. Mol Cell Probes. 2008;22(3):207–11.

Callahan C, Fox K, Fox A. The small acid soluble proteins (SASP alpha and SASP beta) of *Bacillus weihenstephanensis* and *Bacillus mycoides* group 2 are the most distinct among the *Bacillus cereus* group. Mol Cell Probes. 2009;23(6):291–97. doi:10.1016/j.mcp.2009.07.003.

Castanha ER, Fox A, Fox KF. Rapid discrimination of *Bacillus anthracis* from other members of the *B. cereus* group by mass and sequence of "intact" small acid soluble proteins (SASPs) using mass spectrometry. J Microbiol Methods. 2006;67 (2):230–40.

Castanha ER, Vestal M, Hattan S, Fox A, Fox KF, Dickinson D. *Bacillus cereus* strains fall into two clusters (one closely and one more distantly related) to *Bacillus anthracis* according to amino acid substitutions in small acid-soluble proteins as determined by tandem mass spectrometry. Mol Cell Probes. 2007;21(3):190–201.

Chang YH, Shangkuan YH, Lin HC, Liu HW. PCR assay of the groEL gene for detection and differentiation of *Bacillus cereus* group cells. Appl Environ Microbiol. 2003;69(8):4502–10.

Chenau J, Fenaille F, Caro V, Haustant M, Diancourt L, Klee SR, Junot C, Ezan E, Goossens PL, Becher F. Identification and validation of specific markers of *Bacillus anthracis* spores by proteomics and genomics approaches. Mol Cell Proteomics. 2013. doi:10.1074/mcp.M113.032946.

Claydon MA, Davey SN, Edwards-Jones V, Gordon DB. The rapid identification of intact microorganisms using mass spectrometry. Nat Biotechnol. 1996;14(11):1584–86. doi:10.1038/nbt1196-1584.

Daffonchio D, Cherif A, Borin S. Homoduplex and heteroduplex polymorphisms of the amplified ribosomal 16S-23S internal transcribed spacers describe genetic relationships in the "*Bacillus cereus* group". Appl Environ Microbiol. 2000;66(12):5460–68.

Dauwalder O, Carbonnelle E, Benito Y, Lina G, Nassif X, Vandenesch F, Laurent F. Detection of Panton-Valentine toxin in *Staphylococcus aureus* by mass spectrometry directly from colony: time has not yet come. Int J Antimicrob Agents. 2010;36(2):193–94. doi:10.1016/j.ijantimicag.2010.04.005.

de Been M, van Schaik W, Cheng L, Corander J, Willems RJ. Recent recombination events in the core genome are associated with adaptive evolution in *Enterococcus faecium*. Genome Biol Evol. 2013;5(8):1524–35. doi:10.1093/gbe/evt111.

de Regt MJ, van Schaik W, van Luit-Asbroek M, Dekker HA, van Duijkeren E, Koning CJ, Bonten MJ, Willems RJ. Hospital and community ampicillin-resistant *Enterococcus faecium* are evolutionarily closely linked but have diversified through niche adaptation. PloS ONE.2012;7(2):e30319. doi:10.1371/journal.pone.0030319.

Demirev PA, Feldman AB, Kowalski P, Lin JS. Top-down proteomics for rapid identification of intact microorganisms. Anal Chem. 2005;77(22):7455–61. doi:10.1021/ac051419g.

Demirev PA, Ho YP, Ryzhov V, Fenselau C. Microorganism identification by mass spectrometry and protein database searches. Anal Chem. 1999;71(14):2732–38.

Demirev PA, Ramirez J, Fenselau C. Tandem mass spectrometry of intact proteins for characterization of biomarkers from *Bacillus cereus* T spores. Anal Chem. 2001;73(23):5725–31.

Dieckmann R, Helmuth R, Erhard M, Malorny B. Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2008;74(24):7767–78. doi:10.1128/AEM.01402–08.

Du Z, Yang R, Guo Z, Song Y, Wang J. Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Chem. 2002;74(21):5487–91.

Dybwad M, van der Laaken AL, Blatny JM, 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. 2013;79(17):5372–83. doi:10.1128/AEM.01724–13.

Elhanany E, Barak R, Fisher M, Kobiler D, Altboum Z. Detection of specific *Bacillus anthracis* spore biomarkers by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2001;15(22):2110–16.

Fenselau C. Mass spectrometry for the characterization of microorganisms. ACS Symposium Series 541, American Chemical Society: Washington DC 541; 1994.

Fernandez-No IC, Bohme K, Diaz-Bao M, Cepeda A, Barros-Velazquez J, Calo-Mata P. Characterisation and profiling of *Bacillus subtilis*, *Bacillus cereus* and *Bacillus licheniformis* by MALDI-TOF mass fingerprinting. Food Microbiol. 2013;33(2):235–42. doi:10.1016/j.fm.2012.09.022.

File TM, Jr., Wilcox MH, Stein GE. Summary of ceftaroline fosamil clinical trial studies and clinical safety. Clin Infect Dis. 2012;55(Suppl 3):S. 173–80. doi:10.1093/cid/cis559.

Freitas AR, Tedim AP, Novais C, Ruiz-Garbajosa P, Werner G, Laverde-Gomez JA, Canton R, Peixe L, Baquero F, Coque TM. Global spread of the hyl(Efm) colonization-virulence gene in megaplasmids of the *Enterococcus faecium* CC17 polyclonal subcluster. Antimicrob Agents Chemother. 2010;54(6):2660–65. doi:10.1128/AAC.00134-10.

Gagnaire J, Dauwalder O, Boisset S, Khau D, Freydiere AM, Ader F, Bes M, Lina G, Tristan A, Reverdy ME, Marchand A, Geissmann T, Benito Y, Durand G, Charrier JP, Etienne J, Welker M, Van Belkum A, Vandenesch F. Detection of *Staphylococcus aureus* delta-toxin production by whole-cell MALDI-TOF mass spectrometry. PloS ONE. 2012;7(7):e40660. doi:10.1371/journal.pone.0040660.

Galloway-Pena JR, Nallapareddy SR, Arias CA, Eliopoulos GM, Murray BE. Analysis of clonality and antibiotic resistance among early clinical isolates of *Enterococcus faecium* in the United States. J Infect Dis. 2009;200(10):1566–73. doi:10.1086/644790.

Gilmore MS, Lebreton F, van Schaik W. Genomic transition of *Enterococci* from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. Curr Opin Microbiol. 2013;16(1):10–16. doi:10.1016/j.mib.2013.01.006.

Goldstein JE, Zhang L, Borror CM, Rago JV, Sandrin TR. Culture conditions and sample preparation methods affect spectrum quality and reproducibility during profiling of *Staphylococcus aureus* with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Lett App Microbiol. 2013;57(2):144–50. doi:10.1111/lam.12092.

Goodacre R. Characterisation and quantification of microbial systems using pyrolysis mass spectrometry: introducing neural networks to analytical pyrolysis. Microbiol Eur. 1994;2(2):16–22

Goodacre R, Timmins EM, Burton R, Kaderbhai N, Woodward AM, Kell DB, Rooney PJ. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. Microbiology. 1998;144(Pt 5):1157–1170.

Granum PE, Lund T. *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol Lett. 1997;157(2):223–28.

Griffin PM, Price GR, Schooneveldt JM, Schlebusch S, Tilse MH, Urbanski T, Hamilton B, Venter D. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant *Enterococci* and investigate the epidemiology of an outbreak. J Clin Microbiol. 2012;50(9):2918–31. doi:10.1128/JCM.01000-12.

Grundmann H, Aanensen DM, van den Wijngaard CC, Spratt BG, Harmsen D, Friedrich AW, European Staphylococcal Reference Laboratory Working Group. Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. PLoS Med. 2010;7(1):e1000215. doi:10.1371/journal.pmed.1000215.

Harmsen D, Claus H, Witte W, Rothganger J, Claus H, Turnwald D, Vogel U. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. J Clin Microbiol. 2003;41(12):5442–48.

Hathout Y, Demirev PA, Ho YP, Bundy JL, Ryzhov V, Sapp L, Stutler J, Jackman J, Fenselau C. Identification of *Bacillus* spores by matrix-assisted laser desorption ionization-mass spectrometry. Appl Environ Microbiol. 1999;65(10):4313–19

Hathout Y, Setlow B, Cabrera-Martinez RM, Fenselau C, Setlow P. Small, acid soluble proteins as biomarkers in mass spectrometry analysis of *Bacillus* spores. Appl Environ Microbiol. 2003;69(2):1100–07.

Heikens E, van Schaik W, Leavis HL, Bonten MJ, Willems RJ. Identification of a novel genomic island specific to hospital-acquired clonal complex 17 *Enterococcus faecium* isolates. Appl Environ Microbiol. 2008;74(22):7094–97. doi:10.1128/AEM.01378-08.

Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto AB. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. Appl Environ Microbiol. 2000;66(6):2627–30.

Helm D, Labischinski H, Schallehn G, Naumann D. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. J Gen Microbiol. 1991;137(1):69–79.

Hill KK, Ticknor LO, Okinaka RT, Asay M, Blair H, Bliss KA, Laker M, Pardington PE, Richardson AP, Tonks M, Beecher DJ, Kemp JD, Kolsto AB, Wong AC, Keim P, Jackson PJ. Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. Appl Environ Microbiol. 2004;70(2):1068–1080.

Hiramatsu K, Ito T, Tsubakishita S, Sasaki T, Takeuchi F, Morimoto Y, Katayama Y, Matsuo M, Kuwahara-Arai K, Hishinuma T, Baba T. Genomic basis for methicillin resistance in *Staphylococcus aureus*. Infect Chemother. 2013;45(2):117–36. doi:10.3947/ic.2013.45.2.117.

Holland RD, Wilkes JG, Rafii F, Sutherland JB, Persons CC, Voorhees KJ, Lay JO Jr. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1996;10(10):1227–32.

John PA, Catherine F. Identification of bacteria using mass spectrometry. Anal Chem. 1975;47(2):219–25

Josten M, Reif M, Szekat C, Al-Sabti N, Roemer T, Sparbier K, Kostrzewa M, Rohde H, Sahl HG, Bierbaum G. Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of *Staphylococcus aureus* identifies mutations that allow differentiation of the main clonal lineages. J Clin Microbiol. 2013;51(6):1809–1817. doi:10.1128/JCM.00518-13.

Jung JS, Eberl T, Sparbier K, Lange C, Kostrzewa M, Schubert S, Wieser A. Rapid detection of antibiotic resistance based on mass spectrometry and stable isotopes. Eur J Clin Microbiol Infect Dis. 2014;33(6):949–955. doi:10.1007/s10096-013-2031-5.

Kempka M, Sjodahl J, Bjork A, Roeraade J. Improved method for peak picking in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2004;18(11):1208–12. doi:10.1002/rcm.1467.

Kennedy H. Daschle letter bombshell—billions of anthrax spores. New York Daily News. 2001 Oct 31;5.

Keys CJ, Dare DJ, Sutton H, Wells G, Lunt M, McKenna T, McDowall M, Shah HN. Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. Infect Genet Evol. 2004;4(3):221–42.

Klee SR, Brzuszkiewicz EB, Nattermann H, Bruggemann H, Dupke S, Wollherr A, Franz T, Pauli G, Appel B, Liebl W, Couacy-Hymann E, Boesch C, Meyer FD, Leendertz FH, Ellerbrok H, Gottschalk G, Grunow R, Liesegang H. The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of B. cereus with *B. anthracis* virulence plasmids. PloS ONE. 2010;5(7):e10986. doi:10.1371/journal.pone.0010986.

Klee SR, Ozel M, Appel B, Boesch C, Ellerbrok H, Jacob D, Holland G, Leendertz FH, Pauli G, Grunow R, Nattermann H. Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. J Bacteriol. 2006;188(15):5333–44. doi:10.1128/JB.00303–06.

Kolsto AB, Tourasse NJ, Okstad OA. What sets *Bacillus anthracis* apart from other *Bacillus* species? Ann Rev Microbiol. 2009;63:451–76. doi:10.1146/annurev.micro.091208.073255.

Kostrzewa M, Sparbier K, Maier T, Schubert S. MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. Proteomics Clin Appl. 2013;(36):197–203. doi:10.1002/prca.201300042.

Krishnamurthy T, Ross PL, Rajamani U. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1996;10(8):883–8. doi:10.1002/(SICI)1097-0231(19960610)10:8<883::AID-RCM594>3.0.CO;2-V.

Kuch A, Willems RJ, Werner G, Coque TM, Hammerum AM, Sundsfjord A, Klare I, Ruiz-Garbajosa P, Simonsen GS, van Luit-Asbroek M, Hryniewicz W, Sadowy E. Insight into antimicrobial susceptibility and population structure of contemporary human *Enterococcus faecalis* isolates from Europe. J Antimicrob Chemother. 2012;67(3):551–8. doi:10.1093/jac/dkr544.

Lasch P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. Chemom Intell Lab Syst. 2012;117:100–14.

Lasch P. MicrobeMS: a Matlab Toolbox for analysis of microbial MALDI-TOF mass spectra. 2015. http://www.microbe-ms.com

Lasch P, Beyer W, Nattermann H, Stämmler M, Siegbrecht E, Grunow R, Naumann D. Identification of *Bacillus anthracis* by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. Appl Environ Microbiol. 2009;75(22):7229–42. doi:10.1128/AEM.00857-09.

Lasch P, Drevinek M, Nattermann H, Grunow R, Stammler M, Dieckmann R, Schwecke T, Naumann D. Characterization of *Yersinia* using MALDI-TOF mass spectrometry and chemometrics. Anal Chem. 2010;82(20):8464–75. doi:10.1021/ac101036s.

Lasch P, Fleige C, Stammler M, Layer F, Nübel U, Witte W, Werner G. Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of *Enterococcus faecium* and *Staphylococcus aureus* isolates. J Microbiol Methods. 2014;100:58–69. doi:10.1016/j.mimet.2014.02.015.

Lasch P, Nattermann H, Erhard M, Stammler M, Grunow R, Bannert N, Appel B, Naumann D. MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. Anal Chem. 2008;80(6):2026–34. doi:10.1021/ac701822j.

Lasch P, Naumann D. MALDI-TOF mass spectrometry for the rapid identification of highly pathogenic microorganisms. Proteomics, glycomics and antigenicity of BSL3 and BSL4 agents. First Edition Edited by Jiri Stulik, Rudolf Toman, Patrick Butaye, Robert G Ulrich 2011 Wiley-VCH Verlag GmbH & Co KGaA Published 2011 by Wiley-VCH Verlag GmbH & Co KGaA:219–2. 2011.

Lasch P, Wahab T, Hufnagl P, Pályi B, Tomaso H, Zange S, Kiland Langerud B, Drevinek M, Kokotovic B, Wittwer M, Pflüger V, Di Caro A, Stämmler M, Grunow R, Jacob D. Identification of highly pathogenic microorganisms using MALDI-TOF mass spectrometry—results of an Inter-Laboratory Ring Trial. J Clin Microbiol. 2015;53(8):2632-40. doi:10.1128/JCM.00813–15.

Laverde Gomez JA, van Schaik W, Freitas AR, Coque TM, Weaver KE, Francia MV, Witte W, Werner G. A multiresistance megaplasmid pLG1 bearing a hylEfm genomic island in hospital *Enterococcus faecium* isolates. Int J Med Microbiol. 2011;301(2):165–75. doi:10.1016/j.ijmm.2010.08.015.

Lay JO Jr. MALDI-TOF mass spectrometry of bacteria. Mass Spectrom Rev. 2001;20(4):172–94.

Leavis HL, Willems RJ, Top J, Bonten MJ. High-level ciprofloxacin resistance from point mutations in gyrA and parC confined to global hospital-adapted clonal lineage CC17 of *Enterococcus faecium*. J Clinical Microbiol. 2006;44(3):1059–64. doi:10.1128/JCM.44.3.1059-1064.2006.

Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V, Corander J, Cheng L, Saif S, Young S, Zeng Q, Wortman J, Birren B, Willems RJ, Earl AM, Gilmore MS. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. MBio. 2013;4(4). doi:10.1128/mBio.00534-13.

Leendertz FH, Ellerbrok H, Boesch C, Couacy-Hymann E, Matz-Rensing K, Hakenbeck R, Bergmann C, Abaza P, Junglen S, Moebius Y, Vigilant L, Formenty P, Pauli G. Anthrax kills wild chimpanzees in a tropical rainforest. Nature. 2004;430(6998):451–52. doi:10.1038/nature02722

Leendertz FH, Yumlu S, Pauli G, Boesch C, Couacy-Hymann E, Vigilant L, Junglen S, Schenk S, Ellerbrok H. A new *Bacillus anthracis* found in wild chimpanzees and a gorilla from West and Central Africa. PLoS Pathog. 2006;2(1):e8. doi:10.1371/journal.ppat.0020008.

Lester CH, Sandvang D, Olsen SS, Schonheyder HC, Jarlov JO, Bangsborg J, Hansen DS, Jensen TG, Frimodt-Moller N, Hammerum AM, DANRES Study Group. Emergence of ampicillin-resistant *Enterococcus faecium* in Danish hospitals. J Antimicrob Chemother. 2008;62(6):1203–06. doi:10.1093/jac/dkn360.

Maier T, Klepel S, Renner Z, Kostrzewa M. Fast and reliable MALDI-TOF MS-based microorganism identification. Nat Methods. 2006;3:324–34.

Makino S, Uchida I, Terakado N, Sasakawa C, Yoshikawa M. Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in *Bacillus anthracis*. J Bacteriol. 1989;171(2):722–30

MALDI BioTyper 3.0 User Manual (2012). Bruker Daltonic GmbH

Maquelin K, Choo-Smith LP, van Vreeswijk T, Endtz HP, Smith B, Bennett R, Bruining HA, Puppels GJ. Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. Anal Chem. 2000;72(1):12–9

Marques de Sa JP. Pattern recognition: concepts, methods, and applications. Springer Science & Business Media; 2001. p. 318.

Mellmann A, Weniger T, Berssenbrugge C, Rothganger J, Sammeth M, Stoye J, Harmsen D. Based upon repeat pattern (BURP): an algorithm to characterize the long-term evolution of

Staphylococcus aureus populations based on spa polymorphisms. BMC Microbiol. 2007;7:98. doi:10.1186/1471-2180-7-98.

Mikalsen T, Pedersen T, Willems R, Coque TM, Werner G, Sadowy E, van Schaik W, Jensen LB, Francia MV, Sundsfjord A, Hegstad K. Investigating the mobilome in clinically important lineages of Enterococcus faecium and Enterococcus faecalis. BMC Genomics. 2015;16:282. doi:10.1186/s12864-015-1407-6.

Moeller R, Setlow P, Reitz G, Nicholson WL. Roles of small, acid-soluble spore proteins and core water content in survival of *Bacillus* subtilis spores exposed to environmental solar UV radiation. Appl Environ Microbiol. 2009;75(16):5202–8. doi:10.1128/AEM.00789-09.

Moon HW, Lee SH, Chung HS, Lee M, Lee K. Performance of the Vitek MS matrix-assisted laser desorption ionization time-of-flight mass spectrometry system for identification of Gram-positive cocci routinely isolated in clinical microbiology laboratories. J Med Microbiol. 2013;62(Pt 9):1301–6. doi:10.1099/jmm.0.062950-0.

Nakano S, Matsumura Y, Kato K, Yunoki T, Hotta G, Noguchi T, Yamamoto M, Nagao M, Ito Y, Takakura S, Ichiyama S. Differentiation of vanA-positive *Enterococcus faecium* from vanA-negative *E. faecium* by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Int J Antimicrob Agents. 2014;44(3):256–9. doi:10.1016/j.ijantimicag.2014.05.006.

Nimmo GR. USA300 abroad: global spread of a virulent strain of community-associated methicillin-resistant *Staphylococcus aureus*. Clin Microbiol Infect. 2012;18(8):725–34. doi:10.1111/j.1469-0691.2012.03822.x.

O'Hara FP, Amrine-Madsen H, Mera RM, Brown ML, Close NM, Suaya JA, Acosta CJ. Molecular characterization of *Staphylococcus aureus* in the United States 2004–2008 reveals the rapid expansion of USA300 among inpatients and outpatients. Microbial Drug Resist. 2012;18(6):555–61. doi:10.1089/mdr.2012.0056.

Okinaka R, Cloud K, Hampton O, Hoffmaster A, Hill K, Keim P, Koehler T, Lamke G, Kumano S, Manter D, Martinez Y, Ricke D, Svensson R, Jackson P. Sequence, assembly and analysis of pX01 and pX02. J Appl Microbiol. 1999;87(2):261–2

Otto M. Community-associated MRSA: what makes them special? Int J Med Microbiol. 2013;303(6–7):324–30. doi:10.1016/j.ijmm.2013.02.007.

Pineda FJ, Antoine MD, Demirev PA, Feldman AB, Jackman J, Longenecker M, Lin JS. Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. Anal Chem. 2003;75(15):3817–22

Pineda FJ, Lin JS, Fenselau C, Demirev PA. Testing the significance of microorganism identification by mass spectrometry and proteome database search. Anal Chem. 2000;72(16):3739–44.

Quintela-Baluja M, Bohme K, Fernandez-No IC, Morandi S, Alnakip ME, Caamano-Antelo S, Barros-Velazquez J, Calo-Mata P. Characterization of different food-isolated *Enterococcus* strains by MALDI-TOF mass fingerprinting. Electrophoresis. 2013;34(15):2240–50. doi:10.1002/elps.201200699.

Radnedge L, Agron PG, Hill KK, Jackson PJ, Ticknor LO, Keim P, Andersen GL. Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*. Appl Environ Microbiol. 2003;69(5):2755–64

Rasko DA, Altherr MR, Han CS, Ravel J. Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol Rev. 2005;29 (2):303–29

Renard BY, Kirchner M, Steen H, Steen JA, Hamprecht FA. NITPICK: peak identification for mass spectrometry data. BMC Bioinformatics. 2008;9:355. doi:10.1186/1471-2105-9-355.

Robert J, Tristan A, Cavalie L, Decousser JW, Bes M, Etienne J, Laurent F, Onerba. Panton-valentine leukocidin-positive and toxic shock syndrome toxin 1-positive methicillin-resistant *Staphylococcus aureus*: a French multicenter prospective study in 2008. Antimicrob Agents Chemother. 2011;55(4):1734–9. doi:10.1128/AAC.01221-10.

Rösch P, Harz M, Schmitt M, Peschke KD, Ronneberger O, Burkhardt H, Motzkus HW, Lankers M, Hofer S, Thiele H, Popp J. Chemotaxonomic identification of single bacteria by micro-Raman spectroscopy: application to clean-room-relevant biological contaminations. Appl Environ Microbiol. 2005;71(3):1626–37

Ruiz-Garbajosa P, Bonten MJ, Robinson DA, Top J, Nallapareddy SR, Torres C, Coque TM, Canton R, Baquero F, Murray BE, del Campo R, Willems RJ. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. J Clin Microbiol. 2006;44(6):2220–28. doi:10.1128/JCM.02596-05.

Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal Chem. 2001;73(4):746–50.

Ryzhov V, Hathout Y, Fenselau C. Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl Environ Microbiol. 2000;66(9):3828–34.

Sandrin TR, Goldstein JE, Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom Rev. 2013;32(3):188–217. doi:10.1002/mas.21359.

Sauer S, Freiwald A, Maier T, Kube M, Reinhardt R, Kostrzewa M, Geider K. Classification and identification of bacteria by mass spectrometry and computational analysis. PloS ONE. 2008;3(7):e2843. doi:10.1371/journal.pone.0002843.

Sauer S, Kliem M. Mass spectrometry tools for the classification and identification of bacteria. Nat Rev Microbiol. 2010;8(1):74–82. doi:10.1038/nrmicro2243.

Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 1964;36(8):1627–39.

Schmitt J, Udelhoven T, Löchte T, Flemming HC, Naumann D. Artificial neural network applied to FTIR and FT-Raman spectra in biomedical applications. AIP Conf Proc. 1998;430:260–63

Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol Mol Biol Rev. 1998;62(3):775–806

Schulthess B, Brodner K, Bloemberg GV, Zbinden R, Bottger EC, Hombach M. Identification of Gram-positive cocci by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry: comparison of different preparation methods and implementation of a practical algorithm for routine diagnostics. J Clin Microbiol. 2013;51(6):1834–40. doi:10.1128/JCM.02654-12.

Sedo O, Vavrova A, Vad'urova M, Tvrzova L, Zdrahal Z. The influence of growth conditions on strain differentiation within the *Lactobacillus acidophilus* group using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling. Rapid Commun Mass Spectrom. 2013;27(24):2729–36. doi:10.1002/rcm.6741.

Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49(4):543–51.

Setlow P. Small, acid-soluble spore proteins of *Bacillus* species: structure, synthesis, genetics, function, and degradation. Ann Rev Microbiol. 1988;42:319–38. doi:10.1146/annurev.mi.42.100188.001535.

Shin H, Sampat MP, Koomen JM, Markey MK. Wavelet-based adaptive denoising and baseline correction for MALDI TOF MS. Omics. 2010;14(3):283–95. doi:10.1089/omi.2009.0119.

Strommenger B, Braulke C, Heuck D, Schmidt C, Pasemann B, Nübel U, Witte W. spa Typing of *Staphylococcus aureus* as a frontline tool in epidemiological typing. J Clin Microbiol. 2008;46(2):574–81

Szabados F, Becker K, von Eiff C, Kaase M, Gatermann S. The matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS)-based protein peaks of 4448 and 5302 Da are not associated with the presence of Panton-Valentine leukocidin. Int J Med Microbiol. 2011;301(1):58–63. doi:10.1016/j.ijmm.2010.05.005.

Szabados F, Kaase M, Anders A, Gatermann SG. Identical MALDI TOF MS-derived peak profiles in a pair of isogenic SCCmec-harboring and SCCmec-lacking strains of *Staphylococcus aureus*. J Infect. 2012;65(5):400–05. doi:10.1016/j.jinf.2012.06.010.

Tedim AP, Ruiz-Garbajosa P, Corander J, Rodriguez CM, Canton R, Willems R, Baquero F, Coque TM. Population biology of *Enterococcus* from intestinal colonization in hospitalized and non-hospitalized individuals in different age groups. Appl Environ Microbiol. 2014. doi:10.1128/AEM.03661-14

Top J, Willems R, Bonten M. Emergence of CC17 *Enterococcus faecium*: from commensal to hospital-adapted pathogen. FEMS Immunol Med Microbiol. 2008;52(3):297–308. doi:10.1111/j.1574-695X.2008.00383.x.

Udelhoven T, Naumann D, Schmitt J. Development of a hierarchical classification system with artificial neural networks and FT-IR spectra for the identification of bacteria. Appl Spectrosc. 2000;54(10):1471–9

Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. Appl Environ Microbiol. 2005;71(1):58–64

van Schaik W, Top J, Riley DR, Boekhorst J, Vrijenhoek JE, Schapendonk CM, Hendrickx AP, Nijman IJ, Bonten MJ, Tettelin H, Willems RJ. Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. BMC Genomics. 2010;11:239. doi:10.1186/1471-2164-11-239.

van Schaik W, Willems RJ. Genome-based insights into the evolution of enterococci. Clin Microbiol Infect. 2010;16(6):527–32. doi:10.1111/j.1198-743X.2010.03201.x.

Vilas-Bôas GT, Peruca AP, Arantes OM. Biology and taxonomy of Bacillus cereus, *Bacillus anthracis*, and *Bacillus thuringiensis*. Can J Microbiol. 2007;53(6):673–87

VITEK® MS User Manual (2011). bioMérieux SA.

Wang YR, Chen Q, Cui SH, Li FQ. Characterization of *Staphylococcus aureus* isolated from clinical specimens by matrix assisted laser desorption/ionization time-of-flight mass spectrometry. Biomed Environ Sci. 2013;26(6):430–6. doi:10.3967/0895-3988.2013.06.003.

Weisser M, Oostdijk EA, Willems RJ, Bonten MJ, Frei R, Elzi L, Halter J, Widmer AF, Top J. Dynamics of ampicillin-resistant *Enterococcus faecium* clones colonizing hospitalized patients: data from a prospective observational study. BMC Infect Dis. 2012;12:68. doi:10.1186/1471-2334-12-68.

Wenning M, Breitenwieser F, Konrad R, Huber I, Busch U, Scherer S. Identification and differentiation of food-related bacteria: a comparison of FTIR spectroscopy and MALDI-TOF mass spectrometry. J Microbiol Methods. 2014. doi:10.1016/j.mimet.2014.05.011

Werner G, Fleige C, Ewert B, Laverde-Gomez JA, Klare I, Witte W. High-level ciprofloxacin resistance among hospital-adapted *Enterococcus faecium* (CC17). Int J Antimicrob Agents. 2010;35(2):119–25. doi:10.1016/j.ijantimicag.2009.10.012.

Werner G, Freitas AR, Coque TM, Sollid JE, Lester C, Hammerum AM, Garcia-Migura L, Jensen LB, Francia MV, Witte W, Willems RJ, Sundsfjord A. Host range of enterococcal *vanA* plasmids among Gram-positive intestinal bacteria. J Antimicrob Chemother. 2011;66(2):273–82. doi:10.1093/jac/dkq455.

Wieme AD, Spitaels F, Aerts M, De Bruyne K, Van Landschoot A, Vandamme P. The effects of the growth medium on matrix-assisted laser desorption/ionization time-of-flight mass spectra: a case study of acetic acid bacteria. Appl Environ Microbiol. 2013. doi:10.1128/AEM.03708-13.

Willems RJ, Top J, van Schaik W, Leavis H, Bonten M, Siren J, Hanage WP, Corander J. Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. MBio. 2012;3(4):e00151–00112. doi:10.1128/mBio.00151-12.

Willems RJ, van Schaik W. Transition of *Enterococcus faecium* from commensal organism to nosocomial pathogen. Future Microbiol. 2009;4(9):1125–35. doi:10.2217/fmb.09.82.

Williams B, Cornett S, Crecelius A, Caprioli R, Dawant B, Bodenheimer B. An algorithm for baseline correction of MALDI mass spectra. Proceedings of the 43rd annual Southeast regional conference New York, NY, USA 1:137–142; 2005.

Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, Aepfelbacher M, Christner M. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. Int J Med Microbiol. 2011;301(1):64–68. doi:10.1016/j.ijmm.2010.06.002.

Yang C, He Z, Yu W. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. BMC Bioinformatics. 2009;10:4. doi:10.1186/1471-2105-10-4.

Yu W, Wu B, Lin N, Stone K, Williams K, Zhao H. Detecting and aligning peaks in mass spectrometry data with applications to MALDI. Comput Biol Chem. 2006;30(1):27–38. doi:10.1016/j.compbiolchem.2005.10.006.

Zhang X, Paganelli FL, Bierschenk D, Kuipers A, Bonten MJ, Willems RJ, van Schaik W. Genome-wide identification of ampicillin resistance determinants in *Enterococcus faecium*. PLoS Genetics. 2012;8(6):e1002804. doi:10.1371/journal.pgen.1002804.

Zhang X, Top J, de Been M, Bierschenk D, Rogers M, Leendertse M, Bonten MJ, van der Poll T, Willems RJ, van Schaik W. Identification of a genetic determinant in clinical *Enterococcus faecium* strains that contributes to intestinal colonization during antibiotic treatment. J Infect Dis. 2013;207(11):1780–6. doi:10.1093/infdis/jit076.

Zischka M, Künne CT, Blom J, Wobser D, Sakınç T, Schmidt-Hohagen K, Dabrowski PW, Nitsche A, Hübner J, Hain T, Chakraborty T, Linke B, Goesmann A, Voget S, Daniel R, Schomburg D, Hauck R, Hafez HM, Tielen P, Jahn D, Solheim M, Sadowy E, Larsen J, Jensen LB, Ruiz-Garbajosa P, Quiñones Pérez D, Mikalsen T, Bender J, Steglich M, Nübel U, Witte W, Werner G. Comprehensive molecular, genomic and phenotypic analysis of a major clone of Enterococcus faecalis MLST ST40. BMC Genomics. 2015; 16(1):175. doi:10.1186/s12864-015-1367-x.

Chapter 9
MALDI-TOF MS as a Novel Tool for Dereplication and Characterization of Microbiota in Bacterial Diversity Studies

Freek Spitaels, Anneleen D. Wieme and Peter Vandamme

An Introduction to Dereplication in Bacterial Diversity Studies

Determining the number and identity of all cultivable species of an ecosystem and, if possible, distinguishing among different strains of each of these species is a major challenge in bacterial diversity studies. A main reason for this is that there are no comprehensive databases that allow rapid, affordable, and accurate species-level identification of large numbers of isolates, irrespective of their origin. This is even more true for the recognition of different strains within a species. Publicly available 16S rRNA gene sequence databases cover (virtually) the entire known bacterial species diversity, but this gene lacks resolution to allow differentiation between closely related species, let alone strains belonging to the same species (Vandamme et al. 1996). Accurate species-level identification of numerous isolates obtained in the frame of biodiversity studies is therefore either not attempted, or it is based on polyphasic approaches in which methods with different taxonomic resolution, cost, and workload are applied sequentially to examine different subsets of isolates in a stepwise manner. Very often such a procedure involves first a "dereplication" step in which all isolates are examined to group those that represent the same taxon in a rapid and cost-effective manner. Dereplication is thus the assessment of novelty and aims to reduce a (large) number of isolates to a smaller, nonredundant set for further, more labor-intensive identification. It involves a rapid and affordable screening of all isolates to recognize and eliminate those that represent the same taxon.

P. Vandamme (✉) · F. Spitaels · A. D. Wieme
Laboratory of Microbiology, Faculty of Sciences, Ghent University, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium
e-mail: Peter.Vandamme@UGent.be

A. D. Wieme
Laboratory of Biochemistry and Brewing, Faculty of Bioscience Engineering, Ghent University, Valentin Vaerwyckweg 1, 9000 Ghent, Belgium

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_9

In a second step, representative isolates of these unidentified taxa are selected for further identification. The latter very often consists of determining the 16S rRNA gene sequence of these representative isolates to generate a tentative identification or to reveal the genera or species clusters the unknowns belong to. Especially the universal polymerase chain reaction (PCR) primers and the near-complete public database render this "16S rRNA approach" very appealing and very commonly more accurate identification is not required. However, 16S rRNA gene sequence analysis may be complemented by the application of validated methods that can be used for accurate species-level identification of the bacteria concerned (Prakash et al. 2007; Russo et al. 2010; Vandamme et al. 1996; Vandamme and Peeters 2014). For instance, until recently, many studies of fermented food ecosystems such as cocoa, sourdough, and others applied repetitive element (rep) primed PCR to group all isolates in a first step after which representative strains were selected and analyzed by partial 16S rRNA gene sequence analysis to obtain a tentative identification; members of the lactic acid bacteria (LAB) were subsequently accurately identified to the species level through sequence analysis of protein-encoding genes such as *pheS, rpoA,* and *atpA* using LAB-specific PCR primers. The latter was feasible and scientifically correct, because earlier taxonomic studies had generated PCR primers and *pheS, rpoA,* and *atpA* sequences of taxonomic reference strains of all LAB species and demonstrated that comparative sequencing of these housekeeping genes indeed allowed accurate species-level identification of LAB isolates, whereas comparative 16S rRNA gene sequence analysis did not (De Bruyne et al. 2007, 2008; Naser et al. 2005, 2007). The identification results of these representative strains are then extrapolated to all members of the same taxon (Doan et al. 2012; Papalexandratou et al. 2011; Scheirlinck et al. 2008).

The differentiation of genetically different strains belonging to the same species is mostly referred to as "typing" and is commonly done in medical microbiology, for instance, to reveal the epidemiology of outbreak strains, or in food microbiology in the frame of source tracking. Ideally, the dereplication method applied in diversity studies is universally applicable so that prior knowledge of the microorganisms present in a sample is not needed, and it has a resolution that allows to distinguish species and strains simultaneously. This ideal method has not yet been discovered. Several dereplication techniques have been described and include cellular fatty acid methyl ester analysis and $(GTG)_5$-PCR fingerprinting (Coorevits et al. 2008; De Clerck and De Vos 2002; De Vuyst et al. 2008; Faimali et al. 2010; Gevers et al. 2001; Ishii and Sadowsky 2009; Vandecandelaere et al. 2010; Versalovic et al. 1994). After identification of representative strains of the clusters delineated by dereplication, a cumulative database can be constructed which will allow to compare profiles generated in subsequent dereplication studies with those delineated and identified earlier. As a result, a growing dereplication database will increasingly allow identification as well, and further polyphasic identification efforts can be limited to those representative strains with patterns not observed earlier.

The Use of MALDI-TOF MS for Dereplication in Environmental Microbiology

To understand the taxonomic resolution of a dereplication method used is key to knowing how useful it is in diversity studies. For more than a decade, $(GTG)_5$-PCR fingerprinting has been used as a dereplication standard for many bacteria. This method can be used for both species- and strain-level differentiation, but it does not consistently differentiate all strains in a species (De Vuyst et al. 2008; Švec et al. 2005). It also suffers from limited long-term reproducibility of PCR fingerprints and has a limited throughput capacity (Gevers et al. 2001; Ghyselinck et al. 2011). The introduction of matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) in diagnostic microbiology provided a technically appealing alternative. Indeed, sample preparation can be as simple as the deposition of (inactivated) biological material on a ground steel plate and the addition of a suitable matrix. Furthermore, universal sample preparation methods preclude the need for prior knowledge of the identity of the isolates studied and allow analysis of both bacteria and yeasts. Moreover, provided sample preparation is properly standardized, the spectra are highly reproducible (Freiwald and Sauer 2009) and enable the creation of a mass spectral database that can be shared between users. The consumable cost is low, the generation of the mass spectra is fast, and the preparation and deposition of the samples can be automated (Cherkaoui et al. 2010; Seng et al. 2009)—all of these contribute to an impressive and affordable throughput capacity.

Today, MALDI-TOF MS has primarily been used as a dereplication tool aiming at grouping isolates at the species level. Similarity values used for delineating groups of spectra are mostly generated using algorithms based on the presence or absence of peaks (e.g., the Jaccard or Dice coefficient), or taking peak intensity into account (e.g., Pearson product-moment correlation coefficient). Alternatively, spectra can be analyzed using a curve-based algorithm without prior peak picking, using the Pearson product-moment correlation coefficient. Finally, more advanced similarity coefficients that yield normalized similarity values can be used, such as correlation coefficient index analysis (Koubek et al. 2012). The delineation of individual clusters can be based on visual inspection of the tree topology, preset cluster cutoff values, or calculated cluster cutoff values. These cluster cutoff values are similarity levels that are used as thresholds below which the spectra are regarded as belonging to different clusters, and therefore they are expected to represent different species. Preset cluster cutoff values are mostly derived from preliminary experiments and are used as a rule of thumb for the cluster delineation. Follow-up experiments performed on representative isolates of each cluster must yield the species-level identification of all isolates (see above) and will intrinsically confirm or contradict the validity of the cutoff values used. Calculated cluster cutoff values are mostly obtained through the analysis of a limited set of technical and biological replicates, and through the analysis of taxonomically well-characterized reference strains of some species. The latter will allow to quantitate the variance of similarity between spectra of different strains of those species. The degree of variance is however often species dependent and the delineation of clusters is very often based on results of preceding experiments, trial and error, and personal expertise (see below).

MALDI-TOF MS was first used as a dereplication tool in a study of bacterial isolates from marine sponges, in which mass spectra that shared five of the most intense peaks were grouped together (Dieckmann et al. 2005). Subsequent analyses of the isolates by partial 16S rRNA gene sequence analysis revealed a congruence between the obtained MALDI-TOF MS groups and the tentative identifications based on partial 16S rRNA gene sequences (Dieckmann et al. 2005). Ichiki et al. (2008) used MALDI-TOF MS to delineate groups of bacteria that were able to degrade alkylphenol polyethoxylate. The isolates were studied using PCR-restriction fragment length polymorphism (RFLP), MALDI-TOF MS, and *gyrB* sequence analysis. The latter two methods yielded the same results and grouped isolates on the basis of their ability to degrade this compound (Ichiki et al. 2008). MALDI-TOF MS was also used for the dereplication of halophilic prokaryotes from solar saltern sediments (Munoz et al. 2011). In this study, the mass spectra were clustered and divided into groups delineated using the peak-based Jaccard algorithm (Munoz et al. 2011). Some representative strains were selected for 16S rRNA gene sequence analysis and DNA–DNA hybridization experiments which demonstrated that the groups delineated consisted of isolates representing a single species (Munoz et al. 2011). Yet, random amplified polymorphic DNA (RAPD) typing revealed that such isolates could represent genetically distinct strains (Munoz et al. 2011). Similarly, MALDI-TOF MS was also used as a species-level dereplication tool in a study aiming to isolate soil and endophytic *Burkholderia caledonica* isolates (Verstraete et al. 2014). Field isolates were dereplicated by means of MALDI-TOF MS, after which clusters representing *B. caledonica* were identified by sequence analysis of the *recA* gene of representative isolates (Verstraete et al. 2014).

The use of a preset cutoff value was also applied to nine isolates from polychlorinated biphenyl (PCB)-contaminated sediment (Koubek et al. 2012). The MALDI-TOF mass spectra were subjected to a correlation coefficient index analysis and the cluster cutoff was adopted from previous tests and set at 0.6 (Koubek et al. 2012). Isolates were investigated using a commercial biochemical test kit, MALDI-TOF MS, and 16S rRNA gene sequence analysis, of which the latter two resulted in the same number of taxonomical units delineated in the sample analyzed (Koubek et al. 2012). Similarly, bacterial isolates were screened using MALDI-TOF MS for novel carotenoids with UVA-blue light absorbing properties (Stafsnes et al. 2013). The mass spectral clusters were delineated using a preset distance cutoff value of 500, although the authors noticed that not all species were delineated appropriately (Stafsnes et al. 2013). Cluster delineation was evaluated by comparison with the results of the pigment profiles and the inclusion of previously identified strains (Stafsnes et al. 2013). A preset distance cutoff value of 500 was used as a rule of thumb in other studies as well (Kopcakova et al. 2014). This cutoff value was, however, determined for the delineation of MALDI-TOF mass spectra of *Erwinia* species (Sauer et al. 2008), and it seems not appropriate to use the same value without reevaluation for other groups of bacteria as species-level cutoff values can vary, depending on the identity of the isolates studied (Christensen et al. 2012; Stafsnes et al. 2013; Wang et al. 2012).

In contrast, a carefully determined cluster cutoff value was used by Ghyselinck et al. (2011), who were the first to report a detailed comparison of the taxonomic resolution of MALDI-TOF MS and $(GTG)_5$-PCR fingerprinting in a study of potato rhizosphere isolates. They aimed to distinguish between strains rather than species and based the cluster delineation upon the reproducibility of both techniques by analyzing some of the isolates in triplicate (Ghyselinck et al. 2011). The technical variance of the triplicate measurements was then used for the calculation of the most appropriate cutoff level (Ghyselinck et al. 2011).

The Use of MALDI-TOF MS for Dereplication in Food Microbiology

In food microbiology, dereplication of both bacterial and yeast isolates has been reported but the taxonomic resolution aimed at was primarily species-level differentiation. Doan et al. (2012) compared $(GTG)_5$-PCR and MALDI-TOF MS for the dereplication of LAB isolates from Vietnamese fermented meat and used *pheS* sequence analysis for accurate species-level identification of representative strains. Although the dendrograms based on the $(GTG)_5$-PCR and MALDI-TOF MS fingerprints were not identical, both methods yielded the same species-level identification for all 119 isolates studied. These authors further confirmed the potential of MALDI-TOF MS as a dereplication tool by analyzing the LAB diversity of a fermented mustard sample and by again using *pheS* sequence analysis as an identification method (Doan et al. 2012). They also analyzed the LAB diversity of Vietnamese fermented mustard, beet, and eggplant using the same approach (Nguyen et al. 2013). Similarly, MALDI-TOF MS was applied for the dereplication of 348 beer spoilage isolates obtained from multiple types of beer cultivated onto different growth media and in different growth conditions (Wieme et al. 2014b). All isolates were subsequently identified at the species level using sequence analysis of various protein-encoding genes.

Spitaels and coworkers used MALDI-TOF MS as a dereplication tool for yeast and bacterial isolates in their studies of traditional lambic beer, an acidic beverage which is the result of a spontaneous fermentation that lasts for up to 3 years (Spitaels et al. 2014c). A total of over 2000 bacterial and yeast isolates, taken from consecutive samples during the first 2 years of fermentation, were dereplicated using a cumulative, purpose-built database. MALDI-TOF MS clusters were delineated visually, and representative isolates of each cluster were identified using sequence analysis of 16S rRNA (bacteria), 26S rRNA (yeasts), and/or housekeeping genes (both groups). The combination of MALDI-TOF MS-based dereplication coupled to sequence analysis-based identification was paramount for the thorough characterization of the microbial community during this long fermentation process. The same approach was applied on industrially produced lambic beer fermentations of which over 1300 bacterial and yeast isolates were dereplicated and/or identified using MALDI-TOF MS (Spitaels et al. 2015). The resulting mass spectra were used for the construction of a mass spectral database that is now used for the identifica-

tion of numerous bacterial and yeast isolates from other traditional acidic Belgian beers (unpublished data). The long-term application of this method demonstrated that isolates with highly similar mass spectra and therefore grouping in one cluster consistently represented the same species, but it also revealed that very distinct clusters may also represent the same species. The latter was mostly due to genuine differences between mass spectra and therefore revealed differences between strains of the same species, but it was occasionally caused by varying quality of the spectra.

MALDI-TOF MS Can Combine Dereplication and Identification

We mentioned earlier that the use of MALDI-TOF MS as a dereplication tool in biodiversity studies coupled to accurate species-level identification by means of validated taxonomic methods increasingly facilitates direct identification through MALDI-TOF MS and therefore increasingly precludes the need for further polyphasic identification. This is especially true when using commercial MALDI-TOF MS instruments and databases to which the mass spectra of unknowns are matched (Clark et al. 2013; Croxatto et al. 2012). Most of the recent biodiversity studies have been performed using instruments with commercial identification databases that can be extended in-house with mass spectra of organisms that are not in the database because they lack clinical or pharmaceutical interest (Calderaro et al. 2013, 2014; Christensen et al. 2012; Edouard et al. 2012; Ferreira et al. 2011; Kopcakova et al. 2014; Vavrova et al. 2014). MALDI-TOF MS therefore has the technical capacity to simplify dereplication and identification dramatically (Bille et al. 2012; Seng et al. 2009), and recent biodiversity studies that use MALDI-TOF MS are increasingly skipping the dereplication step (Dec et al. 2014; Egert et al. 2014; Ferreira et al. 2011; Uhlik et al. 2011; Vavrova et al. 2014). In such studies, mass spectra of isolates that cannot be identified directly can be clustered and representative isolates can be chosen for further polyphasic identification (Dubourg et al. 2014; Kopcakova et al. 2014; Plenz et al. 2014). With this approach, the cost of additional sequencing is minimized, and unreliable identifications caused by poor-quality mass spectra can be detected. This way, the mass spectral database can be further completed and application-specific databases can be constructed (Campos et al. 2010; Plenz et al. 2014; Wieme et al. 2014b).

To facilitate direct and accurate species-level identification, it is important to have an exhaustive reference database containing high-quality mass spectra generated from a sufficient number of reference strains of the same species, grown on multiple media and in different growth conditions (Bille et al. 2012; Seng et al. 2009). The number of shared peaks appeared to decrease with an increasing number of strains per species examined, and a growth medium-dependent species-specific core set of peaks and therefore peptides has been reported (Wieme et al. 2014a). For instance, the core set of peaks of two *Lactobacillus malefermentans* strains reported earlier (Wieme 2014) was reduced considerably through the acquisition and analysis of three additional *L. malefermentans* strains (Fig. 9.1). This confirmed that the intraspecies diversity

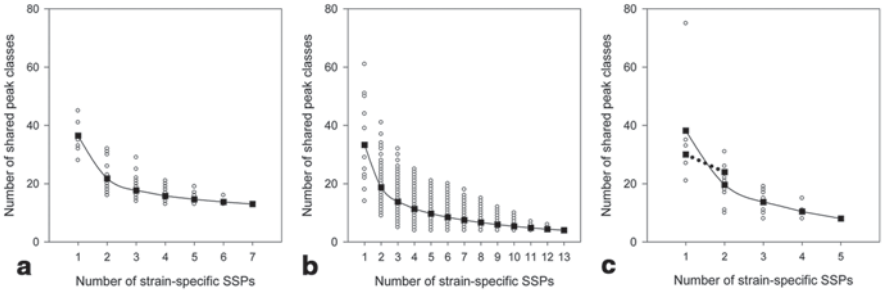


Fig. 9.1 The decrease in shared peak classes among summary spectra (SSPs; theoretical spectra that are generated by peak matching of multiple spectra per strain and that include only peaks that are present in all individual spectra, binned in peak classes) generated from 7 *L. brevis* strains grown on MRS (**a**), 13 *L. paracasei* strains grown on MRS (**b**), and 2 *(dotted curve)* and 5 *(full curve) L. malefermentans* strains grown on MRS (**c**). The number of shared peak classes is plotted as a function of the number of strain-specific SSPs sequentially added. *Squares* represent the average of shared peak classes per sequential addition of a strain-specific SSP. *MRS* de Man, Rogosa, and Sharpe medium

should be properly represented in an identification database before a reasonably accurate representation of the core peptides can be obtained (Lartigue et al. 2009). An inadequate representation of the intraspecies diversity in a MALDI-TOF MS identification database may be the cause of misidentification or lack of discriminatory power in some groups of microorganisms (Seng et al. 2009; Wieme et al. 2014a).

Detection of Taxonomic Novelty

To date, only a fraction of the bacterial diversity has been taxonomically characterized and only about 11,000 species are given formal binomial names; hence, the isolation of novel bacteria in the frame of bacterial diversity studies is fairly common (Tamames and Rosselló-Móra 2012; Vandamme and Peeters 2014; Yarza et al. 2014). The absence of matching mass spectra in a MALDI-TOF MS identification database might point to a missing organism in the database as discussed above, but might also reveal that an isolate represents a novel species. MALDI-TOF MS has aided in the detection and description of several new species in a range of genera (Kadri et al. 2014; Korczak et al. 2014; Li et al. 2014; Snauwaert et al. 2013; Spitaels et al. 2014a, b), and mass spectral data are increasingly added to novel species descriptions (Li et al. 2014; Spitaels et al. 2014a, b). Alternatively, spectra can be made available in public online repositories, such as Spectra, an initiative of the public health agency of Sweden (http://spectra.folkhalsomyndigheten.se), SpectraBank (Böhme et al. 2012), or via a private laboratory database shared online (Mishra et al. 2012; Djelouadji et al. 2012). When using commercial MALDI-TOF MS identification systems such novel species cannot be detected when solely looking at the identification scores that are produced; the raw spectral data as well as the clustering of these mass spectra should be carefully analyzed (Srinivas et al. 2014).

Furthermore, MALDI-TOF MS not only facilitates the detection of novel species but also can reveal taxonomic anomalies (Wieme et al. 2014b). The clustering of the *Pediococcus lolii* type strain among *Pediococcus acidilactici* strains suggested that it might have been misclassified (Wieme et al. 2012). The latter was indeed confirmed by using a polyphasic taxonomic analysis which demonstrated that both organisms represented the same species (Wieme et al. 2012). Similarly, Buddruhs et al. (2013) used MALDI-TOF MS in a polyphasic study to prove that strain DSM 17395 did not correspond with the type strain of *Phaeobacter gallaeciensis*.

MALDI-TOF MS As a Dereplication Tool in Culturomics

The potential of MALDI-TOF MS as a dereplication tool is maximally exploited in so-called culturomics studies. Culturomics is a term used for the exhaustive application of many culture media and growth conditions for a maximal recovery of cultivable microorganisms from a biological sample (Lagier et al. 2012). Although metagenomics (performed by either shotgun sequencing or target enrichment strategy) is a well-established cultivation-independent technique for characterizing the microbial diversity in samples (Gilbert and Dupont 2011), it is increasingly recognized that it has pitfalls too, as rare community members may remain undetected and sequence data generated prove insufficient for identification (Dubourg et al. 2013b; Lagier et al. 2012). There is a renewed interest in cultivation-dependent analyses of microbial diversity to complement other "omics" studies (Lagier et al. 2012). Culturomics approaches aimed at isolating the whole microbial community used more than 200 isolation conditions (Dubourg et al. 2013a, b, 2014; Lagier et al. 2012) and yielded several thousands of isolates. This huge number of isolates necessitated the application of a fast and cost-effective dereplication and identification technique to rapidly process these isolates in order to reduce the risk of losing part of the isolates and therefore valuable resources and information. The dereplication and subsequent identification of all isolates is a limiting factor in these studies, rather than the number of bacterial species that can be cultivated (Lagier et al. 2012). Not surprisingly, MALDI-TOF MS is currently used in such culturomics studies (Bittar et al. 2014; Dubourg et al. 2013a, b, 2014; Lagier et al. 2012). Since these studies addressed the microbial diversity of the human gut microbiome, the databases of commercial MALDI-TOF MS systems were fairly complete and many isolates were identified during dereplication (Samb-Ba et al. 2014). Unidentified isolates have been subjected to 16S rRNA gene sequence analysis, which revealed the presence of many novel species (Dubourg et al. 2013b).

The Influence of Experimental Factors on MALDI-TOF Mass Spectra Generated

Reproducibility is a major factor to consider when characterizing microorganisms using MALDI-TOF MS (Carbonnelle et al. 2011; Hinse et al. 2011; Welker 2011). Therefore, standardized growth conditions (e.g., culture medium, incubation time)

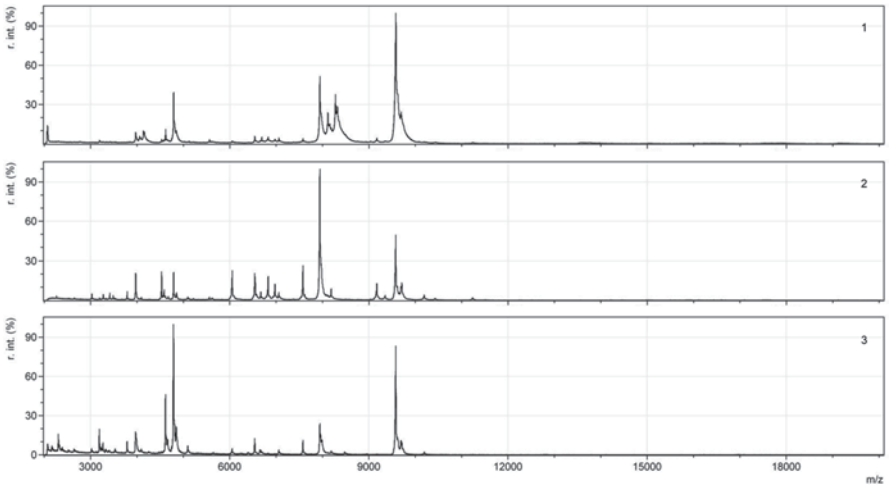


Fig. 9.2 *Pediococcus acidilactici* strain LMG 25667 grown on a commonly used growth medium for the detection of beer spoilage bacteria, known as "Nachweismedium für bierschädliche Bakterien" *(panel 1),* on de Man, Rogosa, and Sharpe medium *(panel 2),* and Raka-Ray medium *(panel 3)*. Spectra were visualized using the free mMass software. (Strohalm et al. 2010)

and optimized data acquisition for the generation of mass spectra have been recommended (Hazen et al. 2009; Sedo et al. 2011; Seibold et al. 2010; Siegrist et al. 2007; Williams et al. 2003). Yet, in biodiversity studies, multiple selective and non-selective isolation media are commonly used to maximize the isolation of different microbial populations. In order to use MALDI-TOF MS as an efficient and effective dereplication tool, the effects of the culture conditions used on the mass spectra generated should be minimized. The mass spectra of bacteria consist of signals derived mostly from ribosomal and other abundant proteins (Alispahic et al. 2010; Arnold et al. 1999; Barbuddhe et al. 2008; Dieckmann et al. 2008; Guo et al. 2009; Ryzhov and Fenselau 2001; Teramoto et al. 2007; Welker 2011). For ribosomal proteins, the effect of growth conditions on the mass spectra and therefore the identification result is expected to be minimal (Cherkaoui et al. 2010; Welker 2011). However, the growth medium used can influence the expression pattern of other proteins and hence alter the mass spectrum (Carbonnelle et al. 2011; Dieckmann et al. 2008; Giebel et al. 2010; Valentine et al. 2005; Walker et al. 2002; Welker 2011). Medium compounds can interfere with the ionization of the bacterial biomolecules, especially if the bacterial cells have the tendency to adhere onto the culture medium surface (Alispahic et al. 2010; Dieckmann et al. 2008; Giebel et al. 2010; Hettick et al. 2004; Walker et al. 2002) and culture media that do not sustain optimal growth can strongly affect the mass spectra generated (Wieme et al. 2014a). For instance, a commonly used growth medium for the detection of beer spoilage bacteria, known as the "Nachweismedium für bierschädliche Bakterien," did not support the growth of *P. acidilactici* strain LMG 25667 very well. Hence, the mass spectra generated from cells grown on this medium differed considerably from those obtained when the strain was grown on de Man, Rogosa, and Sharpe medium or Raka-Ray medium (Fig. 9.2). Besides growth conditions, other experimental factors such as the age of

the cell culture at the time of analysis, cell concentration, sample treatment, spotting method, and data acquisition all can contribute to variation in mass spectra and alter their reproducibility (Arnold et al. 1999; Balážová et al. 2014; Chean et al. 2014; Chen et al. 2008; Giebel et al. 2010; Goldstein et al. 2013; Hettick et al. 2004; Holland et al. 1996; Horneffer et al. 2004; Hsu and Burnham 2014; Karger et al. 2013; Lotz et al. 2010; Sedo et al. 2011; Toh-Boyo et al. 2012; Veloo et al. 2014; Williams et al. 2003; Usbeck et al. 2013). Also, as described earlier, mass spectra may be compared using either peak- or curve-based algorithm (Croxatto et al. 2012; Welker 2011). With a peak-based algorithm, the presence, and possibly also the intensity, of specific biomarker peaks in the unknown isolate's mass spectrum is verified, or peak lists of mass spectra of different organisms are compared (Barbuddhe et al. 2008; Böhme et al. 2011). A curve-based method considers the complete spectrum, that is, not only the presence of certain peaks but also the variation in peak signal intensity, taking into account all data points of the mass spectrum and not only the data points that describe peaks. The latter approach requires less data processing and thus allows a higher throughput, but is more prone to experimental variation (Dieckmann et al. 2005). In dereplication studies, curve-based analysis therefore introduces more variation into the clustering of isolates.

In general, however, the experimental factors used for the generation of the mass spectra do not modify the species-level identification of unknown microorganisms, irrespective of the identification algorithm used (Bille et al. 2012; Conway et al. 2001; De Bruyne et al. 2011; Dieckmann et al. 2008; Grosse-Herrenthey et al. 2008; Lartigue et al. 2009; McElvania TeKippe et al. 2013; Rezzonico et al. 2010; Seibold et al. 2010; Valentine et al. 2005; Wieme et al. 2014a). The inclusion of biological and technical replicates of strains grown at specific culture conditions reduces the biological or technically induced variations of the mass spectra generated. Commercial MALDI-TOF MS systems use robust algorithms and adequately build the proprietary database to enable a robust identification under different experimental conditions. Nonetheless, a drawback of the wide distribution and user-friendly interfaces of commercial databases is that most users report only the output identification scores of the mass spectral software, without publishing the mass spectral data or the parameters applied during mass spectral quality control, thus preventing to evaluate the quality of the spectra generated.

The culture medium used, in particular, can have a profound effect on the mass spectra generated as revealed by the presence and varying signal intensities of the peaks observed (Wieme et al. 2014a). Therefore, strain-level differentiation can be strongly affected by the growth medium selected as it may be based on minor discernible mass spectral differences (Sandrin et al. 2013; van Baar 2000; Wieme et al. 2014a). Consequently, the growth medium used is likely far more important if strain-level differentiation is aimed at, compared to species-level differentiation (Sandrin et al. 2013).

Dereplication at an Infraspecific Level?

The use of MALDI-TOF MS for the dereplication and identification of bacteria and yeasts is now well established. The resolution of MALDI-TOF MS for the identification of bacteria and yeasts is comparable or superior to the taxonomic resolution that can be obtained using rRNA gene sequence analysis (Benagli et al. 2011; Mellmann et al. 2008). Very commonly, MALDI-TOF MS will allow accurate species-level differentiation in groups of bacteria where comparative 16S rRNA gene sequence analysis lacks taxonomic resolution (Böhme et al. 2013). However, for some genera, the taxonomic resolution of MALDI-TOF MS is insufficient when mass spectra are analyzed using conventional identification algorithms only (He et al. 2010; Tan et al. 2012; Werno et al. 2012). The mass spectra of such species show too few discriminating peaks to be sufficiently differentiated when using conventional identification algorithms; yet, more rigorous peak analysis can facilitate species-level identification of some of these bacteria, as was illustrated for *Listeria* species (Barbuddhe et al. 2008).

In an ideal scenario, dereplication is not limited to the differentiation and/or identification of different species in a sample, it also allows to distinguish among strains of the same species. At present, only a limited number of studies have addressed strain-level resolution. Some of these studies aimed to distinguish among subspecies of, for example, *Bifidobacterium* or *Leuconostoc* species, but did not consistently succeed in doing so (Ruiz-Moyano et al. 2012; Pennanec et al. 2010; Zeller-Péronnet et al. 2013; De Bruyne et al. 2011), a result which may also be influenced by the data analysis algorithm used. Other studies aimed to distinguish among strains with or without certain characteristics such as antimicrobial resistance determinants (Wolters et al. 2011) or beer spoilage potential (i.e., their tolerance towards iso-alpha acids and capacity to grow in beer; Kern et al. 2014). These studies also did not prove consistently successful.

The ultimate goal however is genuine strain-level differentiation or typing, as is aimed at, for instance, when characterizing starter cultures, when source-tracking contaminants in food microbiology, or when performing epidemiological studies of outbreak strains. A growing number of studies are addressing this application of MALDI-TOF MS but commonly report an insufficient resolution. For instance, Lasch et al. (2014) could not differentiate among several clinical isolates of *Enterococcus faecium* and *Staphylococcus aureus,* whereas Schirmeister et al. (2014) could not identify reliable biomarker peaks to differentiate among isolates of *Vibrio cholerae.* Similarly, MALDI-TOF MS allowed to differentiate only a few β-lactam-resistant *Klebsiella pneumoniae* strains (Sachse et al. 2014). However, some other applications proved feasible. *Gluconobacter cerevisiae* strains isolated from different lambic breweries and from a spoiled yeast starter from a third brewery had different RAPD fingerprints and their MALDI-TOF MS spectra comprised differentiating peaks at 7491, 8102, and 10483 Da (Fig. 9.3; Spitaels et al. 2014b). Similarly, MALDI-TOF MS was recently used for quality control monitoring of a brewer's yeast: the mass spectra allowed to distinguish the brewing yeast from a nonbrewing

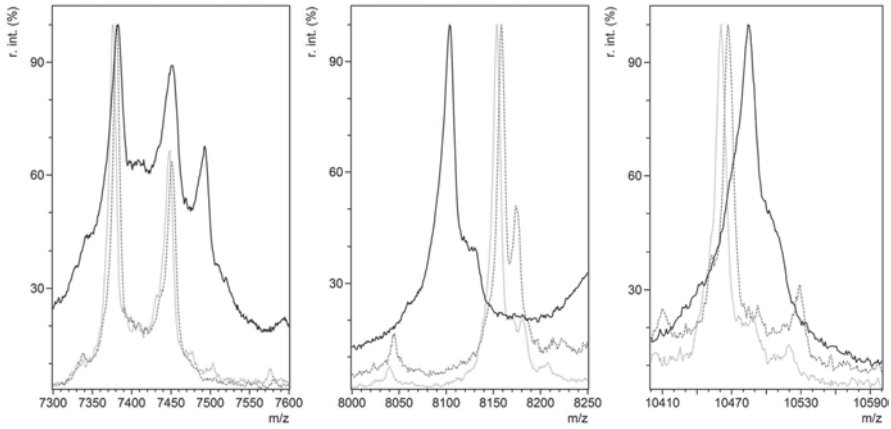


Fig. 9.3 Peaks differentiating *G. cerevisiae* LMG 27749 *(full black line)* from *G. cerevisiae* LMG 27748T *(dotted gray line)* and *G. cerevisiae* LMG 27882 *(dashed gray line):* 7491, 8102, and 10483 Da. Spectra were visualized using the free mMass software. (Strohalm et al. 2010)

wild-type yeast, although both yeasts were identified as *Saccharomyces cerevisiae* (Wieme et al. 2014c). Also, Moothoo-Padayachie et al. (2013) optimized and applied a MALDI-TOF MS protocol for the rapid biotyping of *S. cerevisiae* strains applied in different industrial fermentation backgrounds, and Usbeck et al. (2014) used MALDI-TOF MS for typing commercial yeast starters used in wine fermentation. The approaches for using MALDI-TOF MS for typing have been reviewed (Sandrin et al. 2013; Spinali et al. 2014), but strain typing using MALDI-TOF MS as the sole technique remains difficult, not the least because current algorithms are designed for the identification of microorganisms, while strain typing requires more, in depth peak-based analysis of the spectra in which the common peaks within a species can be filtered out (Sandrin et al. 2013; Spinali et al. 2014). The latter allows to focus on often subtle but reproducible differences in mass spectra of different strains (Giebel et al. 2008; Schumaker et al. 2012; Siegrist et al. 2007). Such subtle differences should represent genuine and reproducible characteristics of the mass spectra rather than experimental variation. Hence, it is plausible to assume that the use of MALDI-TOF MS for the dereplication of both species and strains in biodiversity studies will not be achieved in a single step. Strain-level identification may become feasible when isolates picked up from different cultivation media are reanalyzed after growth in standardized cultivation conditions using biological and technical replicates and the exclusion of common peaks. Indeed, the mass spectral reproducibility and quality are equally important to enable accurate and reliable infraspecific-level discrimination (Goldstein et al. 2013).

Future Perspectives

MALDI-TOF MS has become a standard method for microbial identification in clinical laboratories (Croxatto et al. 2012), while in other fields of microbiology, it is often used as a dereplication tool. Commercial systems with accompanying databases allow the direct identification of isolates during dereplication, but fail to identify many or most species in diversity studies of food or environmental samples. Commercial databases will therefore have to be expanded with appropriate numbers of reference strains to facilitate direct identification of isolates during dereplication studies of such samples as well. Although the advantages of using MALDI-TOF MS in other microbiology disciplines are obvious, the initial investment cost for both instrument and database represents a serious hurdle towards its wider application. To maximize the exploitation of the throughput capacity of MALDI-TOF MS in biodiversity studies, especially those using a culturomics approach, automation of colony picking, direct smearing or extract preparation, and subsequent MALDI-TOF MS analysis will be needed. The increased use of MALDI-TOF MS in biodiversity studies will lead to more comprehensive databases, which could also benefit from the inclusion of MALDI-TOF MS data and/or description of specific biomarkers in the taxonomic descriptions of novel species (Tindall et al. 2010).

Dereplication of microbial communities to the strain level requires further studies but may become a two-step procedure that involves the recultivation of all isolates that must be compared in identical cultivation conditions to be able to evaluate the value of individual peaks as strain markers. There is a need for flexible algorithms that allow to filter out common peaks and focus on differential peaks in the mass spectra.

References

Alispahic M, Hummel K, Jandreski-Cvetkovic D, Nöbauer K, Razzazi-Fazeli E, Hess M, Hess C. Species-specific identification and differentiation of *Arcobacter, Helicobacter* and *Campylobacter* by full-spectral matrix-associated laser desorption/ionization time of flight mass spectrometry analysis. J Med Microbiol. 2010;59(3):295–301. doi:10.1099/jmm.0.016576-0.

Arnold RJ, Karty JA, Ellington AD, Reilly JP. Monitoring the growth of a bacteria culture by MALDI-MS of whole cells. Anal Chem. 1999;71(10):1990–6.

Balážová T, Makovcová J, Šedo O, Slaný M, Faldyna M, Zdráhal Z. The influence of culture conditions on the identification of *Mycobacterium* species by MALDI-TOF MS profiling. FEMS Microbiol Lett. 2014;353(1):77–84. doi:10.1111/1574-6968.12408.

Barbuddhe SB, Maier T, Schwarz G, Kostrzewa M, Hof H, Domann E, Chakraborty T, Hain T. Rapid identification and typing of *Listeria* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2008;74(17):5402–7. doi:10.1128/AEM.02689-07.

Benagli C, Rossi V, Dolina M, Tonolla M, Petrini O. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the identification of clinically relevant bacteria. PLoS ONE. 2011;6(1):e16424. doi:10.1371/journal.pone.0016424.

Bille E, Dauphin B, Leto J, Bougnoux ME, Beretti JL, Lotz A, Suarez S, Meyer J, Join-Lambert O, Descamps P, Grall N, Mory F, Dubreuil L, Berche P, Nassif X, Ferroni A. MALDI-TOF MS

Andromas strategy for the routine identification of bacteria, mycobacteria, yeasts, *Aspergillus* spp. and positive blood cultures. Clin Microbiol Infect. 2012;18(11):1117–25. doi:10.1111/j.1469-0691.2011.03688.x.

Bittar F, Keita MB, Lagier J-C, Peeters M, Delaporte E, Raoult D. Gorilla gorilla gorilla gut: a potential reservoir of pathogenic bacteria as revealed using culturomics and molecular tools. Sci Rep. 2014;4. doi:10.1038/srep07174.

Böhme K, Fernández-No I, Gallardo J, Cañas B, Calo-Mata P. Safety assessment of fresh and processed seafood products by MALDI-TOF mass fingerprinting. Food Bioprocess Technol. 2011;4(6):907–18. doi:10.1007/s11947-010-0441-2.

Böhme K, Fernández-No IC, Barros-Velázquez J, Gallardo JM, Cañas B, Calo-Mata P. SpectraBank: an open access tool for rapid microbial identification by MALDI-TOF MS fingerprinting. Electrophoresis. 2012;33(14):2138–42. doi:10.1002/elps.201200074.

Böhme K, Fernández-No IC, Pazos M, Gallardo JM, Barros-Velázquez J, Cañas B, Calo-Mata P. Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing versus MALDI-TOF MS fingerprinting. Electrophoresis. 2013;34(6):877–87. doi:10.1002/elps.201200532.

Buddruhs N, Pradella S, Göker M, Päuker O, Pukall R, Spröer C, Schumann P, Petersen J, Brinkhoff T. Molecular and phenotypic analyses reveal the non-identity of the *Phaeobacter gallaeciensis* type strain deposits CIP 105210^T and DSM 17395. Int J Syst Evol Microbiol. 2013;63(Pt 11):4340–9. doi:10.1099/ijs.0.053900-0.

Calderaro A, Piccolo G, Gorrini C, Montecchini S, Buttrini M, Rossi S, Piergianni M, De Conto F, Arcangeletti M, Chezzi C, Medici M. *Leptospira* species and serovars identified by MALDI-TOF mass spectrometry after database implementation. BMC Res Notes. 2014;7(1):330.

Calderaro A, Piccolo G, Montecchini S, Buttrini M, Gorrini C, Rossi S, Arcangeletti MC, De Conto F, Medici MC, Chezzi C. MALDI-TOF MS analysis of human and animal *Brachyspira* species and benefits of database extension. J Proteomics. 2013;78(0):273–80. doi:http://dx.doi.org/10.1016/j.jprot.2012.09.027.

Campos VL, Valenzuela C, Yarza P, Kämpfer P, Vidal R, Zaror C, Mondaca M-A, Lopez-Lopez A, Rosselló-Móra R. *Pseudomonas arsenicoxydans* sp nov., an arsenite-oxidizing strain isolated from the Atacama desert. Syst Appl Microbiol 2010;33(4):193–7. doi:http://dx.doi.org/10.1016/j.syapm.2010.02.007.

Carbonnelle E, Mesquita C, Bille E, Day N, Dauphin B, Beretti J-L, Ferroni A, Gutmann L, Nassif X. MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. Clin Biochem. 2011;44(1):104–9. doi:http://dx.doi.org/10.1016/j.clinbiochem.2010.06.017.

Chean R, Kotsanas D, Francis MJ, Palombo EA, Jadhav SR, Awad MM, Lyras D, Korman TM, Jenkin GA. Comparing the identification of *Clostridium* spp. by two Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) mass spectrometry platforms to 16S rRNA PCR sequencing as a reference standard: A detailed analysis of age of culture and sample preparation. Anaerobe. 2014;30:85–9. doi:10.1016/j.anaerobe.2014.09.007.

Chen P, Lu Y, Harrington PB. Biomarker profiling and reproducibility study of MALDI-MS measurements of *Escherichia coli* by analysis of variance-principal component analysis. Anal Chem. 2008;80(5):1474–81. doi:10.1021/ac7018798.

Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, Francois P, Schrenzel J. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. J Clin Microbiol. 2010;48(4):1169–75. doi:10.1128/JCM.01881-09.

Christensen JJ, Dargis R, Hammer M, Justesen US, Nielsen XC, Kemp M, Group TDM-TMS. Matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis of Gram-positive, catalase-negative cocci not belonging to the *Streptococcus* or *Enterococcus* genus and benefits of database extension. J Clin Microbiol. 2012;50(5):1787–91. doi:10.1128/jcm.06339-11.

Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. 2013;26(3):547–603. doi:10.1128/CMR.00072-12.

Conway GC, Smole SC, Sarracino DA, Arbeit RD, Leopold PE. Phyloproteomics: species identification of *Enterobacteriaceae* using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Mol Microbiol Biotechnol. 2001;3(1):103–12.

Coorevits A, De Jonghe V, Vandroemme J, Reekmans R, Heyrman J, Messens W, De Vos P, Heyndrickx M. Comparative analysis of the diversity of aerobic spore-forming bacteria in raw milk from organic and conventional dairy farms. Syst Appl Microbiol. 2008;31(2):126–40. doi:10.1016/j.syapm.2008.03.002.

Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012;36(2):380–407. doi:10.1111/j.1574-6976.2011.00298.x.

De Bruyne K, Franz CM, Vancanneyt M, Schillinger U, Mozzi F, de Valdez GF, De Vuyst L, Vandamme P. *Pediococcus argentinicus* sp. nov. from Argentinean fermented wheat flour and identification of *Pediococcus* species by *pheS, rpoA* and *atpA* sequence analysis. Int J Syst Evol Microbiol. 2008;58:2909–16.

De Bruyne K, Schillinger U, Caroline L, Boehringer B, Cleenwerck I, Vancanneyt M, De Vuyst L, Franz CM, Vandamme P. *Leuconostoc holzapfelii* sp. nov., isolated from Ethiopian coffee fermentation and assessment of sequence analysis of housekeeping genes for delineation of *Leuconostoc* species. Int J Syst Evol Microbiol. 2007;57(Pt 12):2952–9. doi:10.1099/ijs.0.65292-0.

De Bruyne K, Slabbinck B, Waegeman W, Vauterin P, De Baets B, Vandamme P. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. Syst Appl Microbiol. 2011;34(1):20–9. doi:10.1016/j.syapm.2010.11.003.

De Clerck E, De Vos P. Study of the bacterial load in a gelatine production process focussed on *Bacillus* and related endosporeforming genera. Syst Appl Microbiol. 2002;25(4):611–7. doi:10.1078/07232020260517751.

De Vuyst L, Camu N, De Winter T, Vandemeulebroecke K, Van de Perre V, Vancanneyt M, De Vos P, Cleenwerck I. Validation of the $(GTG)_5$-rep-PCR fingerprinting technique for rapid classification and identification of acetic acid bacteria, with a focus on isolates from Ghanaian fermented cocoa beans. Int J Food Microbiol. 2008;125(1):79–90. doi:10.1016/j.ijfoodmicro.2007.02.030.

Dec M, Urban-Chmiel R, Gnat S, Puchalski A, Wernicki A. Identification of *Lactobacillus* strains of goose origin using MALDI-TOF mass spectrometry and 16S-23S rDNA intergenic spacer PCR analysis. Res Microbiol. 2014;165(3):190–201. doi:http://dx.doi.org/10.1016/j.resmic.2014.02.003.

Dieckmann R, Graeber I, Kaesler I, Szewzyk U, von Dohren H. Rapid screening and dereplication of bacterial isolates from marine sponges of the sula ridge by intact-cell-MALDI-TOF mass spectrometry (ICM-MS). Appl Microbiol Biotechnol. 2005;67(4):539–48. doi:10.1007/s00253-004-1812-2.

Dieckmann R, Helmuth R, Erhard M, Malorny B. Rapid classification and identification of *Salmonellae* at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2008;74(24):7767–78. doi:10.1128/Aem.01402-08.

Djelouadji Z, Roux V, Raoult D, Kodjo A, Drancourt M. Rapid MALDI-TOF mass spectrometry identification of Leptospira organisms. Vet Bicrobiol. 2012;158(1–2): 142–6. DOI: 10.1016/j.vetmic.2012.01.02.

Doan NT, Van Hoorde K, Cnockaert M, De Brandt E, Aerts M, Le Thanh B, Vandamme P. Validation of MALDI-TOF MS for rapid classification and identification of lactic acid bacteria, with a focus on isolates from traditional fermented foods in Northern Vietnam. Lett Appl Microbiol. 2012;55(4):265–73. doi:10.1111/j.1472-765X.2012.03287.x.

Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, Raoult D. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. Eur J Clin Microbiol Infect Dis. 2013a;32(5):637–45. doi:10.1007/s10096-012-1787-3.

Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, Raoult D. The proof of concept that culturomics can be superior to metagenomics to study atypical stool samples. Eur J Clin Microbiol Infect Dis. 2013b;32(8):1099. doi:10.1007/s10096-013-1843-7.

Dubourg G, Lagier JC, Robert C, Armougom F, Hugon P, Metidji S, Dione N, Dangui NPM, Pfleiderer A, Abrahao J, Musso D, Papazian L, Brouqui P, Bibi F, Yasir M, Vialettes B, Raoult D. Culturomics and pyrosequencing evidence of the reduction in gut microbiota diversity in patients with broad-spectrum antibiotics. Int J Antimicrob Agents. 2014;44(2):117–24. doi:10.1016/j.ijantimicag.2014.04.020.

Edouard S, Couderc C, Raoult D, Fournier P-E. Mass spectrometric identification of *Propionibacterium* isolates requires database enrichment. Adv Microbiol. 2012;2(4):497–504. doi:10.4236/aim.2012.24063.

Egert M, Späth K, Weik K, Kunzelmann H, Horn C, Kohl M, Blessing F. Bacteria on smartphone touchscreens in a German university setting and evaluation of two popular cleaning methods using commercially available cleaning products. Folia Microbiol (Praha). 2014;1–6. doi:10.1007/s12223-014-0350-2.

Faimali M, Chelossi E, Pavanello G, Benedetti A, Vandecandelaere I, De Vos P, Vandamme P, Mollica A. Electrochemical activity and bacterial diversity of natural marine biofilm in laboratory closed-systems. Bioelectrochemistry. 2010;78(1):30–8. doi:http://dx.doi.org/10.1016/j.bioelechem.2009.04.012.

Ferreira L, Sanchez-Juanes F, Garcia-Fraile P, Rivas R, Mateos PF, Martinez-Molina E, Gonzalez-Buitrago JM, Velazquez E. MALDI-TOF mass spectrometry is a fast and reliable platform for identification and ecological studies of species from family *Rhizobiaceae*. PLoS ONE. 2011;6(5):e20223. doi:10.1371/journal.pone.0020223.

Freiwald A, Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry. Nat Protoc. 2009;4(5):732–42.

Gevers D, Huys G, Swings J. Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. FEMS Microbiol Lett. 2001;205(1):31–6.

Ghyselinck J, Van Hoorde K, Hoste B, Heylen K, De Vos P. Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. J Microbiol Methods. 2011;86(3):327–36. doi:10.1016/j.mimet.2011.06.004.

Giebel R, Worden C, Rust SM, Kleinheinz GT, Robbins M, Sandrin TR. Chapter 6—Microbial fingerprinting using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS): applications and challenges. In: Allen IL, Sima S, Geoffrey MG, editors. Advances in applied microbiology. vol. 71. Academic; 2010. pp 149–84. doi:http://dx.doi.org/10.1016/S0065-2164(10)71006-6.

Giebel RA, Fredenberg W, Sandrin TR. (2008) Characterization of environmental isolates of *Enterococcus* spp. by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Water Res. 42(4–5):931–40. doi:http://dx.doi.org/10.1016/j.watres.2007.09.005.

Gilbert JA, Dupont CL. Microbial metagenomics: beyond the genome. Ann Rev Mar Sci. 2011;3(1):347–71. doi:10.1146/annurev-marine-120709-142811.

Goldstein JE, Zhang L, Borror CM, Rago JV, Sandrin TR. Culture conditions and sample preparation methods affect spectrum quality and reproducibility during profiling of *Staphylococcus aureus* with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Lett Appl Microbiol. 2013;57(2):144–50. doi:10.1111/lam.12092.

Grosse-Herrenthey A, Maier T, Gessler F, Schaumann R, Bohnel H, Kostrzewa M, Kruger M. Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Anaerobe. 2008;14(4):242–9. doi:10.1016/j.anaerobe.2008.06.002.

Guo Z, Liu Y, Li S, Yang Z. Interaction of bacteria and ion-exchange particles and its potential in separation for matrix-assisted laser desorption/ionization mass spectrometric identification

of bacteria in water. Rapid Commun Mass Spectrom. 2009;23(24):3983–93. doi:10.1002/rcm.4338.

Hazen TH, Martinez RJ, Chen Y, Lafon PC, Garrett NM, Parsons MB, Bopp CA, Sullards MC, Sobecky PA. Rapid identification of *Vibrio parahaemolyticus* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2009;75(21):6745–56. doi:10.1128/aem.01171-09.

He Y, Li H, Lu X, Stratton CW, Tang Y-W. Mass spectrometry biotyper system identifies enteric bacterial pathogens directly from colonies grown on selective stool culture media. J Clin Microbiol. 2010;48(11):3888–92. doi:10.1128/JCM.01290-10.

Hettick JM, Kashon ML, Simpson JP, Siegel PD, Mazurek GH, Weissman DN. Proteomic profiling of intact mycobacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Chem. 2004;76(19):5769–76. doi:10.1021/ac049410m.

Hinse D, Vollmer T, Erhard M, Welker M, Moore ER, Kleesiek K, Dreier J. Differentiation of species of the *Streptococcus bovis*/*equinus*-complex by MALDI-TOF Mass Spectrometry in comparison to *sodA* sequence analyses. Syst Appl Microbiol. 2011;34(1):52–7. doi:10.1016/j.syapm.2010.11.010.

Holland RD, Wilkes JG, Rafii F, Sutherland JB, Persons CC, Voorhees KJ, Lay JO. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1996;10(10):1227–32. doi:10.1002/(Sici)1097-0231(19960731)10:101227::Aid-Rcm6593.0.Co;2-6.

Horneffer V, Haverkamp J, Janssen HG, ter Steeg PF, Notz R. MALDI-TOF-MS analysis of bacterial spores: wet heat-treatment as a new releasing technique for biomarkers and the influence of different experimental parameters and microbiological handling. J Am Soc Mass Spectrom. 2004;15(10):1444–54. doi:10.1016/j.jasms.2004.06.017.

Hsu YM, Burnham CA. MALDI-TOF MS identification of anaerobic bacteria: assessment of pre-analytical variables and specimen preparation techniques. Diagn Microbiol Infect Dis. 2014;79(2):144–8. doi:10.1016/j.diagmicrobio.2014.02.007.

Ichiki Y, Ishizawa N, Tamura H, Teramoto K, Sato H, Yoshikawa H. Environmental distribution and novel high-throughput screening of APEO-degrading bacteria using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-MS). J Pestic Sci. 2008;33(2):122–7.

Ishii S, Sadowsky MJ. Applications of the rep-PCR DNA fingerprinting technique to study microbial diversity, ecology and evolution. Environ Microbiol. 2009;11(4):733–40. doi:10.1111/j.1462-2920.2008.01856.x.

Kadri Z, Amar M, Ouadghiri M, Cnockaert M, Aerts M, El Farricha O, Vandamme P. *Streptococcus moroccensis* sp. nov. and *Streptococcus rifensis* sp. nov., isolated from raw camel milk. Int J Syst Evol Microbiol. 2014;64(Pt 7):2480–5. doi:10.1099/ijs.0.062893-0.

Karger A, Melzer F, Timke M, Bettin B, Kostrzewa M, Nockler K, Hohmann A, Tomaso H, Neubauer H, Al Dahouk S. Interlaboratory comparison of intact-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry results for identification and differentiation of *Brucella* spp. J Clin Microbiol. 2013;51(9):3123–6. doi:10.1128/JCM.01720-13.

Kern CC, Vogel RF, Behr J. Differentiation of *Lactobacillus brevis* strains using Matrix-Assisted-Laser-Desorption-Ionization-Time-Of-Flight Mass Spectrometry with respect to their beer spoilage potential. Food Microbiol. 2014;40(0):18–24. doi:http://dx.doi.org/10.1016/j.fm.2013.11.015.

Kopcakova A, Stramova Z, Kvasnova S, Godany A, Perhacova Z, Pristas P. Need for database extension for reliable identification of bacteria from extreme environments using MALDI-TOF mass spectrometry. Chem Pap. 2014;68(11):1435–42. doi:10.2478/s11696-014-0612-0.

Korczak B, Bisgaard M, Christensen H, Kuhnert P. *Frederiksenia canicola* gen. nov., sp. nov. isolated from dogs and human dog-bite wounds. Antonie Van Leeuwenhoek. 2014;105(4):731–41. doi:10.1007/s10482-014-0129-0.

Koubek J, Uhlik O, Jecna K, Junkova P, Vrkoslavova J, Lipov J, Kurzawova V, Macek T, Mackova M. Whole-cell MALDI-TOF: rapid screening method in environmental microbiology. Int Biodeterior Biodegrad. 2012;69:82–6. doi:10.1016/j.ibiod.2011.12.007.

Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin EV, La Scola B, Raoult D. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012;18(12):1185–93. doi:10.1111/1469-0691.12023.

Lartigue MF, Hery-Arnaud G, Haguenoer E, Domelier AS, Schmit PO, van der Mee-Marquet N, Lanotte P, Mereghetti L, Kostrzewa M, Quentin R. Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2009;47(7):2284–7. doi:10.1128/JCM.00175-09.

Lasch P, Fleige C, Stämmler M, Layer F, Nübel U, Witte W, Werner G. Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of *Enterococcus faecium* and *Staphylococcus aureus* isolates. J Microbiol Methods. 2014;100(0):58–69. doi:http://dx.doi.org/10.1016/j.mimet.2014.02.015.

Li L, Wieme A, Spitaels F, Balzarini T, Nunes OC, Manaia CM, Van Landschoot A, De Vuyst L, Cleenwerck I, Vandamme P. *Acetobacter sicerae* sp. nov., isolated from cider and kefir, and identification of species of the genus *Acetobacter* by *dnaK, groEL* and *rpoB* sequence analysis. Int J Syst Evol Microbiol. 2014;64(Pt 7):2407–15. doi:10.1099/ijs.0.058354-0.

Lotz A, Ferroni A, Beretti JL, Dauphin B, Carbonnelle E, Guet-Revillet H, Veziris N, Heym B, Jarlier V, Gaillard JL, Pierre-Audigier C, Frapy E, Berche P, Nassif X, Bille E. Rapid identification of mycobacterial whole cells in solid and liquid culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2010;48(12):4481–6. doi:10.1128/jcm.01397-10.

McElvania TeKippe E, Shuey S, Winkler DW, Butler MA, Burnham C-AD. Optimizing identification of clinically relevant Gram-positive organisms by use of the Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system. J Clin Microbiol. 2013;51(5):1421–7. doi:10.1128/jcm.02680-12.

Mellmann A, Cloud J, Maier T, Keckevoet U, Ramminger I, Iwen P, Dunn J, Hall G, Wilson D, Lasala P, Kostrzewa M, Harmsen D. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. J Clin Microbiol. 2008;46(6):1946–54. doi:10.1128/JCM.00157-08.

Mishra AK, Lagier J-C, Robert C, Raoult D, Fournier P-E. Non contiguous-finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. Stand Genom Sci. 2012;7(1):1–11. doi:10.4056/sigs.2956294.

Moothoo-Padayachie A, Kandappa HR, Krishna SBN, Maier T, Govender P. Biotyping *Saccharomyces cerevisiae* strains using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Eur Food Res Technol. 2013;236(2):351–64. doi:10.1007/s00217-012-1898-1.

Munoz R, Lopez-Lopez A, Urdiain M, Moore ER, Rossello-Mora R. Evaluation of matrix-assisted laser desorption ionization-time of flight whole cell profiles for assessing the cultivable diversity of aerobic and moderately halophilic prokaryotes thriving in solar saltern sediments. Syst Appl Microbiol. 2011;34(1):69–75. doi:10.1016/j.syapm.2010.11.012.

Naser SM, Dawyndt P, Hoste B, Gevers D, Vandemeulebroecke K, Cleenwerck I, Vancanneyt M, Swings J. (2007) Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. Int J Syst Evol Microbiol. 57(Pt 12):2777–89. doi:10.1099/ijs.0.64711-0.

Naser SM, Thompson FL, Hoste B, Gevers D, Dawyndt P, Vancanneyt M, Swings J. Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. Microbiology. 2005;151(Pt 7):2141–50. doi:10.1099/mic.0.27840-0.

Nguyen DTL, Van Hoorde K, Cnockaert M, De Brandt E, Aerts M, Binh Thanh L, Vandamme P. A description of the lactic acid bacteria microbiota associated with the production of traditional fermented vegetables in Vietnam. Int J Food Microbiol. 2013;163(1):19–27. doi:http://dx.doi.org/10.1016/j.ijfoodmicro.2013.01.024.

Papalexandratou Z, Vrancken G, De Bruyne K, Vandamme P, De Vuyst L. Spontaneous organic cocoa bean box fermentations in Brazil are characterized by a restricted species diversity of lactic acid bacteria and acetic acid bacteria. Food Microbiol. 2011;28(7):1326–38. doi:10.1016/J.Fm.2011.06.003.

Pennanec X, Dufour A, Haras D, Rehel K. A quick and easy method to identify bacteria by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun In Mass Spectrom. 2010;24(3): 384–2. doi: 10.1002/rcm.4404.

Plenz B, Schmidt V, Grosse-Herrenthey A, Krüger M, Pees M. Characterisation of the aerobic bacterial flora of boid snakes: application of MALDI-TOF mass spectrometry. Vet Rec. 2014. doi:10.1136/vr.102580.

Prakash O, Verma M, Sharma P, Kumar M, Kumari K, Singh A, Kumari H, Jit S, Gupta SK, Khanna M, Lal R. Polyphasic approach of bacterial classification—An overview of recent advances. Indian J Microbiol. 2007;47(2):98–108. doi:10.1007/s12088-007-0022-x.

Rezzonico F, Vogel G, Duffy B, Tonolla M. Application of whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification and clustering analysis of *Pantoea* species. Appl Environ Microbiol. 2010;76(13):4497–509. doi:10.1128/aem.03112-09.

Ruiz-Moyano S, Tao NN, Underwood MA, Mills DA. Rapid discrimination of Bifidobacterium animalis subspecies by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Food Microbiol. 2012;30(2):432–7. doi: 10.1016/j.fm.2011.12.012.

Russo P, Beleggia R, Ferrer S, Pardo I, Spano G. A polyphasic approach in order to identify dominant lactic acid bacteria during pasta manufacturing. LWT Food Sci Technol. 2010;43(6):982–6. doi:http://dx.doi.org/10.1016/j.lwt.2010.01.013.

Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal Chem. 2001;73(4):746–50. doi:10.1021/Ac0008791.

Sachse S, Bresan S, Erhard M, Edel B, Pfister W, Saupe A, Rödel J. Comparison of multilocus sequence typing, RAPD, and MALDI-TOF mass spectrometry for typing of β-lactam-resistant *Klebsiella pneumoniae* strains. Diagn Microbiol Infect Dis. 2014;80 (4):267–71. doi:http://dx.doi.org/10.1016/j.diagmicrobio.2014.09.005.

Samb-Ba B, Mazenot C, Gassama-Sow A, Dubourg G, Richet H, Hugon P, Lagier J-C, Raoult D, Fenollar F. MALDI-TOF identification of the human gut microbiome in people with and without diarrhea in Senegal. PLoS ONE. 2014;9(5):e87419. doi:10.1371/journal.pone.0087419.

Sandrin TR, Goldstein JE, Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom Rev. 2013;32(3):188–217. doi:10.1002/mas.21359.

Sauer S, Freiwald A, Maier T, Kube M, Reinhardt R, Kostrzewa M, Geider K. Classification and identification of bacteria by mass spectrometry and computational analysis. PLoS ONE. 2008;3(7):e2843.

Scheirlinck I, Van der Meulen R, Van Schoor A, Vancanneyt M, De Vuyst L, Vandamme P, Huys G. Taxonomic structure and stability of the bacterial community in belgian sourdough ecosystems as assessed by culture and population fingerprinting. Appl Environ Microbiol. 2008;74(8):2414–23. doi:10.1128/AEM.02771-07.

Schirmeister F, Dieckmann R, Bechlars S, Bier N, Faruque SM, Strauch E. Genetic and phenotypic analysis of *Vibrio cholerae* non-O1, non-O139 isolated from German and Austrian patients. Eur J Clin Microbiol Infect Dis. 2014;33(5):767–78. doi:10.1007/s10096-013-2011-9.

Schumaker S, Borror CM, Sandrin TR. Automating data acquisition affects mass spectrum quality and reproducibility during bacterial profiling using an intact cell sample preparation method with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2012;26(3):243–53. doi:10.1002/rcm.5309.

Sedo O, Sedlacek I, Zdrahal Z. Sample preparation methods for MALDI-MS profiling of bacteria. Mass Spectrom Rev. 2011;30(3):417–34. doi:10.1002/mas.20287.

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. 2010;48(4):1061–9. doi:10.1128/JCM.01953-09.

Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49(4):543–51. doi:10.1086/600885.

Siegrist TJ, Anderson PD, Huen WH, Kleinheinz GT, McDermott CM, Sandrin TR. Discrimination and characterization of environmental strains of *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). J Microbiol Methods. 2007;68(3):554–62. doi:10.1016/j.mimet.2006.10.012.

Snauwaert I, Papalexandratou Z, De Vuyst L, Vandamme P. Characterization of strains of *Weissella fabalis* sp. nov. and *Fructobacillus tropaeoli* from spontaneous cocoa bean fermentations. Int J Syst Evol Microbiol. 2013;63(Pt 5):1709–16. doi:10.1099/ijs.0.040311-0.

Spinali S, van Belkum A, Goering RV, Girard V, Welker M, Van Nuenen M, Pincus DH, Arsac M, Durand G. Microbial typing by MALDI-TOF MS: Do we need guidance for data interpretation? J Clin Microbiol. 2014. doi:10.1128/jcm.01635-14.

Spitaels F, Li L, Wieme A, Balzarini T, Cleenwerck I, Van Landschoot A, De Vuyst L, Vandamme P. *Acetobacter lambici* sp. nov. isolated from fermenting lambic beer. Int J Syst Evol Microbiol 2014a;64(Pt 4):1083–9. doi:10.1099/ijs.0.057315-0.

Spitaels F, Wieme A, Balzarini T, Cleenwerck I, De Vuyst L, Van Landschoot A, Vandamme P. *Gluconobacter cerevisiae* sp. nov. isolated from the brewery environment. Int J Syst Evol Microbiol 2014b;64(Pt 4):1134–41. doi:10.1099/ijs.0.059311-0.

Spitaels F, Wieme AD, Janssens M, Aerts M, Daniel HM, Van Landschoot A, De Vuyst L, Vandamme P. The microbial diversity of traditional spontaneously fermented lambic beer. PLoS ONE. 2014c;9(4):e95384. doi:10.1371/journal.pone.0095384.

Spitaels F, Wieme AD, Janssens M, Aerts M, Van Landschoot A, De Vuyst L, Vandamme P. The microbial diversity of an industrially produced lambic beer shares members of a traditionally produced one and reveals a core microbiota for lambic beer fermentation. Food Microbiol. 2015. doi:http://dx.doi.org/10.1016/j.fm.2015.01.008. (in press).

Srinivas TNR, Aditya S, Bhumika V, Kumar PA. *Lunatimonas lonarensis* gen. nov., sp. nov., a haloalkaline bacterium of the family *Cyclobacteriaceae* with nitrate reducing activity. Syst Appl Microbiol. 2014;37(1):10–6. doi:http://dx.doi.org/10.1016/j.syapm.2013.10.003.

Stafsnes MH, Dybwad M, Brunsvik A, Bruheim P. Large scale MALDI-TOF MS based taxa identification to identify novel pigment producers in a marine bacterial culture collection. Antonie Van Leeuwenhoek. 2013;103(3):603–15. doi:10.1007/s10482-012-9844-6.

Strohalm M, Kavan D, Novak P, Volny M, Havlicek V. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Anal Chem. 2010;82(11):4648–51. doi:10.1021/Ac100818g.

Švec P, Vancanneyt M, Seman M, Snauwaert C, Lefebvre K, Sedláček I, Swings J. Evaluation of $(GTG)_5$-PCR for identification of *Enterococcus* spp. FEMS Microbiol Lett. 2005;247(1):59–63. doi:10.1016/j.femsle.2005.04.030.

Tamames J, Rosselló-Móra R. On the fitness of microbial taxonomy. Trends Microbiol. 2012;20(11):514–6. doi:http://dx.doi.org/10.1016/j.tim.2012.08.012.

Tan KE, Ellis BC, Lee R, Stamper PD, Zhang SX, Carroll KC. Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: a bench-by-bench study for assessing the impact on time to identification and cost-effectiveness. J Clin Microbiol. 2012;50(10):3301–8. doi:10.1128/JCM.01405-12.

Teramoto K, Sato H, Sun L, Torimura M, Tao H. A simple intact protein analysis by MALDI-MS for characterization of ribosomal proteins of two genome-sequenced lactic acid bacteria and verification of their amino acid sequences. J Proteome Res. 2007;6(10):3899–907. doi:10.1021/pr070218 l.

Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol. 2010;60(Pt 1):249–66. doi:10.1099/ijs.0.016949-0.

Toh-Boyo GM, Wulff SS, Basile F. Comparison of sample preparation methods and evaluation of intra- and intersample reproducibility in bacteria MALDI-MS profiling. Anal Chem. 2012;84(22):9971–80. doi:10.1021/ac302375e.

Uhlik O, Strejcek M, Junkova P, Sanda M, Hroudova M, Vlcek C, Mackova M, Macek T. Matrix-assisted laser desorption ionization (MALDI)-time of flight mass spectrometry- and MALDI biotyper-based identification of cultured biphenyl-metabolizing bacteria from contaminated horseradish rhizosphere soil. Appl Environ Microbiol. 2011;77(19):6858–66. doi:10.1128/aem.05465-11.

Usbeck J, Wilde C, Bertrand D, Behr J, Vogel R. Wine yeast typing by MALDI-TOF MS. Appl Microbiol Biotechnol. 2014;98(8):3737–52. doi:10.1007/s00253-014-5586-x.

Usbeck JC, Kern CC, Vogel RF, Behr J. Optimization of experimental and modelling parameters for the differentiation of beverage spoiling yeasts by matrix-assisted-laser-desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) in response to varying growth conditions. Food Microbiol. 2013;36(2):379–87. doi:10.1016/j.fm.2013.07.004.

Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. Appl Environ Microbiol. 2005;71(1):58–64. doi:10.1128/aem.71.1.58-64.2005.

van Baar BL. Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. FEMS Microbiol Rev. 2000;24(2):193–219.

Vandamme P, Peeters C. Time to revisit polyphasic taxonomy. Antonie Van Leeuwenhoek. 2014;106(1):57–65. doi:10.1007/s10482-014-0148-x.

Vandamme P, Pot B, Gillis M, De Vos P, Kersters K, Swings J. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev. 1996;60(2):407–38.

Vandecandelaere I, Nercessian O, Faimali M, Segaert E, Mollica A, Achouak W, De Vos P, Vandamme P. Bacterial diversity of the cultivable fraction of a marine electroactive biofilm. Bioelectrochemistry. 2010;78(1):62–6. doi:10.1016/j.bioelechem.2009.07.004.

Vavrova A, Matoulkova D, Balážová T, Šedo O. MALDI-TOF MS analysis of anaerobic bacteria isolated from biofilm-covered surfaces in brewery bottling halls. J Am Soc Brew Chem. 2014;72(2):95–101. doi:10.1094/Asbcj-2014-0324-01.

Veloo AC, Elgersma PE, Friedrich AW, Nagy E, van Winkelhoff AJ. The influence of incubation time, sample preparation and exposure to oxygen on the quality of the MALDI-TOF MS spectrum of anaerobic bacteria. Clin Microbiol Infect. 2014. doi:10.1111/1469-0691.12644.

Versalovic J, Schneider M, De Bruijn FJ, Lupski JR. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol Cell Biol. 1994;5(1):25–40.

Verstraete B, Peeters C, van Wyk B, Smets E, Dessein S, Vandamme P. Intraspecific variation in *Burkholderia caledonica:* Europe vs. Africa and soil vs. endophytic isolates. Syst Appl Microbiol. 2014;37(3):194–9. doi:10.1016/j.syapm.2013.12.001.

Walker J, Fox AJ, Edwards-Jones V, Gordon DB. Intact cell mass spectrometry (ICMS) used to type methicillin-resistant *Staphylococcus aureus:* media effects and inter-laboratory reproducibility. J Microbiol Methods. 2002;48(2–3):117–26. doi:http://dx.doi.org/10.1016/S0167-7012(01)00316-5.

Wang J, Zhou N, Xu B, Hao H, Kang L, Zheng Y, Jiang Y, Jiang H. Identification and cluster analysis of *Streptococcus pyogenes* by MALDI-TOF mass spectrometry. PLoS ONE. 2012;7(11):e47152. doi:10.1371/journal.pone.0047152.

Welker M. Proteomics for routine identification of microorganisms. Proteomics. 2011;11(15):3143–53. doi:10.1002/pmic.201100049.

Werno AM, Christner M, Anderson TP, Murdoch DR. Differentiation of *Streptococcus pneumoniae* from Nonpneumococcal *Streptococci* of the *Streptococcus mitis* group by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50(9):2863–7. doi:10.1128/jcm.00508-12.

Wieme A. Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry as a novel tool for detection, identification and typing of spoilage microbiota in the brewing industry. Ghent University; 2014.

Wieme A, Cleenwerck I, Van Landschoot A, Vandamme P. *Pediococcus lolii* DSM 19927^T and JCM 15055^T are strains of *Pediococcus acidilactici*. Int J Syst Evol Microbiol. 2012;62(Pt 12):3105–8. doi:10.1099/ijs.0.046201-0.

Wieme AD, Spitaels F, Aerts M, De Bruyne K, Van Landschoot A, Vandamme P. Effects of growth medium on matrix-assisted laser desorption-ionization time of flight mass spectra: a case study of acetic Acid bacteria. Appl Environ Microbiol. 2014a;80(4):1528–38. doi:10.1128/aem.03708-13.

Wieme AD, Spitaels F, Aerts M, De Bruyne K, Van Landschoot A, Vandamme P. Identification of beer-spoilage bacteria using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Int J Food Microbiol. 2014b;185:41–50. doi:10.1016/j.ijfoodmicro.2014.05.003.

Wieme AD, Spitaels F, Vandamme P, Van Landschoot A. Application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry as a monitoring tool for in-house brewer's yeast contamination: a proof of concept. J Inst Brew. 2014c;120(4):438–43. doi:10.1002/jib.149.

Williams TL, Andrzejewski D, Lay JO, Musser SM. Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. J Am Soc Mass Spectrom. 2003;14(4):342–51. doi:10.1016/s1044-0305(03)00065-5.

Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, Aepfelbacher M, Christner M. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. Int J Med Microbiol. 2011;301(1):64–8. doi:http://dx.doi.org/10.1016/j.ijmm.2010.06.002.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol. 2014;12(9):635–45. doi:10.1038/nrmicro3330.

Zeller-Peronnet V, Brockmann E, Pavlovic M, Timke M, Busch U, Huber I. Potential and limitations of MALDI-TOF MS for discrimination within the species Leuconostoc mesenteroides and Leuconostoc pseudomesenteroides. Journal Fur Verbrauch und Lebensm-Journal of Consum Prot and Food Saf. 2013;8(3): 205–14. doi: 10.1007/s00003-013-0826-z.

Chapter 10
Bacterial Identification at the Serovar Level by Top-Down Mass Spectrometry

Melinda A. McFarland, Denis Andrzejewski and John H. Callahan

Introduction

Members of the *Salmonella enterica enterica* subspecies are the cause of most human salmonellosis and in the USA, most cases are food-borne. *S. enterica enterica* consists of more than 2500 different O and H cell surface antigen combinations, or serovars (FDA 2012). *S. enterica* serovar Typhimurium and *S. enterica* serovar Heidelberg are among the top ten serovars implicated in food-borne *Salmonella* infections (CDC 2014). Although these are distinct serovars, their genomes are 99 % similar (data not shown). Species- and subspecies-level assays are generally adequate for clinical diagnostics. However, localization of the source of a food-borne *Salmonella* contamination requires serovar or strain-level specificity.

Pulsed field gel electrophoresis (PFGE) has become the gold standard for molecular subtyping of *Salmonella,* and polymerization chain reaction (PCR)-based assays built around genomic markers are becoming increasingly popular (Wattiau et al. 2011). Differentiating between two highly similar serovars such as *S.* Typhimurium and *S.* Heidelberg requires multiple enzymes and relies on matching to a previously validated standard. Detection methods that require selection of probe-based assays, such as PCR, are limited by probe selection. Changes to untargeted genes and newly acquired genetic material are likely to be missed. More recently, approaches based on whole-genome sequencing (WGS) have been used to address strain identification (Lienau et al. 2011).

Mass spectrometry is a powerful analytical tool that can be used to probe proteins, peptides, lipids, and metabolites produced by bacteria; mass spectrometers are a ubiquitous, sensitive, specific, and inherently multiplexed platform that can potentially be used to identify and differentiate bacteria. A nontargeted mass spectrometry-based method provides a relatively unbiased snapshot of the expressed

M. A. McFarland (✉) · D. Andrzejewski · J. H. Callahan
Spectroscopy and Mass Spectrometry Branch, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, 5100 Paint Branch Pkwy., College Park, MD 20740, USA
e-mail: melinda.mcfarland@fda.hhs.gov

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_10

proteins in a wide range of bacterial samples and is amenable to both screening and targeted analysis. This facilitates differentiation of closely related bacteria, as well as the detection of un-sequenced or newly acquired non-synonymous SNPs and plasmid proteins that may be specific to a given strain.

Mass spectrometry is commonly used to identify proteins from the bottom-up, using peptides derived from enzymatic digestion of protein lysates (McCormack et al. 1997). However, the cross genome homology present in bacteria limits the feasibility of differentiation across closely related isolates by bottom-up peptide-based analysis. If an MS/MS spectrum is not generated for the SNP (henceforth, this term will be used to mean non-synonymous or non-silent SNPs)-containing peptide, the presence of that SNP will be missed. If the SNP has not been genomically sequenced or is not present in the searched database, the biomarker will also go undetected. The identification of unknown bacterial lysates lacking fully sequenced genomes may be challenging due to a bias toward those species that are most represented in the database. Consequently, there is a distinct advantage of using intact proteins to detect differences induced by non-synonymous SNPs, as the presence of such mutations would result in measurable differences in the mass of the intact protein, with no need for a sequenced genome.

Intact protein mass spectrometry of bacterial lysates provides an inherently multiplexed measurement of the mass of expressed proteins in their intact state, at a given growth stage (Krishnamurthy and Ross 1996; Fenselau and Demirev 2001; Conway et al. 2001). This is particularly useful because bacteria exhibit fewer overall post-translational modifications (PTMs) and, given a controlled growth state, minimal PTM variability as compared to mammalian systems. Bacterial proteins and their modifications are highly conserved across species. Although protein abundances may vary from serovar to serovar, their masses should be highly conserved. Therefore, for bacterial lysates it is a reasonable assumption that the minimal mass shifts found between closely related bacteria are the result of SNPs (Wilcox et al.2001; Dieckmann et al.2008; Arnold and Reilly 1999). These mass-shifted proteins serve as biomarkers for differentiation of bacteria.

Intact protein mass spectrometry has become a commercially available tool for clinical bacterial differentiation based on the matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) technology (Bizzini and Greub 2010; Clark 2013). However, a mass range generally limited to below 15 kDa and a bias toward ribosomal proteins (Ryzhov and Fenselau 2001) often limit MALDI applications to species- and subspecies-level identifications. The increased mass range, improved reproducibility, and greater number of proteins ionized using an electrospray ionization (ESI)-based platform provide access to a more diverse range of proteins and an increased specificity for differentiation of closely related bacteria (Krishnamurthy et al. 1999; Ho and Hsu 2002; Mott et al. 2010). This approach, known as intact protein chromatography electrospray mass spectrometry, has already been used to identify marker masses that differentiate thermophilic versus non-thermophilic groups of *Cronobacter sakazakii* (Williams et al. 2005) to identify proteins characteristic of specific outbreak strains of *Vibrio parahaemolyticus* (Williams et al. 2004), and to differentiate closely related species within the enterobacteriaceae family (Mott et al. 2010; Everley et al. 2008).

The addition of online "top-down" MS/MS fragmentation of the intact proteins provides identification of the proteins containing measured mass differences (Cargile et al. 2001; Lee et al. 2002; Fagerquist et al. 2006; Wynne et al. 2010; McFarland et al. 2014). By identifying which of the most highly expressed bacterial proteins are conserved and which contain amino acid differences, we can differentiate between samples, validate genomically predicted SNPs for sequenced genomes, and for un-sequenced species, determine whether a mass shift in a specific protein represents a novel, and possibly virulent, mutation. This provides a direct link back to genome-sequencing data, facilitating gene-specific marker and sequence validation at an expressed protein level.

The combination of intact protein chromatography ESI-MS with top-down mass spectrometry facilitates the identification of proteins that result from expressed serovar-specific non-synonymous SNPs. This approach is based on deconvoluted ESI-MS generated intact protein expression profiles (Williams et al. 2002) to facilitate rapid differentiation between samples, combined with top-down identification of proteins for marker confirmation. Application of this methodology as a screening method would require sequencing only expression profile masses that show a mass shift when compared to a reference strain, and such an analysis can be done without prior selection of biomarker proteins and without a sequenced genome. Knowledge of which protein sequences are variable across serovars provides a common link to genome sequencing and phylogenetic strain-typing efforts.

Methods

Bacterial Strains

Salmonella enterica enterica serovar Typhimurium strain LT2 and *S.* Heidelberg strain A39 bacterial strains used in the study were obtained from the stock culture collection of the Food and Drug Administration (FDA)/Center for Food Safety and Applied Nutrition. Bacteria were grown for 24 h at 37 °C on lysogeny broth agar plates (Teknova, Hollister, CA). For the multi-isolate study, 36 semi-blinded *Salmonella* isolates from food-borne outbreaks investigated by the FDA were cultured overnight on tryptic soy agar plates. Cell isolates were collected in a 1.5-mL sample tube and washed twice with sterile water and resuspended in 0.5-mL of 70 % ethanol to facilitate sterilization of bacteria (Williams et al. 2003) as well as minimize protease activity. The approximate cell concentration is 8×10^{10} cfu/mL.

Extraction of Cellular Proteins

The sample tube containing bacterial cells suspended in 70 % ethanol was centrifuged at 9800 x g for 5 min. The ethanol solution was removed, and 1.0 mL of a 50:49:1 extraction solution consisting of acetonitrile, high-performance liquid

chromatography (HPLC)-grade water, and formic acid was added and the tube was vortexed to resuspend the cells. The 1.0 mL suspension was transferred to a Barocycler® FT500 pulse tube (Pressure Biosciences, Inc., Boston, MA) along with an additional 0.4 mL of extraction solution and was capped. The Barocycler NEP 3229 was pressure cycled 24 times at 44 °C starting at 35,000 psi for 15 s and then at 0 psi for 10 s. The pulse tube contents were transferred to a 1.5-mL low-binding sample tube and centrifuged at 9800 x *g* for 20 min to pellet the cellular debris. A portion of the supernatant was transferred to an autosampler tube for LC-MS analysis.

HPLC of Intact Proteins

Intact proteins were separated by reverse-phase HPLC using an Agilent (Palo Alto, CA) 1100 system fitted with two ProSphere P-HR (W.R. Grace, MD) 2.1 mm i. d. × 15 cm columns connected in series. Two microliters of the protein extract were injected into the column at an oven temperature of 50 °C and a flow rate of 200 µL/min. Mobile phase A was 95% HPLC-grade water and mobile phase B was 95% acetonitrile, both with 5% acetic acid. The gradient was as follows: 0–5 min 90% A, hold for 1 min, 70 min 50% A, 80 min 10% A, 92 min 10% A, and 94 min 90% A. Identical separation methods were used in-line with both instrument platforms to retain consistent retention times across platforms. For the multi-isolate study, all conditions were the same, except proteins were separated on a Kinetex C8 (Phenomenex, Torrance, CA) 1.7 µm, 100A, 15 cm column, with mobile phase A 98% HPLC-grade water and mobile phase B 98% acetonitrile, both with 2% formic acid.

LC-MS and Data Analysis

The HPLC was interfaced to a Q-TOF Premier (Waters, Beverly, MA) mass spectrometer. The instrument was operated at 3.0 kV capillary voltage, 100 °C source temperature, 150 °C desolvation temperature, desolvation gas 600 L/h, and scanning from 550 to 2000 Da in 1.0 s in single reflectron mode. Data were collected using MassLynx software version 4.1 (Waters, Beverly, MA).

MS Data Analysis

Automated analysis of full-scan (MS) data was performed with ProTrawler6 (previously named Retana) and custom software (BioAnalyte, Inc., Portland, ME). Its function is to automatically process sequential complex, multiply charged mass

spectra obtained during ESI-LC-MS analysis and produce a text file containing the binned uncharged protein mass, retention time, and intensity of all proteins deconvoluted from the LC-MS run. A detailed explanation of the approach has been published (Williams et al. 2002). Briefly, spectra are summed in 30 s windows. In version 6 of ProTrawler the summed spectrum from each time window is baseline subtracted and de-noised using the proprietary ReSpect ™ algorithm (Positive Probability, Shrewsbury, UK). The resultant spectrum is deconvoluted using maximum entropy deconvolution. After generating a protein mass/abundance list for each time window, ProTrawler then bins the data for each time window, determines the time range over which a given mass occurs, and calculates an abundance-weighted time centroid for the mass, which is used to represent the retention time. Masses corresponding to multimers and adducts are also removed. Abundances are then normalized to the summed intensity. The resulting text file contains a cumulative list of all the intact protein masses, abundances, and retention times, of which the mass and abundance information can be represented graphically as mass versus intensity, similar to a traditional mass spectrum. The retention time is also included in the output so that proteins of similar mass can be distinguished based on the retention time.

Top-Down LC-MS/MS

Online intact protein separation was the same as for the Q-TOF LC-MS (above) for consistent protein retention times across platforms. For LC-MS/MS the eluent flow was split to a flow rate of 350 nL/min via the TriVersa NanoMate (Advion BioSciences, Ithaca, NY) chip-based nanospray source and analyzed with a LTQ-Orbitrap XL (Thermo Fisher, San Jose, CA) mass spectrometer. The instrument was operated in a top-three data dependent mode, with both MS spectra and collision-induced dissociation (CID) MS/MS spectra acquired at 60,000 resolving power in the Orbitrap. CID collision energy was operated at 15%. Each MS spectrum was composed of three microscans, and each MS/MS spectrum was the average of 10 microscans. To facilitate the analysis of intact proteins, the instrument was operated with the HCD gas off and the delay before image current detection shortened to 5 ms.

Top-Down Data Analysis

ProSightPC 2.0 (Zamdborg et al. 2007) was used to search MS/MS spectra against a protein sequence library of UniprotKB Swiss-Prot and TrEMBL protein sequence entries for the *Salmonella* Typhimurium fully sequenced strain LT2 or a custom-made *S.* Heidelberg database from fully sequenced strain SL476 (as of the time of this work a fully sequenced A39 genome was not available). Neutral mass deconvoluted precursor and fragment mass lists were generated with the Xtract algorithm

(Thermo Fisher, San Jose, CA) option within ProSightPC 2.0. The precursor mass tolerance was 1000 Da, and the fragment ion tolerance was 20 ppm for the monoisotopic mass. Only disulfide bonds were included as a modification in the primary search. PTMs were inferred from mass differences relative to the theoretical mass. Modifications were subsequently validated by manual addition of the proposed modification followed by re-assignment of fragment ions and rescoring via the sequence gazer option in ProSightPC. Modifications were considered valid if there was an increase in matched fragment ions upon inclusion of the predicted modification. A secondary search was also performed that included the most commonly inferred PTMs as confirmation of the amended modification as the top-scoring identification. Only proteins identified with ProSight *e*-values better than $1e^{-5}$ for a minimum of three MS/MS spectra were considered valid identifications.

Results and Discussion

The power of intact protein analysis is that the mass of the protein is measured with functional modifications intact. This is ideally suited for bacterial proteins because, unlike mammalian systems, bacterial lysates from similar species appear to exhibit highly reproducible and conserved PTMs under similar growth conditions. Although protein abundances may vary, there should be few differences in their masses. Therefore, for bacterial lysates grown under the same conditions, it is reasonable to assume that a small number of mass shifts found across serovars are SNPs, and novel masses are insertions or proteins that have undergone a significant change in the expression level. These mass-shifted proteins serve as markers for differentiation of bacteria at the species, subspecies, and serovar levels.

Intact Protein Expression Profiles

To facilitate nontargeted SNP discovery, the intact accurate mass, retention time, and relative abundance of proteins from the soluble fraction of bacterial lysates are measured and compared using LC-MS. Figure 10.1a shows a representative total ion current chromatogram from a 90-min LC-MS analysis of an intact bacterial protein lysate. Mass spectra were summed in 30-s windows, and each window was deconvoluted using ProTrawler6 software (Williams et al. 2002). Unlike mass spectra of peptides, intact proteins produce broad charge state distributions, effectively splitting the ion current generated for a given protein over multiple structural conformations (Fig. 10.1b). The elution profile of each protein is 1.5 min wide on average, further distributing the ion current, as well as greatly increasing the likelihood of multiple co-eluting proteins. Consequently, software is necessary to deconvolute each spectrum (or summed spectra) (Fig. 10.1c) and merge consecutive abundances into a single protein mass and intensity. The result (Fig. 10.1d)

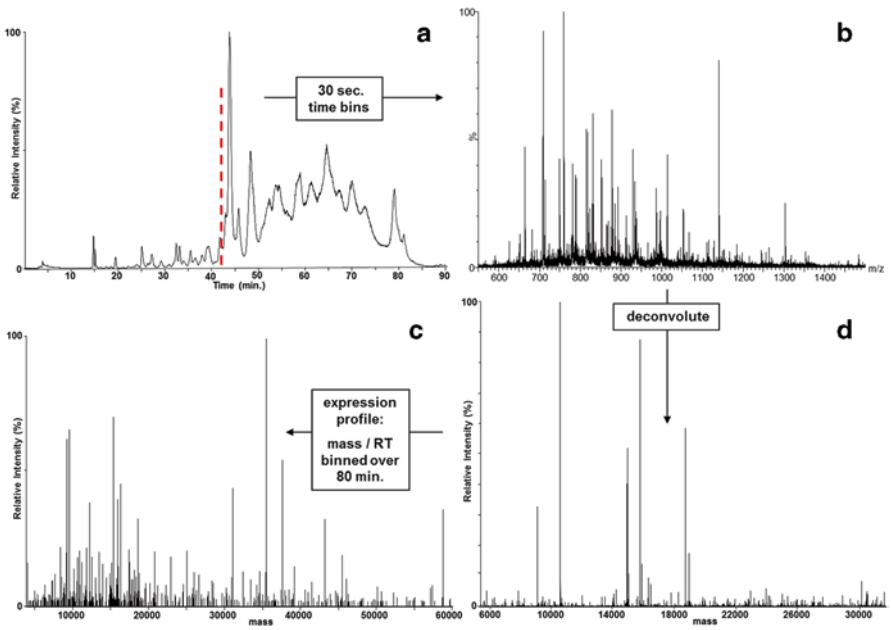


Fig. 10.1 Intact protein expression profile generation. ProTrawler software was used to deconvolute and reconcile all MS scans from the chromatogram into a single mass, retention time, and abundance profile. **a** Representative chromatogram from a 90-min LC-MS analysis of a *S. enterica* strain LT2 intact protein lysate. **b** Mass spectra were summed into 30-s bins across the chromatogram. **c** The resultant spectra at each time interval was deconvoluted to produce a series of neutral mass peak lists consisting of mass, retention time, and intensity. **d** Bins were merged into a single profile based on mass and retention time tolerance. The result is an intact protein expression profile that visually simplifies the assessment of protein differences between lysates. (Reprinted with permission from McFarland et al. 2014. Copyright 2015 American Chemical Society)

is an intact protein expression profile or mass map that represents the masses and intensities of all proteins detected across the chromatogram. This approach has the visual simplicity of a MALDI spectrum but with the greater information content provided by chromatographically resolved ESI spectra. The increase in the number of detectable masses provided by an extended mass range and improved ionization of proteins yields a greater capacity for differentiation as compared to MALDI-MS. The power of our method is the visualization of all proteins detected in an LC-MS experiment in a single spectrum, thus providing a quicker and more complete assessment of differences when compared to relying solely on LC-MS/MS protein or peptide identifications to assess changes between samples (Everley et al. 2008). Intact protein expression profiles facilitate rapid assessment of differential proteins as possible biomarkers and offer a larger dynamic range as compared to chromatographic alignment alone.

Tracing back to the source of a *Salmonella* contamination requires a minimum of serovar-level differentiation. Serovar differentiation is not currently possible on commercially available MALDI-based clinical bacterial typing platforms. *Salmo-*

nella enterica enterica Typhimurium and Heidelberg are closely related serovars that have both been implicated in food-based outbreaks (CDC 2014). Recent phylogenetic and MLST analysis (Bell et al. 2011) confirm that the chosen strains are members of two closely related serovars. Figure 10.2 shows a mirrored comparison of the LC-MS generated intact protein expression profiles of these serovars. Each profile is the result of deconvolution and binning of mass, abundance, and retention time from a representative 90-min LC-MS run. As is expected by the extreme homology across the *Salmonella* species and the similarity of these two serovars, the mass maps look nearly identical, with differences occurring in only a small number of detectable masses.

One can readily observe that the majority of masses detected are conserved across serovars. The observed mass shifts likely represent protein products of SNP-containing genes that differentiate *S. enterica* serovar Typhimurium strain LT2 from *S.* Heidelberg strain A39 and are likely biomarkers for serovar identification. No protein sequencing is required to determine the presence of mass shifts and/or novel masses, and markers do not need to be known prior to analysis.

Top-Down Protein Identification

It has been previously shown that comparisons of intact protein expression profiles are sufficient to differentiate two bacterial serovars (Williams et al. 2004, 2005; Everley et al. 2008). Although the presence of a differential pattern is sufficient for grouping a serovar with a set of previously run samples, it does not readily facilitate identification of uncharacterized strains and provides little to link the result with complementary assays such as targeted PCR probes or genome sequencing. Confirmation of the identity of differential masses as orthologs is necessary to validate the protein as a viable biomarker. The second stage of this method is the addition of top-down MS/MS identification of proteins to the existing LC-MS separation method (Fig. 10.2; McFarland et al. 2014). Proteins maintain the same elution profile but now the most abundant proteins are identified. The recent introduction of faster instruments with improved data-dependent selection increases the number of proteins identified in a single run.

Protein identifications in Fig. 10.2 are represented by the protein name, as assigned for the reference genome of *S.* Typhimurium strain LT2. A complete list of identified proteins and a detailed description of PTM assignments can be found in McFarland et al. (McFarland et al. 2014). Although, in general, the highly conserved protein sequences of related bacterial strains make strain typing challenging, it also means that the vast majority of fragment ions match across proteomes. Searching top-down MS/MS spectra does not require the strict precursor mass accuracy of bottom-up proteomics. In this work, the precursor mass error was permitted to be 1000 Da to account for unpredicted signal peptides and unknown PTMs, such

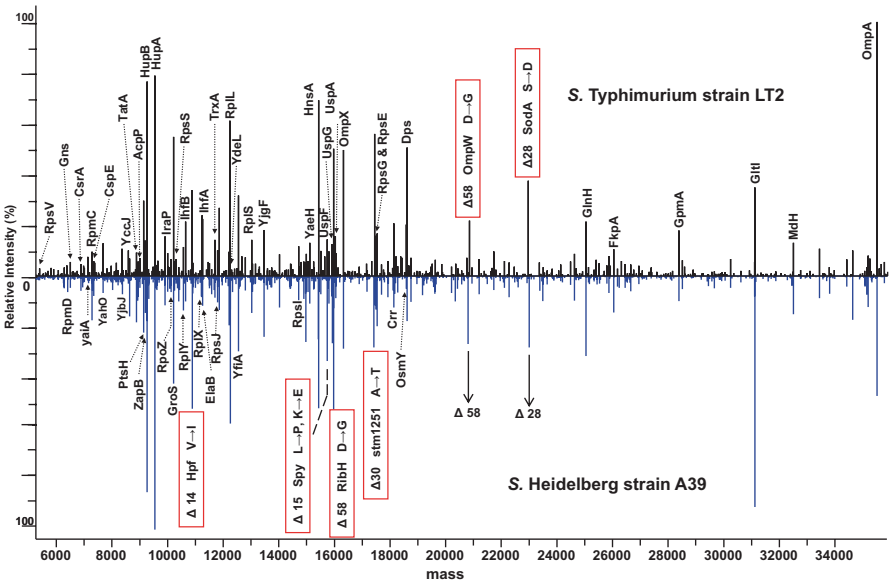


Fig. 10.2 Comparison of intact protein expression profiles for *S.* Typhimurium strain LT2 and *S.* Heidelberg *strain A39*. Profiles for these closely related serovars are similar but a small number of mass differences are evident. Approximately 80 proteins were subsequently identified by top-down LC-MS/MS analysis, each with a minimum of three MS/MS spectra and an e-value better than $1e^{-5}$. A subset of identified proteins is labeled with gene names. Proteins containing serovar-specific SNP-related mass differences are noted and the amino acid substitution shown

as lipidations. A fragment ion mass accuracy requirement of 20 ppm (Meng et al. 2001) provides sufficient specificity to identify sequence tags without an exact precursor mass. Consequently, one can confidently identify enough fragment ions to identify MS/MS fragment ion data to a homologous protein while still retaining the intact mass of the protein. Comparison of the measured intact mass with that of the identified protein readily determines whether the measured protein contains a mass shift.

Most observed masses show no discernable mass difference between the two *Salmonella* strains analyzed. Because we are able to readily identify the most abundant masses by top-down fragmentation, we can confirm that proteins that do produce serovar-specific mass shifts between *S.* Typhimurium and *S.* Heidelberg are indeed products of the same gene. Site-specific fragmentation at the SNP site is not necessary. Because we simultaneously detect the mass of the intact protein and fragment the intact precursor for identification, we can rely on accurate mass and retention time profiles to confirm that the identified proteins are related. Alignment of the in-silico predicted protein sequences can be used to confirm the presence of an amino acid change resulting from a non-synonymous SNP.

While a high-throughput top-down approach identifies fewer proteins and SNPs than a typical bottom-up survey, we gain independence from the need for a strain-specific sequenced genome. Comparison of intact protein expression profiles by mass, retention time, and relative abundance is sufficient for determination of masses that differ across serovars. Reproducible SNP identification in a bottom-up experiment would require the sequenced genome, such that the novel SNP must be present in the searched database. Identification of a SNP-containing peptide that is not in the database would require de-novo sequencing of unassigned peptides. Peptide SNP identification by spectral similarity alignment may be possible, but knowledge of the full degree of genetic drift is difficult without knowledge of the mass of the intact protein because complete peptide sequence coverage is rarely achieved. An obvious strength of the intact protein-based methodology presented here is that any differences as compared to proteins in a reference strain are readily apparent.

Proteogenomics

Maintaining a protein's intact mass while still being able to identify the protein to a homologous protein sequence is also advantageous for proteogenomic-based reconciliation of the mass spectrometric detection of expressed proteins with genome sequencing data. This provides a direct link to complementary genome-based methods as well as a mechanism for the detection of genome sequencing errors. For example, protein ElaB identified in *S.* Typhimurium strain LT2 has a theoretical mass of 418 Da greater than its measured mass. The identity of the measured mass was confirmed by CID fragmentation, with 21 y-ions identified. No b-type fragment ions were identified, and the measured mass differs from the theoretical mass as stated (Fig. 10.3a). The assigned e-value of 3.5 e^{-20} confirms confident protein identification, and the absence of b-ions points to a mass discrepancy at the N-terminus. The measured mass of the same protein in *S.* Heidelberg strain A39 does reconcile with its measured mass (after cleavage of the initiator methionine), strongly suggesting that the large mass discrepancy is not due to an unpredicted PTM. Alignment of the *S.* Typhimurium strain LT2 theoretical protein sequences with that of the same protein from another sequenced *S.* Typhimurium strain (strain U288) shows that the mass discrepancy lies at the translational start site of the protein (Fig. 10.3b). Confirmation of a sequencing start site error is seen in Fig. 10.3c. Removal of the erroneous amino acids increases the precursor mass accuracy to less than 3 ppm and results in the identification of a string of N-terminal containing b-type fragment ions. Identification of protein sequences combined with an intact mass measurement provides a unique link to genome sequencing and phylogenetic stg efforts. As the use of high-throughput genome sequencing annotation pipelines increases, validation of start site errors will minimize the propagation of start site errors through multiple genomes.

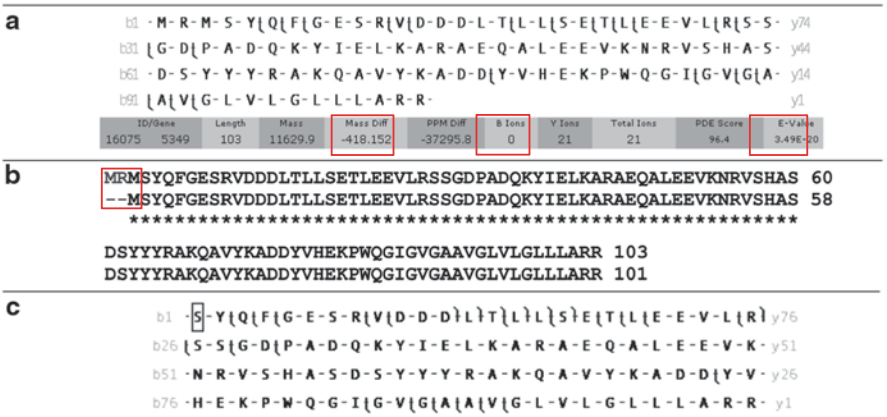


Fig. 10.3 Top-down mass spectrometry to verify genome annotation. **a** The intact mass measured in *S.* Typhimurium strain LT2 for SNP-containing protein ElaB does not agree with the theoretical mass. Genome sequencing predicts a larger mass difference between serovars than is actually expressed. Top-down MS/MS identifies the correct protein but no b-type fragment ions are assigned. **b** Comparison of the predicted protein sequence for strain LT2 against ElaB sequences predicted from other strains (here shown for *S.* Typhimurium strain U288) shows disagreement at the N-terminus. **c** Correction of the N-terminal amino acids in the LT2 sequence results in the additional identification of a substantial sequence tag of b-type ions. (Reprinted with permission from McFarland et al. 2014. Copyright 2015 American Chemical Society)

Multiplexed Serovar Identification of Semi-blinded Isolates

To demonstrate the specificity and scalability of intact protein LC-MS expression profiles for *Salmonella* serovar identification, the method was applied to a semi-blinded study of 36 *Salmonella* isolates originating from food-borne outbreaks (McFarland et al. 2014). Study creators established sample relatedness at the serotype, PFGE, and WGS levels.

Representative LC-MS generated intact protein expression profiles for each serovar are shown in Fig. 10.4. Labeled masses are SNP-containing proteins, SodA, YfeA, and OmpA. Combinations of these markers were sufficient to correctly identify the serovar type for all 36 *Salmonella* isolates, four serovars represented by nine isolates each. Neither the identity of the isolates nor the differentiating protein markers were known in advance. Markers were picked from the resultant LC-MS expression profiles, based on variable masses in abundant proteins. No one marker was sufficient to differentiate all four serovars. As is expected for blind identification, more than one marker is necessary. It is worth noting that top-down identifications of serovar-specific biomarkers did not need to be performed because protein identifications were known from previous top-down work on *S.* Heidelberg and *S.* Typhimurium (McFarland et al. 2014) and were confirmed based on the retention time. Differentiating protein markers were then used to confirm serovar assignments by comparing the measured masses with in-silico protein sequences from publically available protein databases, providing a direct link to genome sequencing data.

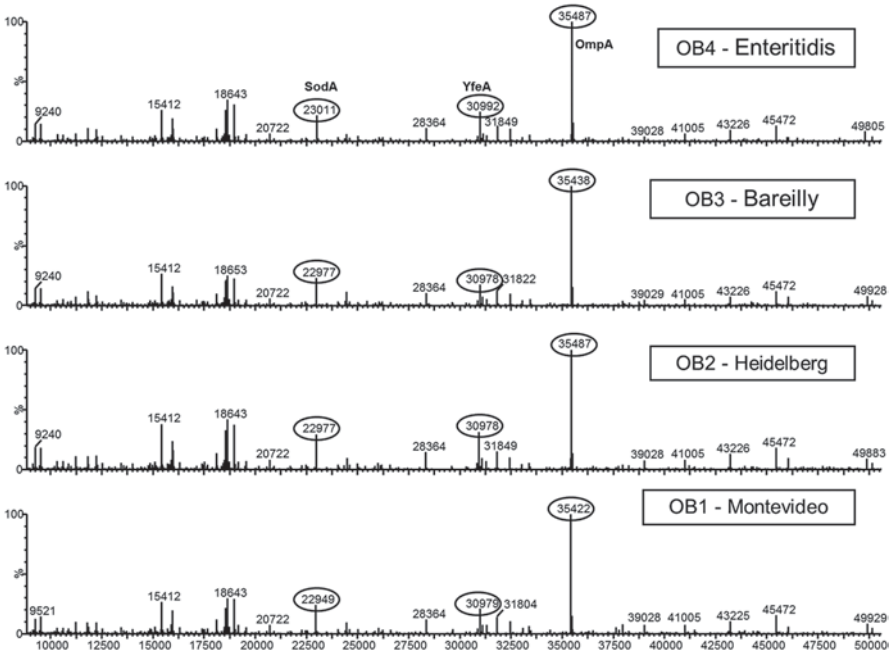


Fig. 10.4 Representative LC-MS generated intact protein expression profile for each serovar. Circled masses are SNP-containing proteins, SodA, YfeA, and OmpA. Combinations of these markers were sufficient to correctly identify serovar type for all 36 semi-blinded isolates. Markers did not need to be selected in advance. LC-MS profiles were acquired and markers were chosen based on the resultant data

LC-MS intact protein expression profiles assigned the correct serovar type for all 36 isolates, as determined by the study key based on PFGE and WGS of the outbreak samples. Mass and abundance profiles generated from triplicate analysis of each strain were used for PCA analysis. Each of the 36 isolates clustered into one of four distinct clusters corresponding to each of the four serovars. Although LC-MS may not provide the strain-level specificity of WGS, LC-MS should offer the same level of specificity as any marker-based method but without the need for preselection of markers. This offers flexibility given that different combinations of markers will be required depending on the serovar in question.

Conclusion

As the speed of whole genome sequencing increases and its cost decreases, strain-level bacterial differentiation will be decided at the genome level, rather than by expressed proteins. While the specificity required for strain-level typing may remain

the purview of phylogenetics, the use of mass spectrometry to track intact protein biomarkers at a serovar level would provide a cheaper, inherently multiplexed screen to determine the value of genetic sequencing. LC-MS/MS analysis not only supplies the detectable masses that differ between two samples (within the upper mass limit of the mass spectrometer) but also the identity of those masses. Knowledge of which gene products contain SNPs or which proteins have been newly transferred to a bacterial strain provides a direct link back to genome sequencing data, providing gene-specific validation at an expressed protein level.

The rapid rate of bacterial evolution translates to a moving target for strain and serovar-differentiating SNP-containing proteins. Any method meant to differentiate across multiple serovars would require a combination of multiple SNP-containing proteins. The advantage of nontargeted expression profiles generated in the method presented here is that any unpredicted changes that occur in the most abundant soluble proteins should be detected. Target marker proteins do not need to be known before sample analysis.

Identification of SNP-containing proteins becomes much quicker once initial identification of the most abundant expression profile masses has been established. Because the majority of the most abundant proteins are conserved across bacterial intact protein expression profiles of *Salmonella* serovars, it is not necessary to identify hundreds of proteins in each new isolate. Most abundant masses can be identified by matching the accurate mass and retention time to existing data from a reference strain. Only the compounds that exhibit a mass difference as compared to a standard strain may need to be analyzed by MS/MS for identity confirmation. This small subset of SNP-containing proteins can then be used to query the rapidly growing number of bacterial genomes as a gene name and intact mass (or mass difference) pair. Instead of comparing each new bacterial expression profile to a mass spectral data repository, we can take advantage of bacterial sequencing and alignment efforts and query for only the expressed proteins that show a change in mass. This targeted analysis would be quicker than whole-genome sequencing and more likely to detect genetic changes than multiplexed PCR or targeted mass spectrometry alone because the biomarkers do not need to be known in advance.

Acknowledgments The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the US Food and Drug Administration.

References

Arnold RJ, Reilly JP. Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. Anal Biochem. 1999;269(1):105–12.

Bell RL, Gonzalez-Escalona N, Stones R, Brown EW. Phylogenetic evaluation of the 'Typhimurium' complex of *Salmonella* strains using a seven-gene multi-locus sequence analysis. Infect Genet Evol. 2011;11(1):83–91.

Bizzini A, Greub G. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. Clin Microbiol Infect. 2010;16(11):1614–9.

Cargile BJ, McLuckey SA, Stephenson JL. Identification of bacteriophage MS2 coat protein from *E-coli* lysates via ion trap collisional activation of intact protein ions. Anal Chem. 2001;73(6):1277–85.

CDC. Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet Surveillance Report for 2012 (final report) (trans: U.S. Department of Health and Human Services C). CDC, Atlanta, GA; 2014.

Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. 2013;26(3):547–603.

Conway GC, Smole SC, Sarracino DA, Arbeit RD, Leopold PE. Phyloproteomics: species identification of Enterobacteriaceae using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Mol Microbiol Biotechnol. 2001;3(1):103–12.

Dieckmann R, Helmuth R, Erhard M, Malorny B. Rapid classification and identification of *salmonellae* at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2008;74(24):7767–78.

Everley RA, Mott TM, Wyatt SA, Toney DM, Croley TR. Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species. J Am Soc Mass Spectrom. 2008;19(11):1621–8.

Fagerquist CK, Bates AH, Heath S, King BC, Garbus BR, Harden LA, Miller WG. Sub-speciating *Campylobacter jejuni* by proteomic analysis of its protein biomarkers and their posttranslational modifications. J Proteome Res. 2006;5(10):2527–38.

FDA. Bad Bug Book: foodborne pathogenic microorganisms and natural toxins. 2nd ed.; 2012. pp. 9–13.

Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev. 2001;20(4):157–71.

Ho YP, Hsu PH. Investigating the effects of protein patterns on microorganism identification by high-performance liquid chromatography-mass spectrometry and protein database searches. J Chromatogr A. 2002;976(1–2):103–11.

Krishnamurthy T, Ross PL. Rapid identification of bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. Rapid Commun Mass Spectrom. 1996;10(15):1992–6.

Krishnamurthy T, Davis MT, Stahl DC, Lee TD. Liquid chromatography microspray mass spectrometry for bacterial investigations. Rapid Commun Mass Spectrom. 1999;13(1):39–49.

Lee SW, Berger SJ, Martinovic S, Pasa-Tolic L, Anderson GA, Shen YF, Zhao R, Smith RD. Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. Proc Natl Acad Sci U S A. 2002;99(9):5942–7.

Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao GJ, Meng JH, Stones R. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med. 2011;364(10):981–2.

McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. Anal Chem. 1997;69(4):767–76.

McFarland MA, Andrzejewski D, Musser SM, Callahan JH. Platform for identification of *Salmonella* serovar differentiating bacterial proteins by top-down mass spectrometry: *S.* Typhimurium vs *S.* Heidelberg. Anal Chem. 2014;86(14):6879–86.

Meng FY, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. Informatics and multiplexing of intact protein identification in bacteria and the archaea. Nat Biotechnol. 2001;19(10):952–7.

Mott TM, Everley RA, Wyatt SA, Toney DM, Croley TR. Comparison of MALDI-TOF/MS and LC-QTOF/MS methods for the identification of enteric bacteria. Int J Mass spectrom. 2010;291(1–2):24–32.

Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal Chem. 2001;73(4):746–50.

Wattiau P, Boland C, Bertrand S. Methodologies for *Salmonella enterica* subsp. enterica subtyping: gold standards and alternatives. Appl Environ Microbiol. 2011;77(22):7877–85.

Wilcox SK, Cavey GS, Pearson JD. Single ribosomal protein mutations in antibiotic-resistant bacteria analyzed by mass spectrometry. Antimicrob Agents Chemother. 2001;45(11):3046–55.

Williams TL, Leopold P, Musser S. Automated postprocessing of electrospray LC/MS data for profiling protein expression in bacteria. Anal Chem. 2002;74(22):5807–13.

Williams TL, Andrzejewski D, Lay JO, Musser SM. Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. J Am Soc Mass Spectrom. 2003;14(4):342–51.

Williams TL, Musser SM, Nordstrom JL, DePaola A, Monday SR. Identification of a protein biomarker unique to the pandemic O3: K6 clone of *Vibrio parahaemolyticus*. J Clin Microbiol. 2004;42(4):1657–65.

Williams TL, Monday SR, Edelson-Mammel S, Buchanan R, Musser SM. A top-down proteomics approach for differentiating thermal resistant strains of *Enterobacter sakazakii*. Proteomics. 2005;5(16):4161–9.

Wynne C, Edwards NJ, Fenselau C. Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an Orbitrap. Proteomics. 2010;10(20):3631–43.

Zamdborg L, LeDuc RD, Glowacz KJ, Kim Y-B, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S, Kelleher NL. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. Nucleic Acids Res. 2007;35:W701–6.

Part III
Drug Resistance Monitoring and Assays

Chapter 11
Rapid Profiling of Human Pathogenic Bacteria and Antibiotic Resistance Employing Specific Tryptic Peptides as Biomarkers

David Drissner, René Brunisholz, Ralph Schlapbach and Maria-Theresia Gekenidis

Introduction

Identification of human pathogens by matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry, based on profiling of mainly taxonomic relevant ribosomal proteins and comparison to a reference mass spectra database, has developed into a robust cutting-edge diagnostic technology and has *revolutionized work in microbiological laboratories in recent years* (Seng et al. 2009). *This is due to the* high speed of analysis allowing a short time to result and the streamlined protocol, enabling an accurate and cost-effective identification within less than 20 min. *Application fields include* clinical and veterinary diagnostics, food safety control, outbreak tracking, environmental microbiology, biotechnology, and biodefense. A major challenge to MS-based identification has been to reliably increase the taxonomic resolution to the below-species level. This challenge originates from the fact that strains of one species exhibit substantial genetic overlap and thus high protein similarity. Two main approaches have been applied to resolve mass peak variations below species level: the library-based and the proteomics-based approach. Within the library-based approach, both sample pre-treatment and data reduction strategies have been developed. Proteomics-based approaches comprise bottom-up and top-down characterization of biomarkers applying large databases available to the public.

The focus of this chapter is on the state of MALDI-TOF- and MALDI-TOF/TOF MS-based identification of human pathogens below the species level and specifically on the application of tryptic peptides as a recent development in enhancing the

D. Drissner (✉) · M.-T. Gekenidis
Agroscope Institute for Food Sciences, Food Microbiology and Analytics, Schloss 1, 8820 Waedenswil, Switzerland
e-mail: david.drissner@agroscope.admin.ch

R. Brunisholz · R. Schlapbach
Functional Genomics Center Zurich, ETH Zurich and University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_11

discriminatory power for bacterial profiling and determination of bacterial antibiotic resistances. Rapid identification at the below-species level is highly important in identification and diagnosis of pathogens, to determine appropriate drug therapy, to reliably trace back contamination sources in elucidation of epidemics, to improve food production and processing, or to develop better clinical practices.

Library-Based Approaches

In library-based approaches, peak lists extracted from a profile mass spectrum of unknown microorganisms are compared to the peak lists of reference spectra deposited in a database containing a large collection of well-characterized strains. This approach based on the detection of subtle and reproducible differences in spectra has been applied in most studies reporting successful profiling of pathogens below the species level using MALDI-TOF MS (Table 11.1). A comprehensive review on MALDI-TOF MS profiling of bacteria at the strain level has recently been published by Sandrin et al. (2013). The prerequisite for MALDI-TOF MS profiling is cultivation of pathogens on solid or liquid culture media and subsequently direct smearing of inactivated whole cells onto the MALDI target or short chemical extraction with formic acid and acetonitrile and spotting of supernatants onto the MALDI target. In general, the number of proteins detected increases with the level of separation and fractions collected which helps to increase taxonomic resolution. In both cases (direct smear or extract), the sample is covered with a standard MALDI matrix, for example, α-cyano-4-hydroxycinnamic acid. Mass spectra are acquired in positive ion mode from random locations on the target spot and comprise a mass range of 2–20 kDa (Ghyselinck et al. 2011; Ilina et al. 2010), of a broader (Hettick et al. 2006; Jackson et al. 2005; Teramoto et al. 2009) or narrower mass range (Keys et al. 2004; Rajakaruna et al. 2009). In particular, ions with high masses are promising for differentiation in below-species level due to rarity of these ions and to the absence of background signals in that mass range. Single mass peaks of spectra in library-based approaches are not given proof of identity; however, most of the peaks are attributed to basic, abundant, and conserved proteins, in particular ribosomal proteins (Sauer and Kliem 2010; Fenselau and Demirev 2001) and to a certain degree to proteins associated with bacterial cell walls (Evason et al. 2001). Ribosomal proteins comprise approximately 30 % of total proteins in a cell being in the exponential growth phase. Success in identification below the species level using library-based approaches requires robust software, reliable algorithms as well as databases in order to precisely compare acquired spectra to database entries and to calculate the similarity. Furthermore, mass spectral quality (resolution, accuracy, and reproducible acquisition of spectra) is key, and standardized experimental conditions including culture conditions need to be strictly followed in order to ensure reproducibility of the MALDI mass spectra and to detect specific protein biomarker masses for organisms below the species level.

Table 11.1 Overview of studies for profiling human pathogens with MALDI-TOF MS at the strain level: Examples of strain categorization (A), strain differentiation (B), and strain identification (C). *MRSA* methicillin-resistant *Staphylococcus aureus, MSSA* methicillin-sensitive *S. aureus, L* library-based approach, *P* proteomics-based approach, *CHCA* a-cyano-4-hydroxycinnamic acid, *CMBT* 5-chloro-2-mercaptobenzothiazole, *DHB* 2-hydroxy-5-methoxy benzoic acid, *FA* ferulic acid, *HABA* 2,4-hydroxyphenylazobenzoic acid, *SA* 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), *THAP* 2,4,6-trihydroxyacetophenon. Adapted and modified from Sandrin et al. (2013)

Bacterium	Approach	Cell preparation	Matrix	Software	Instrument	Reference
A						
Cyanobacteria	L	E	CHCA	ns	Voyager Elite; PerSeptive Biosystems	Erhard et al. 1997
Haemophilus	L	I/E	SA	ns	Voyager-DE; PerSeptive Biosystems	Haag et al. 1998
MRSA/MSSA	L	E	CMBT	ns	Kompact MALDI 2; Kratos Analytical	Edwards-Jones et al. 2000
Salmonella enterica	L	I	CHCA	ns	Voyager DE STR 4071; Perseptive, Biosystems	Leuschner et al. 2003
Francisella tularensis	L	E	SA	Bionumerics; Applied Maths	PBS II; Ciphergen Biosystems	Lundquist et al. 2005
MRSA	L	I	CMBT	in-house	Kratos Kompact MALDI 2; Shimadzu Biotech	Jackson et al. 2005
Streptococcus	L	I	CHCA	MATLAB; Math Works	Biflex III; Bruker Daltonics	Rupf et al. 2005
Neisseria gonorrhoeae	L	E	SA	Neuroshell; Ward Systems Group	PBS II Protein Chip Array Reader; Ciphergen Biosystems	Schmid et al. 2005
Moraxella catarrhalis	P	I/E	CHCA	MASCOT	Ultraflex; Bruker Saxonia Analytik	Schaller et al. 2006
Escherichia coli	L	I	SA	Diversity Database; Bio-Rad	Reflex IV; Bruker Daltonics	Siegrist et al. 2007
Escherichia coli	P	E	SA	ns	Voyager-DE STR; Applied Biosystems	Camara and Hays 2007
Enterococcus	L	E	SA	GelCompar II; Applied Maths	Reflex IV; Bruker Daltonics	Giebel et al. 2008

Table 11.1 (continued)

Bacterium	Approach	Cell preparation	Matrix	Software	Instrument	Reference
Listeria	L	E	CHCA	BioTyper, Bruker Daltonics	Microflex LT; Bruker Daltonics	Barbuddhe et al. 2008
Clostridia	L	E	CHCA	BioTyper; Bruker Daltonics	Microflex LT; Bruker Daltonics	Grosse-Herrenthey et al. 2008
Rhodococcus erythropolis	L	E	SA	Bionumerics; Applied Maths	Axima CFR; Shimadzu/Kratos	Teramoto et al. 2009
Yersinia pestis	L	I	CHCA	BioTyper; Bruker Daltonics	Autoflex II; Bruker Daltonics	Ayyadurai et al. 2010
Staphylococcus	L	E	CHCA		Ultraflex II TOF-TOF; Bruker Daltonics	Dubois et al. 2010
MRSA/MSSA	L	E	CHCA	BioTyper; Bruker Daltonics	Microflex LT; Bruker Daltonics	Szabados et al. 2010
Escherichia coli	L	E	CHCA	BioTyper; Bruker Daltonics	Ultraflex I; Bruker Daltonics	Karger et al. 2011
MRSA	L	E	CHCA	BioTyper; Bruker Daltonics	Microflex LT; Bruker Daltonics	Wolters et al. 2011
MRSA	L	E	SA	ns	PBS II Protein Chip Array Reader; Ciphergen Biosystems	Shah et al. 2011
Yersinia enterocolitica	L, P	I	SA	ns	Ultraflex II TOF-TOF; Bruker Daltonics	Kraushaar et al. 2011
Yersinia enterocolitica	L	I	SA	SARAMIS; bioMérieux	Axima Confidence; Shimadzu-Biotech	Stephan et al. 2011
B						
Gram (−)/Gram (+) enterobacteria	L	I	CHCA	ns	Kompact MALDI III; Kratos Analytical	Claydon et al. 1996
Various pathogenic and non-pathogenic	L	E	CHCA, SA	ns	Vestec 2000; Vestec Instruments	Krishnamurthy et al. 1996

Table 11.1 (continued)

Bacterium	Approach	Cell preparation	Matrix	Software	Instrument	Reference
Cyanobacteria	L	I	CHCA	ns	Voyager Elite; PerSeptive Biosystems	Erhard et al. 1997
Enterobacteriaceae	L	I	CHCA	ns	In-house	Lynn et al. 1999
Helicobacter pylori	L	E	CHCA, SA, FA	ns	Reflex; Bruker-Franzen	Nilsson 1999
MRSA, MSSA	L	I	CBMT	ns	Kompact MALDI 2; Kratos Analytical	Edwards-Jones et al. 2000
Bacillus	L	I,E	SA	ns	Kompact MALDI 4; Kratos Analytical	Ryzhov et al. 2000
Bacillus	L	E	CHCA, SA	ns	TofSpec 2E; Micromass Ltd.	Elhanany et al. 2001
MRSA/MSSA	L	E	CHCA, SA, CMBT, MCA	ns	Bruker Reflex III; Bruker Saxonia Analytic	Bernardo et al. 2002
MRSA	L	I	CMBT	ns	Kompact MALDI 2; Kratos Analytica	Walker et al. 2002
Bacillus	L	E	FA	ns	Reflex II; Bruker Daltonics	Dickinson et al. 2004
Bacillus	L	I, E	CHCA, SA, DHB, THAP	ns	TofSpec 2E; Micromass Ltd.	Horneffer et al. 2004
Acinetobacter, Escherichia coli, Salmonella	L	E	CHCA, SA, CMBT, FA	ns	Tofspec 2E, Micromass Ltd.	Ruelle et al. 2004
Campylobacter	B	E	FA	GPM; Global Proteome Machine	Reflex II; Bruker Daltonics	Fagerquist et al. 2005
Campylobacter	L	E	FA	ns	Reflex II; Bruker Daltonics	Mandrell et al. 2005
Escherichia coli	P	I	CHCA, SA, FA	Masslynx; Waters Corp.	MALDI LR; Micromass Ltd.	Ochoa and Harrington 2005
Aeromonas	L	I	SA	ns	BiFlex III; Bruker Daltonics	Donohue et al. 2006
Campylobacter	P	E	FA	in-house	Reflex II; Bruker Daltonics	Fagerquist et al. 2006

Table 11.1 (continued)

Bacterium	Approach	Cell preparation	Matrix	Software	Instrument	Reference
Mycobacterium	L	E	CHCA	ns	PBS IIC; Ciphergen Biosystems	Hettick et al. 2006
MRSA	L	I	CMBT	MicrobeLynx, Micromass Ltd.	ns; Micromass Ltd.	Majcherczyk et al. 2006
Bacillus anthracis, Bacillus cereus	P	I,E	CHCA	Data Explorer; Applied Biosystems	Ultraflex I, Bruker Daltonics; 4700 Proteomics Analyzer TOF-TOF; Applied Biosystems	Castanha et al. 2007
Coxiella burnetii	L,P	I	CHCA, SA, CMBT	MATLAB; Math Works Inc.	4700 Proteomics Analyzer TOF-TOF; Applied Biosystems	Pierce et al. 2007
Bacillus	L	I	DHB	ns	Omniflex; Bruker Daltonics	Price et al. 2007
Francisella tularensis	L	E	SA	CIPHERGENEX-PRESS, Ciphergen Biosystems	4000 Enterprise Edition; Ciphergen Biosystems	Seibold et al. 2007
Escherichia coli	L	I	SA	MATLAB; Math Works	MALDI LR; Micromass Ltd.	Chen et al. 2008
Escherichia coli, Shigella	L	E	SA	ns	Ultraflex II; Bruker Daltonics	Everley et al. 2008
Streptococcus pyogenes	L	I	CHCA, SA, DHB	ns	AB 4700 Proteomics Analyzer TOF-TOF; Applied Maths	Moura et al. 2008
Streptococcus	L	I	SA	ns	AB 4700 Proteomics Analyzer TOF-TOF; Applied Maths	Williamson et al. 2008
Escherichia coli	P	E	CHCA, SA	In-house	4800 TOF-TOF; Applied Biosystems	Fagerquist et al. 2010
Pseudomonas	P	I	SA	ns	Axima Performance; Shimadzu/Kratos	Hotta et al. 2010
Enterobacteriaceae	L	E	SA	ns	Ultraflex II; Bruker Daltonics	Mott et al. 2010
Legionella	L	E	CHCA	BioTyper, Bruker Daltonics	Autoflex II; Bruker Daltonics	Fujinami et al. 2011

Table 11.1 (continued)

Bacterium	Approach	Cell preparation	Matrix	Software	Instrument	Reference
C						
Bordetella, as-yet-uncultured	P	E	3-HPA	SEQUENOM; Sequenom		Von Wintzingerode et al. 2002
Micrococcaceae (CoNS)	L	I	DHB	BGP database	Autoflex; Bruker Daltonics	Carbonnelle et al. 2007
Neisseria meningitidis	P	E	3-HPA	SEQUENOM; Sequenom	ns	Honisch et al. 2007
Stenotrophomonas maltophilia	L	E	3-HPA & DAC	Data Explorer; Applied Biosystems		Jackson et al. 2007
Salmonella	P	I	SA, CHCA, DHB	SARAMIS; bioMérieux	Ultraflex II TOF-TOF; Bruker Daltonics	Dieckmann et al. 2008
Streptococcus agalactiae	L	E	CHCA	BioTyper; Bruker Daltonics	Ultraflex III TOF-TOF; Bruker Daltonics	Lartigue et al. 2009
MRSA	P	E	3-HPA	MassARRAY Typer; Sequenom	MassArray Compact Analyzer; Sequenom	Syrmis et al. 2011

E extract; *I* intact; *ns* not specified

Regarding strain categorization, serovars of *Salmonella enterica* subsp. *enterica* have been successfully categorized by comparison of their MALDI mass spectra, which contained up to 500 mass peaks in that study (Leuschner et al. 2003). Karger et al. (2011) employed a library-based approach to categorize STEC serovars (Karger et al. 2011), and Stephan et al. (2011) categorized *Yersinia enterocolitica* as pathogenic or non-pathogenic strains (Stephan et al. 2011). Further studies showed that strains of *Yersinia pestis* could be categorized according to their biotypes (Ayyadurai et al. 2010), strains of *Escherichia coli* according to their environmental origin, and clinical strains of *Moraxella catarrhalis* have been categorized at the subpopulation level (Schaller et al. 2006). *Listeria monocytogenes* was categorized at the level of clonal lineage, whereby the MALDI MS-derived lineage agreed with those from pulsed-field gel electrophoresis (Barbuddhe et al. 2008). Categorization of methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin-susceptible *Staphylococcus aureus* (MSSA) strains using characteristic markers for the methicillin resistance status has been achieved by Edwards-Jones et al. (2000) and by Shah et al. (2011) using artificial neural networks. Wolters et al. (2011) and recently, Josten et al. (2013) categorized strains according to the major clonal complexes of MRSA.

In order to differentiate between single bacterial strains, MALDI-TOF mass spectra have been used to identify mass peaks as biomarkers for the respective strains. Such an approach has been applied to *Helicobacter pylori* (Nilsson 1999), *E. coli* (Lynn et al. 1999), *Campylobacter* (Mandrell et al. 2005), *Mycobacterium* (Hettick et al. 2006), and MRSA (Majcherczyk et al. 2006). Williamson et al. (2008) differentiated strains of *Streptococcus pneumoniae* by using unique mass peaks (Williamson et al. 2008). Masses in the range of 5000–11,000 Da matched ribosomal proteins of *S. pneumoniae,* and spectrum clustering revealed the relationship between an outbreak of *S. pneumoniae* conjunctivitis and their corresponding isolates. Similarly, *Streptococcus pyogenes* strains could be differentiated by Moura et al. (2008) into invasive and non-invasive isolates using specific biomarkers (Moura et al. 2008). Many of the biomarker masses in the range of 4000–14,000 Da matched *S. pyogenes* ribosomal proteins. Differentiation of *Enterococcus faecium* and *E. faecalis* at the strain level has been described by Albesharat et al. (2011). Intact mycobacteria could be differentiated at the strain level by linear discriminant analysis (Hettick et al. 2006). Differentiation has been successful even when the intensity of the mass peaks was considered additionally to the presence or absence data. Pierce et al. (2007) demonstrated differentiation of *Coxiella burnetii* strains using partial least squares discriminant analysis of MALDI-TOF mass spectral peaks (Pierce et al. 2007).

For identification of single, unknown strains—compared to categorization or differentiation—the entire mass spectrum is usually used and compared to a library of reference spectra of known strains. Distinct algorithms have been applied in correlations calculated and often small spectral differences between strains that have been given more weight (weighted pattern matching) increased the sensitivity of such small differences and thus contributed to successful identification, for example, in studies using *E. coli* (Arnold and Reilly 1998), or Micrococcaceae (Carbonnelle

et al. 2007). In the study by Arnold and Reilly (1998), strains exhibited both peaks in common and also strain-specific peaks in the range of 3.5–10 kDa. By applying an algorithm calculating both cross-correlation and auto-correlation values for each of 13 intervals, 25 strains could be distinguished. Bright et al. (2002) applied a pattern recognition algorithm to the mass spectra (m/z 500–10,000), and each spectrum was translated into a point vector in an n-dimensional space. Data of 35 strains from 20 species and mainly enterobacteria were included in a reference library and correct identification on the strain level was achieved for 79% of the samples. The algorithm succeeded even in the distinction of species for which biochemical typing fails, for example, for *E. coli* O122 and *Citrobacter freundii*. A hierarchical cluster algorithm combined with analysis of variance (ANOVA) was used in a study by Hsieh et al. to extract biomarkers from several isolates of six human pathogens (Hsieh et al. 2008).

In general, two kinds of algorithms exist: one includes intensities of peaks and the other uses the presence and absence of mass peaks. It is worth mentioning that spectral mass signals exhibit an analytical error due to slight variability of acceleration voltage, to status of matrix crystals, and to peak recognition by the software. With respect to linear MALDI-TOF MS, an analytical error of approximately 500 ppm, meaning a 5 Da deviation for a signal at $m/z = 10{,}000$, is generally regarded acceptable. Besides software applications developed in-house, two main commercially available and automated softwares including validated reference databases are available (BioTyper, Bruker Daltonics (Sauer et al. 2008) and SARAMIS, bioMérieux (Kallow et al. 2000)), which also allow analysis of MALDI-TOF mass spectra on the below-species level as shown, for example, by Grosse-Herrenthey et al. (2008) using BioTyper to identify clostridia at the strain level or by Stephan et al. (2011) using SARAMIS for characterization of *Y. enterocolitica* strains according to their biotype. Such databases are constantly improving by inclusion of new bacteria relevant to clinical diagnostics, veterinary medicine, food safety, and environmental microbiology. To obtain more mass peaks serving as putative biomarkers and to increase sensitivity, in several studies samples have been treated by enzymes, detergent, sonication, corona plasma discharge, or heat (Nilsson 1999; Horneffer et al. 2004; Krishnamurthy et al. 1996; Ryzhov et al. 2000). Furthermore, in some studies mass spectra that contained less peaks have been applied for discrimination of strains as shown, for example for *M. catarrhalis* strains (Schaller et al. 2006), *S. aureus* (Shah et al. 2011), or *Francisella tularensis* (Seibold et al. 2007). In the latter study, a method applying surface-enhanced laser desorption/ionization has been used.

Proteomics-Based Approaches

The rapid increase in the availability of full genomes of bacteria in public databases boosted research of proteomics-based approaches comprising identification of single peaks in mass spectra in order to profile pathogens below the species level. Both

application of MALDI-TOF MS and MALDI-TOF/TOF MS have been described for identification of intact proteins serving as biomarkers. This comprises the use of their masses which are compared to in silico-generated protein databases derived from genomic databases. Intact protein identification has been successfully used to identify strain-specific protein biomarkers, for example, for *E. coli* O157:H7 (Ochoa and Harrington 2005), *Campylobacter* (Mandrell et al. 2005), and *Salmonella* (Dieckmann et al. 2008).

In bottom-up approaches, proteins extracted from bacterial cultures are digested enzymatically at specific sites and resulting peptides are identified by MS/MS (post-source decay, laser-induced dissociation, or collision-induced dissociation). Site-specific digestion is generally performed using proteolytic enzymes such as trypsin (Aebersold and Mann 2003; Yao et al. 2002). In order to accelerate digestion, microwave heating has been successfully applied (Lill et al. 2007). Non-enzymatic protein digestion by acid hydrolysis accelerated through microwave heating has been performed for analyzing spores of *Bacillus* (Swatkoski et al. 2006). Bottom-up approaches often include a separation and purification step prior to digestion. Fagerquist et al. (2005) applied high-performance liquid chromatography (HPLC) and 1D sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to proteins from *Campylobacter* before identifying strain-specific biomarker proteins. Two-dimensional SDS-PAGE has been used by Schaller et al. (2006) prior to biomarker identification for *M. catarrhalis* strains.

Regarding further optimization and speeding-up of bottom-up identification workflows, a proteomics-based approach that was developed recently to identify subspecies of *Salmonella enterica* (Gekenidis et al. 2014) is described in detail in the next chapter. This approach comprises whole-cell protein extracts produced via an established extraction procedure (MALDI biotyping) and high-intensity focused ultrasound (HIFU)-assisted trypsin digestion prior to identification of specific peptides and proteins.

In contrast, top-down proteomics approaches in profiling bacteria comprise accurate measurement of the mass of intact proteins and fragmentation of these by MS/MS yielding partial amino acid sequences and/or peptide fragments. Fragmentation is achieved by collision-induced dissociation, laser-induced dissociation, electron capture dissociation, or electron transfer dissociation. Resulting MS/MS spectra are compared to a database in order to identify the protein and ultimately—in the case of sufficiently unique protein sequence—the source strain. Software applications compare the masses of MS/MS fragment ions to a database of in silico fragment ions (a-, b-, and y-fragment ions) derived from a large number of protein sequences which exhibit the same mass as that of the biomarker. An algorithm calculates the probability of identification. MALDI-TOF/TOF MS has been used for identification of intact spores that were treated with 10 % formic acid on-target to facilitate extraction of small acid soluble proteins (Demirev et al. 2005). In another study, proteins were extracted with water–acetonitrile–TFA under bead-beating using 0.1 mm zirconia/silica beads for 1 min prior to biomarker identification of *E. coli* O157:H7 via MALDI-TOF/TOF MS (Fagerquist et al. 2010). Furthermore, shiga toxins of *E. coli* O157:H7 have been identified by this approach (Fagerquist and Sultan 2011) .

Future applications of MALDI in top-down approaches will need further developments to make fragmentation of large proteins more efficient (McLuckey 2010). Compared to library-based approaches, proteomics-based approaches are at an advantage with higher level of specificity and independence of producing mass spectral profiles with reproducible relative intensities of mass peaks.

Approaches Based on Tryptic Peptides Toward Identification and Typing of Pathogens Below the Species Level

Exploiting MALDI-TOF/TOF MS for Discrimination of Subspecies: In Search of Microorganism-Specific Tryptic Peptides

One approach to increase the taxonomic resolution of classical MALDI-TOF biotyping is by analyzing protein digests in the so-called bottom-up approach (as reviewed above). We have recently described a method for discrimination of bacterial subspecies relying on the ultra-fast generation of tryptic peptides enabling the identification of subspecies-specific biomarker peptides (Fig. 11.1). For the proof of concept, we used a model system consisting of the three *Salmonella enterica*

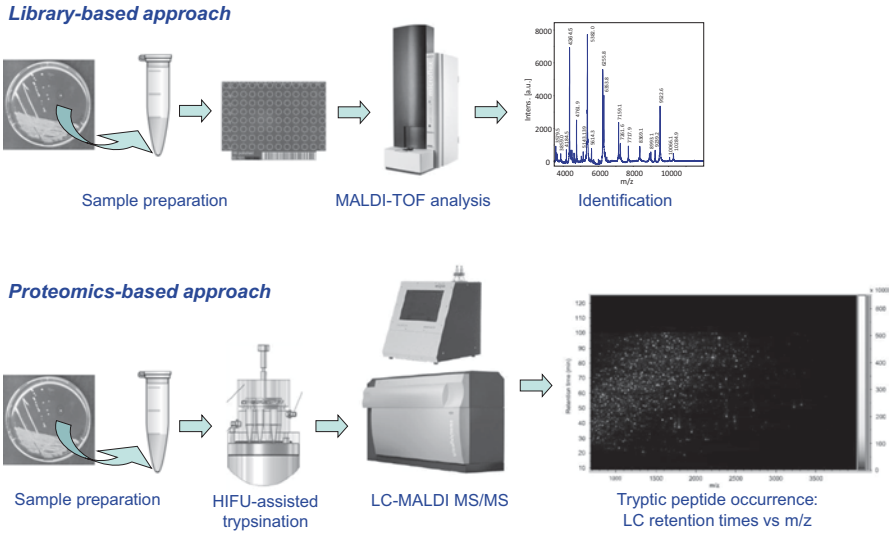


Fig. 11.1 Classical library-based approach for identification of bacteria using MALDI-TOF MS *(top)* and workflow for the novel proteomics-based approach for identification using HIFU-assisted trypsination and LC-MALDI MS/MS *(below)* as described in detail in Sect. 2 of this chapter

subspecies: *arizonae, enterica,* and *houtenae*. It is, to the best of our knowledge, the first study using the classical MALDI biotyping extract directly for proteomic analyses in a bottom-up approach. This rapid procedure allows generation of tryptic peptides within minutes without need for any further processing straight from a simple whole-cell extract. Details of the experimental procedure and data analysis allowing identification of unique biomarker peptides for subspecies discrimination shall be given in the next section.

Experimental Procedure

After selecting a representative strain for each of the subspecies to be discriminated within the genus of interest (e.g., *Salmonella*), a classical MALDI biotyping extraction is performed for each subspecies using a simple acid/organic solvent extraction as previously described (Sauer et al. 2008). We recommend using at least three biological replicates per subspecies. For extraction, an overnight culture of each strain is diluted to $OD_{600\ nm} = 1$. One milliliter of each diluted culture is centrifuged at $5000 \times g$ for 1 min, the medium is discarded, and the resulting pellet is washed with deionized water. The resulting cell suspension is spun again and the pellet is resuspended in 75% ethanol. After centrifuging the sample for 2 min at $16{,}100 \times g$, the supernatant is discarded and the residual ethanol is allowed to evaporate for 5 min at room temperature. In the following steps, the proteins are extracted by consecutive addition and vortexing in 100 µl of 70% formic acid and 100 µl of 100% acetonitrile. Finally, after another spin ($16{,}100 \times g$, 2 min) the supernatant containing the extracted proteins is transferred to a fresh Eppendorf tube where the extract can be stored at −20 °C until further analysis.

In order to generate tryptic digests, each extract should first be dried completely using a SpeedVac concentrator (room temperature, approximately 30–45 min). The remaining pellet which is hardly visible is then resolubilized in 3 µl of 100% acetonitrile, 3 µl of RapiGest, 18 µl of 100 mM Tris-HCl (pH 8.2), and 3 µl of 1 M Tris-HCl (pH 8.2) using 5 min HIFU treatment (UTR200, Hielscher Ultrasonics, Teltow, Germany) (intensity 90%, cycles 0.8). Trypsin (0.1 mg/ml in 10 mM HCl) is then added and the extracted proteins are digested within only 15 min under HIFU treatment, which significantly accelerates the generation of peptides otherwise lasting up to 2–16 h depending on the incubation temperature. Up to six samples can be processed simultaneously. In order to avoid a drastic temperature increase, HIFU treatment should always be performed in an ice water bath. Each digest is subsequently spotted onto a MALDI target (e.g., MTP AnchorChip 1536 TF) using a nano-LC system coupled to a fraction collector. The eluting peptides are directly mixed in the fraction collector with a matrix solution containing α-cyano-4-hydroxycinnamic acid (CHCA). One per eight spots can be manually spotted with a peptide calibration standard diluted in CHCA matrix at a ratio of 1:200. Details of the system, reagents, and operation parameters are given in Table 11.2.

Table 11.2 System requirements, reagents, and operation parameters for nano-LC spotting of tryptic digests onto the MALDI target for MALDI-TOF/TOF MS analysis

nano-LC column	15 cm × 75 µm C18 column; particle size 3 µm, pore size 100 Å
Trap column	2 cm × 100 µm C18 reversed phase column (solid phase extraction); particle size 5 µm, pore size 120 Å
Reagents	
CHCA tryptic digest	748 µl acetonitrile–water–TFA (95:4.9:0.1, vol/vol/vol) 36 µl saturated CHCA (10 mg/ml) in acetonitrile–water–TFA (90:9.9:0.1, vol/vol/vol) 8 µl of 10 % TFA 8 µl of 100 mM $NH_4H_2PO_4$ dissolved in water
CHCA peptide calibration standard	748 µl acetonitrile–water–TFA (85:14.9:0.1, vol/vol/vol), rest as for CHCA tryptic digest
Mobile phase A	0.1 % TFA in water
Mobile phase B	0.1 % TFA in 90 % acetonitrile
Operation parameters	
Spotting interval	Six spots per minute
Elution gradient	Linear, mobile phase B from 2 to 45 %, 64 min, 300 nl/min

After having spotted the tryptic digests, the MALDI-TOF/TOF MS spectra are acquired. The obtained data are then searched on the MASCOT search engine using the National Center for Biotechnology Information (NCBI) or UniProtKB/SwissProt database. The latter should be preferred because of the higher reliability of its reviewed data entries; however, one has to make sure that it contains a satisfactory number of entries for the subspecies under investigation. The NCBI database on the other hand will yield more potential biomarker peptides than the UniProtKB/SwissProt database search. To limit the amount of irrelevant peptide matches, the search can be restricted to the genus under investigation. A peptide decoy database and the MASCOT Percolator algorithm may be used to increase the significance of search results. Further suggested acquisition and search parameters are listed in Table 11.3.

The next step is the search for potential biomarker peptide masses, for example, by applying a Microsoft Excel macro to the data (as described by Gekenidis et al. 2014). The MALDI-TOF/TOF measurement yields compound lists for each digest containing the masses (*m*/*z* values) of all the peptides measured as well as the corresponding signal-to-noise (*S*/*N*) ratios. A biomarker peptide mass is defined as an *m*/*z* value present in all sets of one subspecies but absent from all sets of the others. To identify potential biomarker *m*/*z* values, all lists obtained for the analyzed subspecies and biological replicates are merged in one Excel table and sorted by increasing *m*/*z* values after having added a tag to each entry (see Table 11.4). Then, the *S*/*N* ratio and a ppm value are defined. Note that the *S*/*N* ratio should be equal to or above the value defined for MS/MS precursor selection (see Table 11.3), whereby only masses for which an MS/MS spectrum was acquired will be considered. The macro will select datasets with an *S/N* ratio equal to or greater than the defined value and

Table 11.3 Parameters for MALDI-TOF/TOF MS data acquisition and peptide search

MS/MS precursor selection	Signal-to-noise ratio threshold 10
Compound merging	When separated by less than six fractions, mass tolerance ± 50 ppm
Measurement settings	Laser frequency 1000 Hz Positive reflectron mode Acquisition range 700–4000 Da 3000 shots per spot, 100 shots per raster spot Laser intensity and detector sensitivity: highest peak 10^4–10^5 arbitrary units
Peptide search	
Search restrictions	Tryptic peptides with variable methionine, histidine, tryptophan oxidation One miscleavage Peptide tolerance ± 50 ppm, peptide charge + 1 MS/MS tolerance ± 0.7 Da

Table 11.4 Extract of compiled compound tables from *Salmonella enterica* subsp. *arizonae, S. enterica* subsp. *houtenae,* and *S. enterica* subsp. *enterica* prepared for biomarker peptide mass search (raw data from Gekenidis et al. 2014)

Tag	*m/z*	*S/N*
enterica_a	2703.26023	6.4
enterica_b	2703.27328	4.2
enterica_a	2703.28559	9.0
houtenae_a	2703.29299	10.0
enterica_b	2703.29761	6.8
arizonae_a	2703.30283	5.5
enterica_b	2703.30725	5.6
enterica_b	2703.30853	7.8
houtenae_b	2703.30992	6.8
enterica_a	2703.31167	8.6
arizonae_a	2703.31628	7.4
houtenae_c	2703.32243	8.5
enterica_b	2703.32295	4.4
houtenae_c	2703.33365	106.7
houtenae_a	2703.33666	14.3
arizonae_b	2703.33781	3.9
enterica_b	2703.33957	3.3
enterica_b	2703.34086	4.0
arizonae_a	2703.37732	3.0
houtenae_b	2703.37971	33.7

create a pivot table containing the *m*/*z* values in the row fields and the tag of the different measurements in the column fields. Each *m*/*z* value is then taken as a center, *m*/*z* values within the surrounding ppm window are counted, and the counts of the tag of the different measurements are recorded in the pivot table (see Table 11.5). Such a pivot table will give an overview of the number of *m*/*z* signals within the

Table 11.5 Extract of the pivot table generated from data on *Salmonella enterica* subsp. *arizonae, S. enterica* subsp. *houtenae,* and *S. enterica* subsp. *enterica* (*S/N* = 10, ± 100 ppm). Each *m/z* value is used as center of a ppm window to search for adjacent *m/z* values (see the text). Bold numbers indicate to which biological replicate the respective *m/z* value belongs. Two potential biomarker peptides are highlighted in green and purple (raw data from Gekenidis et al. 2014). In the case of the potential biomarker peptide highlighted in green, four *m/z* values were observed in the three biological replicates of *S. enterica* subsp. *houtenae* (2703.29299 and 2703.33666 in *houtenae_a,* 2703.33365 in *houtenae_b,* and 2703.37971 in *houtenae_c*), but none in the other two subspecies. The average of those four *m/z* values was therefore defined as a potential biomarker peptide mass for *S. enterica* subsp. *houtenae*.

m/z	*arizonae_a*	*arizonae_b*	*arizonae_c*	*houtenae_a*	*houtenae_b*	*houtenae_c*	*enterica_a*	*enterica_b*	*enterica_c*
2701.34674	1	**2**	1	1	2	1	1	2	2
2701.34875	1	2	1	1	2	1	1	**2**	2
2701.39451	1	2	1	1	2	1	**1**	2	2
2702.26381	2			1				**1**	
2702.27090	2			**1**				1	
2702.27861	**2**			1				1	
2702.30037	**2**			1				1	
2703.29299				**2**	1	1			
2703.33365				2	**1**	1			
2703.33666				**2**	1	1			
2703.37971				2	1	**1**			
2704.23789	1	1	**1**						
2704.28700	**1**	1	1						
2704.29833	1	**1**	1						
2705.06338		2	**2**		1	1		1	1
2705.06406		**2**	2		1	1		2	1
2705.13323		2	2		1	**1**	2	3	2

ppm window per set of each subspecies. An extract of a macro used for creation of such a pivot table is given in Fig. 11.2. After having generated the pivot table, the *m/z* values of biomarker peptides are selected and pasted into a new Excel sheet (for examples of such biomarker peptide masses, see marked entries in Table 11.5). Finally, all potential biomarker peptide masses are compared to the results obtained from the MASCOT search in order to find the masses belonging to actual peptides in proteins of the investigated genus. For those actual peptides, a sequence comparison can be made between the biomarker peptide and the corresponding peptides of

```
…
RawSheet = ActiveSheet.Name
NrRows = ActiveSheet.UsedRange.Rows.Count
NrColumns = ActiveSheet.UsedRange.Columns.Count

…
'Determine start row
For Z = 3 To NrRows
    If (Cells(Z, MZColumn) - Cells(2, MZColumn)) > (MZWindow * Cells(Z, MZColumn).Value /
2)
Next Z
StartRow = Z - 1

'Determine end row
For Z = 1 To (NrRows - 1)
    If (Cells(NrRows, MZColumn) - Cells((NrRows - Z), MZColumn)) > (MZWindow *
Cells((NrRows - Z), MZColumn).Value / 2) Then Exit For
Next Z
EndRow = NrRows - Z

'Row-by-row processing of data
PivotRow = 1
For Z = StartRow To EndRow

'Check S/N limit: jump to next row if S/N ratio lies below defined limit
    If Cells(Z, SNColumn).Value < SNLimit Then GoTo NextZ
    PivotRow = PivotRow + 1
    Tag = Sheets(RawSheet).Cells(Z, TagColumn).Value
    For i = 2 To NrColumnsPivot
        Select Case Tag
            Case Pivot(i)
                PivotColumn = i
                Exit For
        End Select
    Next i
    Sheets(PivotSheet).Cells(PivotRow, 1).Value = Sheets(RawSheet).Cells(Z,
MZColumn).Value
    Sheets(PivotSheet).Cells(PivotRow, PivotColumn).Value = 1 +
Sheets(PivotSheet).Cells(PivotRow, PivotColumn).Value
    Sheets(PivotSheet).Cells(PivotRow, PivotColumn).Font.Bold = True

'Analyze ppm window from center to upper m/z limit
    For O = (Z + 1) To NrRows

'Check S/N limit
        If Cells(O, SNColumn).Value < SNLimit Then GoTo NextO

'Check whether m/z value lies within ppm window (upper half)
        If (Cells(O, MZColumn) - Cells(Z, MZColumn)) > (MZWindow * Cells(Z,
MZColumn).Value / 2) Then Exit For
        Tag = Sheets(RawSheet).Cells(O, TagColumn).Value
        For i = 2 To NrColumnsPivot
            Select Case Tag
                Case Pivot(i)
                    PivotColumn = i
                    Exit For
            End Select
        Next i
        Sheets(PivotSheet).Cells(PivotRow, PivotColumn).Value = 1 +
Sheets(PivotSheet).Cells(PivotRow, PivotColumn).Value
NextO:
    Next O

'Analyze ppm window from center to lower m/z limit
    For U = 1 To (Z - StartRow)

'Check S/N limit
        If Cells(Z - U, SNColumn).Value < SNLimit Then GoTo NextU
'Check whether m/z value lies within ppm window (lower half)
        If (Cells(Z, MZColumn) - Cells((Z - U), MZColumn)) > (MZWindow * Cells(Z,
MZColumn).Value / 2) Then Exit For
        Tag = Sheets(RawSheet).Cells((Z - U), TagColumn).Value
        For i = 2 To NrColumnsPivot
            Select Case Tag
                Case Pivot(i)
                    PivotColumn = i
                    Exit For
            End Select
        Next i
        Sheets(PivotSheet).Cells(PivotRow, PivotColumn).Value = 1 +
Sheets(PivotSheet).Cells(PivotRow, PivotColumn).Value
NextU:
    Next U

NextZ:
      Next Z
```

Fig. 11.2 Extract of the Excel macro code used for determination of subspecies-specific peptide masses as applied to process raw data from Gekenidis et al. (2014)

the other subspecies in order to identify the amino acid exchanges responsible for the uniqueness of each biomarker peptide.

Identification of Antibiotic Resistance Mechanisms in Bacteria Using Tryptic Peptides

Since their discovery, antibiotics have been extensively used to fight bacterial infections. This broad use has led to a drastic increase in the occurrence of antibiotic-resistant bacteria, representing one of the major current threats to human health. Consequently, there is a need for discovery of novel antibiotics or drug targets. Identification of the pathways involved in resistance and understanding the underlying molecular mechanisms are important steps toward fulfilling this task.

A recent application of MALDI-TOF MS focuses on the elucidation of resistance mechanisms in bacteria, involving mainly resistance to antibiotics. An extensive recent review on the application of MALDI-TOF MS for detection of antibiotic resistance mechanisms has been published by Hrabak et al. 2013. In the following we shall focus on studies using gel electrophoresis followed by tryptic protein digestion to (a) elucidate changes in expression profiles associated with exposure of microorganisms to antibiotics, (b) identify the proteins being up- or down-regulated as a consequence of antibiotic exposure, and (c) deduce mechanisms involved in bacterial resistance. The vast majority of these studies rely on a principle described as early as 1996 by Shevchenko for in-gel tryptic digestion and mass spectrometric sequencing of proteins (Shevchenko et al. 1996).

Decreased membrane permeability is one major mechanism providing antibiotic resistance to bacteria. In 2001, Dé and coworkers investigated the role of the major porin in *Enterobacter aerogenes* for its resistance to cephalosporins (Dé et al. 2001). They first purified the porin and observed a mass difference in the wild-type and resistant strain porins by MALDI-TOF MS which they hypothesized to rely on a difference in the primary sequence. The SDS-PAGE protein bands of the two proteins were then digested in-gel with trypsin and peptide mapping by MALDI-TOF and nanospray MS/MS identified a G to D mutation in one of the porin's loops, which was suggested to be conferring antibiotic resistance to the clinical *E. aerogenes* strain.

Conejo et al. reported the loss of the outer membrane porin protein D (OprD) in *Pseudomonas aeruginosa* in response to zinc eluting from siliconized latex urinary catheters (SLUC) resulting in carbapenem resistance (Conejo et al. 2003). The outer membrane proteins of *P. aeruginosa* grown in the presence and absence of zinc were prepared, and the expression profiles were compared after separation by SDS-PAGE. Further analysis by MALDI-TOF MS after in-gel tryptic digestion of outer membrane proteins not expressed in the zinc-supplemented extract (and the SLUC eluate) revealed that they matched OprD. The authors concluded that the loss of OprD from *P. aeruginosa* in the presence of zinc is the underlying mechanism for the previously reported increased resistance of this bacterium to imipenem, an antibiotic belonging to the class of carbapenems.

Recently, Khatua et al. elucidated a novel mechanism for how sialic acids on OprD might confer β-lactam antibiotic resistance to *P. aeruginosa* (Khatua et al. 2014). Strains containing α2,3- and α2,6-linked sialic acids have previously been shown to have increased resistance to β-lactam antibiotics. Therefore, after purifying sialoglycoproteins from the membrane fractions of four clinical *P. aeruginosa* isolates and separating them by 2D gel electrophoresis, Khatua et al. digested those proteins in-gel with trypsin and analyzed the resulting peptides by MALDI-TOF/TOF MS. Sialoglycoproteins containing either α2,3-, α2,6-linkages, or both could be identified, among others an OprD precursor. In a subsequent step, sialylated OprD proteins were purified by anion exchange chromatography, and their identity was confirmed by trypsin digestion and MALDI-TOF/TOF MS. Further experiments led the authors to the conclusion that sialic acids on the OprD protein hampered its interaction with β-lactam antibiotics, probably thereby increasing the survival of such strains under antibiotic pressure.

Another study on *P. aeruginosa* (Peng et al. 2005) examined the sarcosine-insoluble outer membrane fraction upon treatment with ampicillin, kanamycin, and tetracycline to identify proteins related to the respective antibiotic resistances. The authors found 11 differential proteins, which were excised from the 2D gel and identified by MALDI-TOF MS after in-gel tryptic digestion of the excised spots. Apart from some known antibiotic resistance proteins, Peng et al. discovered some new proteins and thereby novel potential antibiotic targets.

The same technique of in-gel tryptic digestion and subsequent identification by MALDI-TOF MS was applied by Dupont and coworkers in *Acinetobacter baumannii,* an opportunistic bacillus comprising increasing numbers of resistant strains (Dupont et al. 2005). They compared the outer membrane of different strains and found two differentially expressed proteins, one of which was identified as belonging to the OprD family.

A final example demonstrating the importance of membrane permeability as a mechanism of antibiotic resistance is a study from 2006 which investigated the response of outer membrane proteins in *E. coli* to tetracycline and ampicillin (Xu et al. 2006). Three known and six new outer membrane proteins related to antibiotic resistance in *E. coli* K-12 could be identified.

Another set of studies explored changes occurring in the overall proteome in response to antibiotic treatments (Cordwell et al. 2002). Cordwell et al. used 2D gel electrophoresis to compare the protein profiles of an MSSA and an MRSA strain. A total of 377 proteins were analyzed by MALDI-TOF MS following tryptic digestion of gel-purified proteins. Proteins which could not be identified by MALDI were subjected to tandem electrospray ionization (ESI) MS. In addition, the effect of Triton X-100, a detergent known to reduce methicillin resistance, was investigated. Here, 44 proteins showed altered abundance on the 2D gel with 11 spots found exclusively in the resistant strain. Based on these findings, the authors could conclude that among other factors, products of the σ^B and the SarA regulon (the alternative sigma factor and a regulator of virulence genes) are involved in methicillin resistance of *S. aureus*.

Another study by Cho et al. on MRSA investigated the effect of tea polyphenols (TPP) on the protein expression of a clinical MRSA isolate displaying an excellent synergistic effect of TPP and oxacillin (Cho et al. 2008). Down-regulation of 14 extracellular proteins (chaperone-like and other proteins related to cellular pathogenicity mechanisms as identified by MALDI fingerprinting) and up-regulation of 3 proteins upon TPP exposure were observed. Although the underlying mechanism for this synergy of TPP and oxacillin could not be elucidated, the findings show a clear effect of TPP on the expression of several key MRSA proteins.

Eyraud et al. could show how a small regulatory RNA, SprX, influences antibiotic resistance of *S. aureus* to two glycopeptides, vancomycin and teicoplanin, which are the antibiotics of choice to treat MRSA infections (Eyraud et al. 2014). By constructing a mutant strain lacking expression of SprX (Δ*sprX*) and comparing its expression profile with the wild-type, the authors could identify a SprX target, stage V sporulation protein G, SpoVG, which is significantly down-regulated in the presence of SprX. Of note, SpoVG has been suggested previously to fulfill more general regulatory functions unrelated to sporulation in nonsporulating bacteria such as *S. aureus* (Meier et al. 2007; Schulthess et al. 2011).

Other studies have conducted 2D gel electrophoresis and tryptic peptide-based proteomic surveys on *Mycobacterium tuberculosis*. Sharma et al. analyzed whole-cell extracts of streptomycin-susceptible and streptomycin-resistant clinical isolates of *M. tuberculosis* (Sharma et al. 2010). In 2013, Kumar and coworkers could identify 12 proteins consistently up-regulated in resistant isolates (Kumar et al. 2013). Finally, Truong et al. published results on expression changes related to proteins associated with resistance to rifampicin and isoniazid (RH), the key drugs for tuberculosis treatment (Truong et al. 2014). A comparison of the proteome extracted from RH-resistant and RH-susceptible clinical isolates after 2D gel electrophoresis separation yielded 41 spots with differential expression. After identification of the corresponding proteins by MALDI-TOF/TOF MS analysis of the generated tryptic peptides, 12 proteins involved in virulence, adaptation, and lipid metabolism were identified.

Recent investigations on a cefotaxime-resistant *E. coli* strain WA57 (producing extended-spectrum β-lactamase) revealed 40 differentially expressed proteins from different cell compartments (extracellular, periplasmic, cytoplasmic, membrane, and whole-cell) upon exposure to cefotaxime (Gonçalves et al. 2014). These 40 and additional 275 proteins were all identified by analyzing tryptic protein digests with MALDI-TOF/TOF MS. This study gives a comprehensive overview of the changes occurring in the *E. coli* strain WA57 when stressed with cefotaxime. Chaperone, porin, and export proteins were particularly affected, suggesting an important role of stress response and transport functions in antibiotic resistance of this strain.

Another important principle underlying antibiotic resistance is the inactivation of the agent by either chemically modifying it (e.g., hydrolysis) or directly binding to it (antibiotic trapping) (Goessens et al. 2013). Goessens et al. hypothesized a covalent binding of meropenem to an enzyme in *E. coli* as an underlying resistance mechanism toward carbapenems. A comparison of a carbapenem-susceptible *E. coli* to its carbapenem-resistant successor strain isolated from the same patient

after carbapenem treatment showed that the resistant strain additionally possessed the plasmid-encoded β-lactamase CMY-2. In order to confirm their hypothesis of an acyl–enzyme complex formation, they incubated periplasmic extracts with meropenem, separated those as well as untreated extracts on SDS–PAGE, and analyzed the excised CMY-2 band on a MALDI-TOF/TOF MS after in-gel digestion with trypsin. By comparing the tryptic peptides in which the active site of the enzyme is located from treated and untreated samples, they found the peptide mass corresponding to the peptide containing the active site after modification with meropenem and removal of an acetaldehyde group in the samples treated with meropenem. This finding strongly supports the hypothesis of meropenem being covalently bound to CMY-2 as a possible antibiotic resistance mechanism.

A recent study from 2014 exploited the analysis of in-gel trypsin-digested proteins with MALDI-TOF/TOF MS to identify CMY-2-type cephalosporinases in Enterobacteriaceae (Papagiannitsis et al. 2014). A peak uniquely observed in CMY producing isolates was thereby confirmed to represent a *C. freundii*-like β-lactamase.

In a comparative proteome study with a multi-resistant *E. coli,* 21 differentially expressed proteins under treatment with multiple drugs were identified (Piras et al. 2012). From the identified proteins, the authors concluded that quorum sensing might be involved in the multiple antibiotic resistance observed in this strain.

Hemmerlin et al. investigated by applying 2D gel electrophoresis, tryptic digestion, and MALDI-TOF MS the effect of fosmidomycin on *E. coli* being only shortly exposed to the antibiotic (Hemmerlin et al. 2014). Within the first 3 h after exposure, combined strategies are triggered mainly consisting of adapting metabolism to increase tolerance to oxidative stress and rapidly exporting the antibiotic from the cell. Such insights can aid the development of new efficient drugs by improving the understanding of the underlying defense mechanisms.

Similar studies can be conducted using mass spectrometric methods other than MALDI-TOF. However, being beyond the scope of this brief overview, we shall mention only one study from 2007 by Camara and Hays (2007). A protein with an approximate mass of 29,000 Da found only in ampicillin-resistant *E. coli* was confirmed to be a β-lactamase by in-gel digestion followed by liquid chromatography-mass spectrometry (LC-MS).

Another approach, circumventing the time-consuming preparation of 2D gels, was shown by Wilcox et al. (2001). Instead of digesting protein bands separated on a gel, they analyzed tryptic digests of fractions collected from an HPLC. In particular, three ribosomal proteins responsible for streptomycin, erythromycin, and spectinomycin resistance in three *E. coli* strains were investigated. The mutations responsible for the observed resistance were located by analyzing tryptic peptides on a MALDI-TOF/TOF and a nano-electrospray tandem mass spectrometer.

In conclusion, tryptic digestion of proteins and analysis of the resulting peptides by MALDI-TOF MS or LC-MS have proven to be a potent tool to elucidate mechanisms underlying bacterial resistance to antibiotics. Either proteins of interest or protein fractions such as outer membrane proteins can be analyzed, or the whole proteome of antibiotic-susceptible and -resistant strains can be compared. The proteins of interest or the differentially expressed proteins can be digested to

tryptic peptides and then further analyzed. New proteins related to resistance are thereby identified, and mutations responsible for resistance can be located. Resistance mechanisms are usually deduced from the function of the identified proteins. In the case of antibiotic trapping as the underlying mechanism, that is, the covalent binding of the antibiotic to a target, the actual antibiotic–target complex can be detected. The number of analyzed proteins going up to several hundreds and the broad spectrum of antibiotics and bacteria tested in the aforementioned studies show the global applicability of these approaches.

Limitations and Future Perspectives

Intact protein expression typing of pathogenic bacteria has been continuously improved since in the mid-eighties MALDI-TOF MS and ESI MS have been established as efficient soft-ionization techniques of biomolecules. For rapid identification and classification, the intact protein MALDI-TOF MS approach has to be favored as compared to the ESI MS technique with respect to reproducibility, speed, and robustness of data acquisition, and cost-effectiveness. However, in cases where time to result and increased complexity of analytical workflows and infrastructures are not of primary concern, state-of-the-art technologies aiming at profiling bacteria according to their peptide profiles (proteomics-based approach) give preference clearly to the ESI MS method. Even in the light of the many achievements and advantages of whole-cell mass spectrometry (biotyping), there exist currently still limitations and areas that need to be improved in the future, such as the following:

- Accurate identification of strains (below-species level) for the precise definition of organisms and communities under investigation
- Identification of single or multiple bacteria in bacterial mixtures, opening up a wide range of possibilities to investigate diversity at the biologically relevant level
- Identification without *in vitro* culturing, directly from environmental samples contaminated by pathogenic bacteria or identification of non-culturable bacteria (non-culture-based identification)
- Set-up and publish lists of organism-specific biomarkers (proteins or peptides) as a community resource for quick and reliable identification
- Targeted proteomics approaches using selected reaction monitoring (SRM) methods for increased accuracy and sensitivity in quantification
- Merging proteomic and genomic databases, using so-called proteogenomics approaches allowing to identify and characterize previously undescribed species and proteins
- Enrichment of bacteria applying bead-based technology prior to identification to increase analytical sensitivity, particularly from low abundant specimens

Direct and rapid identification of bacteria from environmental samples and mixtures by MALDI-TOF MS remains challenging. Target bacteria may exist in low

concentrations (below the analytical detection limit) and background from the sample matrix may influence subsequent analyses. As an approach to separate and enrich target bacteria without culturing using standard microbiological enrichment procedures, functionalized magnetic nano-beads, for example, coated with antibodies, are promising for preparation of such samples prior to MALDI-TOF MS analysis (Ho et al. 2004; Schlosser et al. 2007; Madonna et al. 2001). Aiming to implement such approaches in routine diagnostics, future work needs to emphasize increasing cell recovery rates as well as overcoming cross-reaction with non-target bacteria and agglutination of beads. In particular, success in these fields will strongly depend on the development and application of new specific affinity probes.

In order to increase the discriminatory power of MS-based methods of bacterial characterization, we have been investigating into a discovery-based proteomics approach making use of the traditional organic acid/organic solvent extracted biotyping sample (Gekenidis et al. 2014; Drissner et al. 2014). As commented in the sections above, we could clearly demonstrate that this ready-to-use, straightforward preparation—free of any contaminating ingredients, such as detergents—is proving to be a very valuable starting material to perform proteomics experiments using LC-MALDI MS. We have now compared the equivalent tryptic digests with LC-ESI MS and found that, by combining both ionization techniques, the information content regarding the produced peptides significantly increased (unpublished results). The ultimate aim of generating such biomarker peptide lists is to switch from discovery-based proteomics into target-based proteomics allowing in the future to monitor pathogenic bacteria with increased discriminatory power using SRM technology. In brief, the mass spectrometer (triple quadrupole) would be set to monitor only selected, microorganism-typic tryptic peptides, and their absolute abundance could be determined by spiking respective isotopically labeled peptides. Until now, there are only a few reports regarding SRM applications in the field of SRM technology of pathogenic bacteria. Aebersold and Picotti have recently paved the way to accurately employ SRM technology within proteomics approaches (Karlsson et al. 2012; Picotti and Aebersold 2012).

In the context of truly real-time detection, it is essential that, as pre-analytical steps, bacterial sample concentration and elimination from the matrix need to be improved and simplified. As recently shown by Barreiro et al. (2012), non-culture-based identification of bacteria in milk by protein fingerprinting is easily performed by using the recently introduced Sepsityper™ by Bruker.

The problem of identifying non-culturable bacteria and the possibility of accurately identifying individual microorganisms out of a mixture have been addressed so far mainly through genomics-oriented projects, for example, by next-generation sequencing applications and with proteogenomics approaches (Pierce et al. 2012; Woo et al. 2014; Lasken and McLean 2014; Sheynkman et al. 2014). The latter one is an area of research interfacing proteomics and genomics and as such helping to identify novel peptides (not present in reference protein sequence databases) from MS-based proteomics data (Nesvizhskii 2014).

References

Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422:198–207.

Albesharat R, Ehrmann MA, Korakli M, Yazaji S, Vogel RF. Phenotypic and genotypic analyses of lactic acid bacteria in local fermented food, breast milk and faeces of mothers and their babies. Syst Appl Microbiol. 2011;34:148–55.

Arnold RJ, Reilly JP. Fingerprint matching of *E. coli* strains with matrix-assisted laser desorption ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. Rapid Commun Mass Spectrom. 1998;12:630–6.

Ayyadurai S, Flaudrops C, Raoult D, Drancourt M. Rapid identification and typing of *Yersinia pestis* and other *Yersinia* species by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. BMC Microbiol. 2010;10:285.

Barbuddhe SB, Maier T, Schwarz G, Kostrzewa M, Hof H, Domann E, Chakraborty T, Hain T. Rapid identification and typing of *Listeria* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2008;74:5402–7.

Barreiro JR, Braga PAC, Ferreira CR, Kostrzewa M, Maier T, Wegemann B, Boettcher V, Eberlin MN, dos Santos MV. Nonculture-based identification of bacteria in milk by protein fingerprinting. Proteomics. 2012;12:2739–45.

Bernardo K, Pakulat N, Macht M, Krut O, Seifert H, Fleer S, Huenger F, Kroenke M. Identification and discrimination of *Staphylococcus aureus* strains using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. Proteomics. 2002;2:747–53.

Bright JJ, Claydon MA, Soufian M, Gordon DB. Rapid typing of bacteria using matrix-assisted laser desorption ionisation time-of-flight mass spectrometry and pattern recognition software. J Microbiol Methods. 2002;48:127–38.

Camara JE, Hays FA. Discrimination between wild-type and ampicillin-resistant *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Bioanal Chem. 2007;389:1633–8.

Carbonnelle E, Beretti JL, Cottyn S, Quesne G, Berche P, Nassif X, Ferroni A. Rapid identification of staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization—time of flight mass spectrometry. J Clin Microbiol. 2007;45:2156–61.

Castanha ER, Vestal M, Hattan S, Fox A, Fox KF, Dickinson D. *Bacillus cereus* strains fall into two clusters (one closely and one more distantly related) to *Bacillus anthracis* according to amino acid substitutions in small acid-soluble proteins as determined by tandem mass spectrometry. Mol Cell Probes. 2007;21:190–201.

Chen P, Lu Y, Harrington PB. Biomarker profiling and reproducibility study of MALDI-MS measurements of *Escherichia coli* by analysis of variance-principal component analysis. Anal Chem. 2008;80:1474–81.

Cho YS, Schiller NL, Oh KH. Antibacterial effects of green tea polyphenols on clinical isolates of methicillin-resistant *Staphylococcus aureus*. Curr Microbiol. 2008;57:542–6.

Claydon MA, Davey SN, Edwards-Jones V, Gordon DB. The rapid identification of intact microorganisms using mass spectrometry. Nat Biotechnol. 1996;14:1584–6.

Conejo MC, Garcia I, Martinez-Martinez L, Picabea L, Pascual A. Zinc eluted from siliconized latex urinary catheters decreases OprD expression, causing carbapenem resistance in *Pseudomonas aeruginosa*. Antimicrob Agents Ch. 2003;47:2313–5.

Cordwell SJ, Larsen MR, Cole RT, Walsh BJ. Comparative proteomics of *Staphylococcus aureus* and the response of methicillin-resistant and methicillin-sensitive strains to Triton X-100. Microbiology. 2002;148:2765–81.

Dé E, Baslé A, Jaquinod M, Saint N, Malléa M, Molle G, Pagés JM. A new mechanism of antibiotic resistance in Enterobacteriaceae induced by a structural modification of the major porin. Mol Microbiol. 2001;41:189–98.

Demirev PA, Feldman AB, Kowalski P, Lin JS. Top-down proteomics for rapid identification of intact microorganisms. Anal Chem. 2005;77:7455–61.

Dickinson DN, Duc MT L, Haskins WE, Gornushkin I, Winefordner JD, Powell DH, Venkateswaran K. Species differentiation of a diverse suite of *Bacillus* spores by mass spectrometry-based protein profiling. Appl Environ Microbiol. 2004;70:475–82.

Dieckmann R, Helmuth R, Erhard M, Malorny B. Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2008;74:7767–78.

Donohue MJ, Smallwood AW, Pfaller S, Rodgers M, Shoemaker JA. The development of a matrix-assisted laser desorption/ionization mass spectrometry-based method for the protein fingerprinting and identification of *Aeromonas* species using whole cells. J Microbiol Methods. 2006;65:380–9.

Drissner D, Gekenidis MT, Schlapbach R, Brunisholz R. When time-to-result matters: identification of microbes based on MALDI-TOF protein and peptide profiling. Chimia. 2014;68:453–3.

Dubois D, Leyssene D, Chacornac JP, Kostrzewa M, Schmit PO, Talon R, Bonnet R, Delmas J. Identification of a variety of *Staphylococcus* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2010;48:941–5.

Dupont M, Pages JM, Lafitte D, Siroy A, Bollet C. Identification of an OprD homologue in *Acinetobacter baumannii*. J Proteome Res. 2005;4:2386–90.

Edwards-Jones V, Claydon MA, Evason DJ, Walker J, Fox AJ, Gordon DB. Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. J Med Microbiol. 2000;49:295–300.

Elhanany E, Barak R, Fisher M, Kobiler D, Altboum Z. Detection of specific *Bacillus anthracis* spore biomarkers by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2001;15:2110–6.

Erhard M, von Doehren H, Jungblut P. Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry. Nat Biotechnol. 1997;15:906–9.

Evason DJ, Claydon MA, Gordon DB. Exploring the limits of bacterial identification by intact cell-mass spectrometry. J Am Soc Mass Spectrom. 2001;12:49–54.

Everley RA, Mott TM, Wyatt SA, Toney DM, Croley TR. Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species. J Am Soc Mass Spectrom. 2008;19:1621–8.

Eyraud A, Tattevin P, Chabelskaya S, Felden B. A small RNA controls a protein regulator involved in antibiotic resistance in *Staphylococcus aureus*. Nucleic Acids Res. 2014;42:4892–905.

Fagerquist CK, Sultan O. Induction and identification of disulfide-intact and disulfide-reduced beta-subunit of Shiga toxin 2 from *Escherichia coli* O157:H7 using MALDI-TOF-TOF-MS/MS and top-down proteomics. Analyst. 2011;136:1739–46.

Fagerquist CK, Miller WG, Harden LA, Bates AH, Vensel WH, Wang G, Mandrell RE. Genomic and proteomic identification of a DNA-binding protein used in the "fingerprinting" of *Campylobacter* species and strains by MALDI-TOF-MS protein biomarker analysis. Anal Chem. 2005;77:4897–907.

Fagerquist CK, Bates AH, Heath S, King BC, Garbus BR, Harden LA, Miller WG. Sub-speciating *Campylobacter jejuni* by proteomic analysis of its protein biomarkers and their post-translational modifications. J Proteome Res. 2006;5:2527–38.

Fagerquist CK, Garbus BR, Miller WG, Williams KE, Yee E, Bates AH, Boyle S, Harden LA, Cooley MB, Mandrell RE. Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. Anal Chem. 2010;82:2717–25.

Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev. 2001;20:157–71.

Fujinami Y, Kikkawa HS, Kurosaki Y, Sakurada K, Yoshino M, Yasuda J. Rapid discrimination of *Legionella* by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Microbiol Res. 2011;166:77–86.

Gekenidis MT, Studer P, Wuethrich 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. 2014;80:4234–41.

Ghyselinck J, Van Hoorde K, Hoste B, Heylen K, De Vos P. Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. J Microbiol Methods. 2011;86:327–36.

Giebel RA, Fredenberg W, Sandrin TR. Characterization of environmental isolates of *Enterococcus* spp. by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Water Res. 2008;42:931–40.

Goessens WHF, van der Bij AK, van Boxtel R, Pitout JDD, van Ulsen P, Melles DC, Tommassen J. Antibiotic trapping by plasmid-encoded CMY-2 β-lactamase combined with reduced outer membrane permeability as a mechanism of carbapenem resistance in *Escherichia coli*. Antimicrob Agents Chemother. 2013;57:3941–9.

Gonçalves A, Poeta P, Monteiro R, Marinho C, Silva N, Guerra A, Petrucci-Fonseca F, Rodrigues J, Torres C, Vitorino R, Domingues P, Igrejas G. Comparative proteomics of an extended spectrum β-lactamase producing *Escherichia coli* strain from the Iberian wolf. J Proteomics. 2014;104:80–93.

Grosse-Herrenthey A, Maier T, Gessler F, Schaumann R, Boehnel H, Kostrzewa M, Krueger M. Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Anaerobe. 2008;14:242–9.

Haag AM, Taylor SN, Johnston KH, Cole RB. Rapid identification and speciation of *Haemophilus* bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Mass Spectrom. 1998;33:750–6.

Hemmerlin A, Tritsch D, Hammann P, Rohmer M, Bach TJ. Profiling of defense responses in *Escherichia coli* treated with fosmidomycin. Biochimie. 2014;99C:54–62.

Hettick JM, Kashon ML, Slaven JE, Ma Y, Simpson JP, Siegel PD, Mazurek GN, Weissman DN. Discrimination of intact mycobacteria at the strain level: a combined MALDI-TOF MS and biostatistical analysis. Proteomics. 2006;6:6416–25.

Ho KC, Tsai PJ, Lin YS, Chen YC. Using biofunctionalized nanoparticles to probe pathogenic bacteria. Anal Chem. 2004;76:7162–8.

Honisch C, Chen Y, Mortimer C, Arnold C, Schmidt O, van den Boom D, Cantor CR, Shah HN, Gharbia SE. Automated comparative sequence analysis by base-specific cleavage and mass spectrometry for nucleic acid-based microbial typing. Proc Natl Acad Sci U S A. 2007;104:10649–54.

Horneffer V, Haverkamp J, Janssen HG, Notz R. MALDI-TOF-MS analysis of bacterial spores: wet heat-treatment as a new releasing technique for biomarkers and the influence of different experimental parameters and microbiological handling. J Am Soc Mass Spectrom. 2004;15:1444–54.

Hotta Y, Teramoto K, Sato H, Yoshikawa H, Hosoda A, Tamura H. Classification of genus *Pseudomonas* by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* operon at strain level. J Proteome Res. 2010;9:6722–8.

Hrabak J, Chudackova E, Walkova R. Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. Clin Microbiol Rev. 2013;26:103–14.

Hsieh SY, Tseng CL, Lee YS, Kuo AJ, Sun CF, Lin YH, Chen JK. Highly efficient classification and identification of human pathogenic bacteria by MALDI-TOF MS. Mol Cell Proteomics. 2008;7:448–56.

Ilina EN, Borovskaya AD, Serebryakova MV, Chelysheva VV, Momynaliev KT, Maier T, Kostrzewa M, Govorun VM. Application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for the study of *Helicobacter pylori*. Rapid Commun Mass Spectrom. 2010;24:328–34.

Jackson GW, McNichols RJ, Fox GE, Willson RC. Universal bacterial identification by mass spectrometry of 16 S ribosomal RNA cleavage products. Int J Mass Spectrom. 2007;261:218–26.

Jackson KA, Edwards-Jones V, Sutton CW, Fox AJ. Optimisation of intact cell MALDI method for fingerprinting of methicillin-resistant *Staphylococcus aureus*. J Microbiol Methods. 2005;62:273–84.

Josten M, Reif M, Szekat C, Al-Sabti N, Roemer T, Sparbier K, Kostrzewa M, Rohde H, Sahl HG, Bierbaum G. Analysis of the matrix-assisted laser desorption ionization-time of flight mass

spectrum of *Staphylococcus aureus* identifies mutations that allow differentiation of the main clonal lineages. J Clin Microbiol. 2013;51:1809–17.
Kallow W, Dieckmann R, Kleinkauf N, Erhard M, Neuhof T. Method of identifying microorganisms using MALDI-TOF-MS. European Patent. 2000.
Karger A, Ziller M, Bettin B, Mintel B, Schares S, Geue L. Determination of serotypes of shiga toxin-producing *Escherichia coli* isolates by intact cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2011;77:896–905.
Karlsson C, Malmstroem L, Aebersold R, Malmstroem J. Proteome-wide selected reaction monitoring assays for the human pathogen *Streptococcus pyogenes*. Nat Commun. 2012;3:1301.
Keys CJ, Dare DJ, Sutton H, Wells G, Lunt M, McKenna T, McDowall M, Shah HN. Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. Infect Genet E. 2004;4:221–42.
Khatua B, Van Vleet J, Choudhury BP, Chaudhry R, Mandal C. Sialylation of outer membrane porin protein D: a mechanistic basis of antibiotic uptake in *Pseudomonas aeruginosa*. Mol Cell Proteomics. 2014;13:1412–28.
Kraushaar B, Dieckmann R, Wittwer M, Knabner D, Konietzny A, Maede D, Strauch E. Characterization of a *Yersinia enterocolitica* biotype 1A strain harbouring an *ail* gene. J Appl Microbiol. 2011;111:997–1005.
Krishnamurthy T, Ross PL, Rajamani U. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1996;10:883–8.
Kumar B, Sharma D, Sharma P, Katoch VM, Venkatesan K, Bisht D. Proteomic analysis of *Mycobacterium tuberculosis* isolates resistant to kanamycin and amikacin. J Proteomics. 2013;94:68–77.
Lartigue MF, Hery-Arnaud G, Haguenoer E, Domelier AS, Schmit PO, van der Mee-Marquet N, Lanotte P, Mereghetti L, Kostrzewa M, Quentin R. Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2009;47:2284–7.
Lasken RS, McLean JS. Recent advances in genomic DNA sequencing of microbial species from single cells. Nat Rev Genet. 2014;15:577–84.
Leuschner RGK, Beresford-Jones N, Robinson C. Difference and consensus of whole cell *Salmonella enterica* subsp. *enterica* serovars matrix-assisted laser desorption/ionization time-of-flight mass spectrometry spectra. Lett Appl Microbiol. 2003;38:24–31.
Lill JR, Ingle ES, Liu PS, Pham V, Sandoval WN. Microwave-assisted proteomics. Mass Spectrom Rev. 2007;26:657–71.
Lundquist M, Caspersen MB, Wikstroem P, Forsman M. Discrimination of *Francisella tularensis* subspecies using surface enhanced laser desorption ionization mass spectrometry and multivariate data analysis. FEMS Microbiol Lett. 2005;243:303–10.
Lynn EC, Chung MC, Tsai WC, Han CC. Identification of Enterobacteriaceae bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. Rapid Commun Mass Spectrom. 1999;13:2022–7.
Madonna AJ, Basile F, Furlong E, Voorhees KJ. Detection of bacteria from biological mixtures using immunomagnetic separation combined with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2001;15:1068–74.
Majcherczyk PA, McKenna T, Moreillon P, Vaudaux P. The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*. FEMS Microbiol Lett. 2006;255:233–9.
Mandrell RE, Harden LA, Bates A, Miller WG, Haddon WF, Fagerquist CK. Speciation of *Campylobacter coli, C. jejuni, C. helveticus, C. lari, C. sputorum,* and *C. upsaliensis* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2005;71:6292–307.
McLuckey SA. The emerging role of ion/ion reactions in biological mass spectrometry: considerations for reagent ion selection. Eur J Mass Spectrom. 2010;16:429–36.

Meier S, Goerke C, Wolz C, Seidl K, Homerova D, Schulthess B, Kormanec J, Berger-Baechi B, Bischoff M. σ^B and the σ^B-dependent *arlRS* and *yabJ-spoVG* loci affect capsule formation in *Staphylococcus aureus*. Infect Immun. 2007;75:4562–71.

Mott TM, Everley RA, Wyatt SA, Toney DM, Croley TR. Comparison of MALDI-TOF/MS and LC-QTOF/MS methods for the identification of enteric bacteria. Int J Mass Spectrom. 2010;291:24–32.

Moura H, Woolfitt AR, Carvalho MG, Pavlopoulos A, Teixeira LM, Satten GA, Barr JR. MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. FEMS Immunol Med Microbiol. 2008;53:333–42.

Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. Nat Methods. 2014;11:1114–25.

Nilsson CL. Fingerprinting of *Helicobacter pylori* strains by matrix-assisted laser desorption/ionization mass spectrometric analysis. Rapid Commun Mass Spectrom. 1999;13:1067–71.

Ochoa ML, Harrington PB. Immunomagnetic isolation of enterohemorrhagic *Escherichia coli* O157:H7 from ground beef and identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and database searches. Anal Chem. 2005;77:5258–67.

Papagiannitsis CC, Kotsakis SD, Tuma Z, Gniadkowski M, Miriagou V, Hrabak J. Identification of CMY-2-type cephalosporinases in clinical isolates of Enterobacteriaceae by MALDI-TOF MS. Antimicrob Agents Chemother. 2014;58:2952–7.

Peng X, Xu C, Ren H, Lin X, Wu L, Wang S. Proteomic analysis of the sarcosine-insoluble outer membrane fraction of *Pseudomonas aeruginosa* responding to ampicilin, kanamycin, and tetracycline resistance. J Proteome Res. 2005;4:2257–65.

Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods. 2012;9:555–66.

Pierce CY, Barr JR, Woolfitt AR, Moura H, Shaw EI, Thompson HA, Massung RF, Fernandez FM. Strain and phase identification of the U.S. category B agent *Coxiella burnetii* by matrix assisted laser desorption/ionization time-of-flight mass spectrometry and multivariate pattern recognition. Anal Chim Acta. 2007;583:23–31.

Pierce SE, Bell RL, Hellberg RS, Cheng CM, Chen KS, Williams-Hill DM, Martin WB, Allard MW. Detection and identification of *Salmonella enterica*, *Escherichia coli*, and *Shigella* spp. via PCR-electrospray ionization mass spectrometry: isolate testing and analysis of food samples. Appl Environ Microbiol. 2012;78:8403–11.

Piras C, Soggiu A, Bonizzi L, Gaviraghi A, Deriu F, De Martino L, Iovane G, Amoresano A, Roncada P. Comparative proteomics to evaluate multi drug resistance in *Escherichia coli*. Mol Biosyst. 2012;8:1060–7.

Price NPJ, Rooney AP, Swezey JL, Perry E, Cohan FM. Mass spectrometric analysis of lipopeptides from *Bacillus* strains isolated from diverse geographical locations. FEMS Microbiol Lett. 2007;271:83–9.

Rajakaruna L, Hallas G, Molenaar L, Dare D, Sutton H, Encheva V, Culak R, Innes I, Ball G, Sefton AM, Eydmann M, Kearns AM, Shah HN. High throughput identification of clinical isolates of *Staphylococcus aureus* using MALDI-TOF-MS of intact cells. Infect Genet E. 2009;9:507–13.

Ruelle V, Moualij B E, Zorzi W, Ledent P, De Pauw E. Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2004;18:2013–9.

Rupf S, Breitung K, Schellenberger W, Merte K, Kneist S, Eschrich K. Differentiation of mutans streptococci by intact cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Oral Microbiol Immun. 2005;20:267–73.

Ryzhov V, Hathout Y, Fenselau C. Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl Environ Microbiol. 2000;66:3828–34.

Sandrin TR, Goldstein JE, Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom Rev. 2013;32:188–217.

Sauer S, Kliem M. Mass spectrometry tools for the classification and identification of bacteria. Nat Rev Microbiol. 2010;8:74–82.

Sauer S, Freiwald A, Maier T, Kube M, Reinhardt R, Kostrzewa M, Geider K. Classification and identification of bacteria by mass spectrometry and computational analysis. PLOS ONE. 2008;3:e2843.

Schaller A, Troller R, Molina D, Gallati S, Aebi C, Stutzmann Meier P. Rapid typing of *Moraxella catarrhalis* subpopulations based on outer membrane proteins using mass spectrometry. Proteomics. 2006;6:172–80.

Schlosser G, Kacer P, Kuzma M, Szilagyi Z, Sorrentino A, Manzo C, Pizzano R, Malorni L, Pocsfalvi G. Coupling immunomagnetic separation on magnetic beads with matrix-assisted laser desorption ionization-time of flight mass spectrometry for detection of staphylococcal enterotoxin B. Appl Environ Microbiol. 2007;73:6945–52.

Schmid O, Ball G, Lancashire L, Culak R, Shah H. New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionization time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. J Med Microbiol. 2005;54:1205–11.

Schulthess B, Bloes DA, Francois P, Girard M, Schrenzel J, Bischoff M, Berger-Baechi B. The σ^B-dependent *yabJ-spoVG* operon is involved in the regulation of extracellular nuclease, lipase, and protease expression in *Staphylococcus aureus*. J Bacteriol. 2011;193:4954–62.

Seibold E, Bogumil R, Vorderwuelbecke S, Dahouk S A, Buckendahl A, Tomaso H, Splettstoesser W. Optimized application of surface-enhanced laser desorption/ionization time-of-flight MS to differentiate *Francisella tularensis* at the level of subspecies and individual strains. FEMS Immunol Med Microbiol. 2007;49:364–73.

Seng P, Drancourt M, Gouriet F, Scola B L, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49:543–51.

Shah HN, Rajakaruna L, Ball G, Misra R, Al-Shahib A, Fang M, Gharbia SE. Tracing the transition of methicillin resistance in sub-populations of *Staphylococcus aureus*, using SELDI-TOF mass spectrometry and artificial neural network analysis. Syst Appl Microbiol. 2011;34:81–6.

Sharma P, Kumar B, Gupta Y, Singhal N, Katoch VM, Venkatesan K, Bisht D. Proteomic analysis of streptomycin resistant and sensitive clinical isolates of *Mycobacterium tuberculosis*. Proteome Sci. 2010;8:59.

Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Anal Chem. 1996;68:850–8.

Sheynkman GM, Johnson JE, Jagtap PD, Shortreed MR, Onsongo G, Frey BL, Griffin TJ, Smith LM. Using Galaxy-P to leverage RNA-Seq for the discovery of novel protein variations. BMC Genomics. 2014;15:703.

Siegrist TJ, Anderson PD, Huen WH, Kleinheinz GT, McDermott CM, Sandrin TR. Discrimination and characterization of environmental strains of *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). J Microbiol Methods. 2007;68:554–62.

Stephan R, Cernela N, Ziegler D, Pflueger V, Tonolla M, Ravasi D, Fredriksson-Ahomaa M, Haechler H. Rapid species specific identification and subtyping of *Yersinia enterocolitica* by MALDI-TOF Mass spectrometry. J Microbiol Methods. 2011;87:150–3.

Swatkoski S, Russell SC, Edwards N, Fenselau C. Rapid chemical digestion of small acid-soluble spore proteins for analysis of *Bacillus spores*. Anal Chem. 2006;78:181–8.

Syrmis MW, Moser RJ, Whiley DM, Vaska V, Coombs GW, Nissen MD, Sloots TP, Nimmo GR. Comparison of a multiplexed MassARRAY system with real-time allele-specific PCR technology for genotyping of methicillin-resistant *Staphylococcus aureus*. Clin Microbiol Infect. 2011;17:1804–10.

Szabados F, Woloszyn J, Richter C, Kaase M, Gatermann S. Identification of molecularly defined *Staphylococcus aureus* strains using matrix-assisted laser desorption/ionization time of flight mass spectrometry and the Biotyper 2.0 database. J Med Microbiol. 2010;59:787–90.

Teramoto K, Kitagawa W, Sato H, Torimura M, Tamura T, Tao H. Phylogenetic analysis of *Rhodococcus erythropolis* based on the variation of ribosomal proteins as observed by matrix-assisted laser desorption ionization-mass spectrometry without using genome information. J Biosc Bioeng. 2009;108:348–53.

Truong QP, Hammer E, Salazar MG, Do TTH, Nguyen LH, Dang MH, Nguyen TH, Phung TT, Volker U. 2D DIGE proteomic analysis of multidrug resistant and susceptible clinical *Mycobacterium tuberculosis* isolates. J Integr Omics. 2014;4:28–36.

von Wintzingerode F, Boecker S, Schloetelburg C, Chiu NHL, Storm N, Jurinke C, Cantor CR, Goebel UB, van den Boom D. Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: a tool for rapid bacterial identification. Proc Natl Acad Sci U S A. 2002;99:7039–44.

Walker J, Fox AJ, Edwards-Jones V, Gordon DB. Intact cell mass spectrometry (ICMS) used to type methicillin-resistant *Staphylococcus aureus*: media effects and inter-laboratory reproducibility. J Microbiol Methods. 2002;48:117–26.

Wilcox SK, Cavey GS, Pearson JD. Single ribosomal protein mutations in antibiotic-resistant bacteria analyzed by mass spectrometry. Antimicrob Agents Ch. 2001;45:3046–55.

Williamson YM, Moura H, Woolfitt AR, Pirkle JL, Barr JR, Carvalho MG, Ades EP, Carlone GM, Sampson JS. Differentiation of *Streptococcus pneumoniae* conjunctivitis outbreak isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2008;74:5891–7.

Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, Aepfelbacher M, Christner M. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. Int J Med Microbiol. 2011;301:64–8.

Woo S, Cha SW, Merrihew G, He Y, Castellana N, Guest C, MacCoss M, Bafna V. Proteogenomic database construction driven from large scale RNA-seq data. J Proteome Res. 2014;13:21–8.

Xu C, Lin X, Ren H, Zhang Y, Wang S, Peng X. Analysis of outer membrane proteome of *Escherichia coli* related to resistance to ampicillin and tetracycline. Proteomics. 2006;6:462–73.

Yao ZP, Demirev PA, Fenselau C. Mass spectrometry-based proteolytic mapping for rapid virus identification. Anal Chem. 2002;74:2529–34.

Chapter 12
Detection of β-Lactamases and Their Activity Using MALDI-TOF MS

Jaroslav Hrabak, Vladimír Havlicek and Costas C. Papagiannitsis

β-Lactamases

Antibiotics have been used for the treatment of infectious diseases and for a number of non-human applications (agriculture, animal husbandry, and aquaculture) during the past 70 years (Levy and Marshall 2004). Prior to the introduction of antibiotics, natural populations of human/animal bacterial pathogens or commensal bacteria were susceptible to antibiotics (Hughes and Datta 1983). Immediately after the entry of antibiotics in the treatment of infectious diseases, the appearance of antibiotic-resistant bacteria was observed. Today, the overwhelming majority of enterobacteria are resistant to sulfonamides, the first antibacterial chemotherapeutics introduced in clinical practice in 1937. Additionally, a high proportion of bacteria are resistant to a broad range of penicillins, streptomycin, chloramphenicol, and tetracyclines.

Dealing with the problem of antibiotic resistance, older antibiotics were replaced with new antibiotic families or improved versions of old families with more effective action. However, resistance to these new antibiotic molecules emerged quickly (Hede 2014).

Antibiotic resistance is a worldwide problem. Increasing trends of antibiotic-resistant bacteria, causing infections in humans, have been reported by many European countries (ECDC 2012). Emergence of this phenomenon complicated the antibiotic treatment of infectious diseases and significantly increased its cost. As a consequence, several reports describing the failure of antibiotic treatment schemes

J. Hrabak (✉) · C. C. Papagiannitsis
Biomedical Center and Department of Microbiology,
Charles University in Prague, Alej Svobody 76, 323 00 Plzen, Czech Republic
e-mail: Jaroslav.Hrabak@lfp.cuni.cz

V. Havlicek
RCPTM, Department of Analytical Chemistry, Palacky University,
17.listopadu 12, 771 46 Olomouc, Czech Republic

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_12

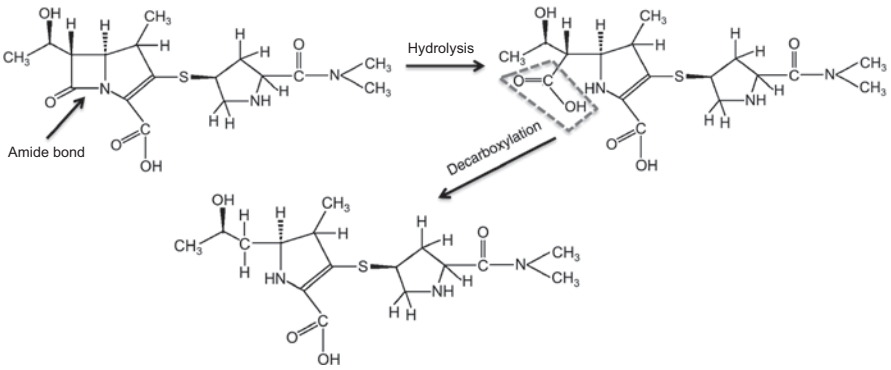


Fig. 12.1 Hydrolysis of the amide bond of the β-lactam ring of β-lactam antibiotics (meropenem)

used in humans and animals have been published (Levy and Marshall 2004). In addition, suffering and mortality due to untreatable infections have increased dramatically. The most alarming rise in antibiotic resistance in humans over the past decade has been observed in *Escherichia coli*, *Klebsiella pneumoniae,* and other members of Enterobacteriaceae family (ECDC 2011). The resistance mechanisms of these bacteria, posing a threat for public health, often involve production of enzymes inactivating antibiotics, such as β-lactamases and aminoglycoside-modifying enzymes (Hrabak et al. 2013). Also, the efflux of antibiotics from the cell and the protection of antibiotic targets from the action of antibiotics [e.g., plasmid-mediated quinolone resistance determinants] are some of the resistance mechanisms harbored by enteric bacteria. Currently, however, the most important and largest group of resistance mechanisms is the production of enzymes degrading the amide bond of the β-lactam ring in β-lactam–β-lactamases (Fig. 12.1).

β-Lactamases were first described in *Staphylococcus aureus* as factors causing resistance to penicillin (Kirby 1944). Later, the first plasmid-encoded β-lactamase, designated TEM based on the patient's name, was described in Greece from a strain of *E. coli* (Datta and Kontomichalou 1965). Currently, hundreds of different types of β-lactamases have been described. These enzymes are categorized based on their molecular properties (Ambler classification) or their hydrolytic pattern (Bush classification)—see Table 12.1 (Bush et al. 1995; Bush and Jacoby 2010).

β-Lactamase-mediated resistance of Gram-negative bacteria (especially of Enterobacteriaceae family) to cephalosporins and recently to carbapenems has continuously increased since 1980s (Cantón et al. 2012; Gniadkowski 2008). At the beginning of the new millennium, multiresistant enterobacteria producing extended-spectrum β-lactamases (ESβLs) and acquired AmpC-type cephalosporinases have spread worldwide, mainly as nosocomial pathogens, but also in the community (Pitout 2010). The high consumption of carbapenem antibiotics under epidemic situation has been followed by a rapid spread of carbapenem resistance, in large measure due to acquisition of genes encoding carbapenemases (Borg et al. 2008; Nordmann et al. 2011). Nowadays, bacteria producing carbapenemases, being resistant to all

Table 12.1 Classification of β-lactamases (Bush and Jacoby 2010; Bush 2013)

Bush classification Group	Subgroup	Ambler classification	Main properties
Cephalosporinases	2a	C	AmpC β-lactamases; hydrolysis of all β-lactams, except of carbapenems (and mostly fourth generation cephalosporins); non-inhibited by conventional serine inhibitors (i.e., clavulanic acid, tazobactam, and sulbactam)
	2b		
	2be		
Penicillinases, cephalosporinases, carbapenemases	2br	A	Staphylococcal penicillinases
	2c		Broad-spectrum enzymes, e.g., TEM-1,2; SHV-1
	2e		Extended-spectrum β-lactamases (ESβL), e.g., TEM-3; SHV-2; CTX-M
	2f		Inhibitor-resistant enzymes (IRT)
	2d		Carbenicillinases
			Cephalosporinases inhibited by clavulanic acid
			Carbapenemases inhibited by clavulanic acid
		D	Cloxacillinases/oxacillinases (OXA)
Metallo-β-lactamases	3a	B	Metallo-β-lactamases (usually Zn^{2+} -dependent)
	3b 3c		

available antibiotics and causing severe infections (e.g., bloodstream infections), have been reported in several countries worldwide. Carbapenemase-producing bacteria express diverse carbapenem-hydrolyzing enzymes, namely metallo-β-lactamases (MβLs), *K. pneumoniae* carbapenemases (KPCs), and OXA-48-type oxacillinases. The clinically most important carbapenemases in Enterobacteriaceae are the class A enzymes of the KPC type. KPC β-lactamases exhibit activity against a wide spectrum of β-lactams, including penicillins, older and newer cephalosporins, aztreonam, and carbapenems. Acquired MβLs are resistance determinants of increasing clinical importance in Gram-negative pathogens. Of these, mainly enzymes of VIM, IMP, and NDM types have been encountered in *K. pneumoniae* and other members of Enterobacteriaceae family. MβLs hydrolyze most β-lactams, including expanded-spectrum cephalosporins (ESCs), and are not inhibited by β-lactam inhibitors. In the past 10 years, Ambler class D enzymes (oxacillinases) able to hydrolyze carbapenems have become a serious problem worldwide. These enzymes are usually abbreviated as carbapenem-hydrolyzing class D β-lactamases (CHDL). Their clinical and epidemiological importance has been primarily recognized in *Acinetobacter* spp. and recently in Enterobacteriaceae (OXA-48-type carbapenemases) (Diene and Rolain 2014).

Carbapenemase-producing Enterobacteriaceae (CPE) have already been detected all over the globe with a marked endemicity according to the enzyme type. In Europe, the most critical situation has been reported in Greece and Italy with 60.5 and 28.8 % incidence, respectively, of carbapenem-resistant *K. pneumoniae* isolates recovered from blood samples within the EARS-Net project in 2012 (ECDC 2012). Infections caused by CPE are connected with a significant mortality. Mortality rates observed in small clinical studies ranged from 22 to 72 % (Hirsch and Tam 2010;

Borer et al. 2009). The reasons for this fact are probably multiple, including underlying diseases, delays in the initiation of antibiotic treatment, and lack of effective antimicrobials (Patel et al. 2008). Yet, there is no equal molecule to substitute carbapenem antibiotics for the treatment of severe infections caused by multidrug-resistant Gram-negative bacteria.

Data from many epidemiological studies show that active surveillance may prevent a spread of ESβL- and carbapenemase-producing bacteria. For successful intervention, however, there is an urgent need for diagnostic laboratories to introduce rapid and sensitive methodologies for the detection of carbapenemase producers (Hrabak et al. 2014).

Detection of β-Lactam Hydrolysis

Therefore, the introduction of rapid and sensitive methodologies for the detection of carbapenemase-producing bacteria is of utmost importance. Contrary to the methods based on susceptibility, inhibition patterns, or molecular-genetic techniques, methods measuring β-lactam hydrolysis are gold standard assays for the detection of β-lactamases (Hrabak et al. 2014). Hydrolysis of β-lactams detected by spectrometric measurement in a UV spectrophotometer has been used as reference by laboratories for many years. This method, however, is labor intensive and needs the preparation of β-lactamase-containing extracts (Cornaglia et al. 2007). Recently, two new methods (colorimetric and matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF MS)-based assays) allowing direct detection of carbapenemase activity were developed. In 2012, the group of Patrice Nordmann developed a colorimetric assay for the detection of carbapenemases and named it Carba NP test (Dortet et al. 2012). The test is based on a detection of acidification of the reaction mixture due to a hydrolysis of the β-lactam ring.

Electrospray ionization was investigated in 2008 by monitoring the concentration and toxicity of meropenem and its degradation products (Mendez et al. 2008). Later, the idea that MALDI-TOF MS also can track the molecular weight changes of β-lactams was described for the first time and used for the detection of carbapenemase-producing bacteria (Hrabak et al. 2011; Burckhardt and Zimmermann 2011). Sparbier et al. showed that MALDI-TOF MS is able to detect not only carbapenems (meropenem and ertapenem), but also other β-lactams (e.g., ampicillin, third-generation cephalosporins). Therefore, β-lactamases able to hydrolyze the ESCs (e.g., ESβLs and AmpC-type enzymes) can also be detected.

For routine detection of β-lactamase activity, a fresh bacterial culture is mixed with a β-lactam solution (meropenem or ertapenem) (Hrabak et al. 2011; Burckhardt and Zimmermann 2011; Hrabak et al. 2013, 2014; Studentova et al. 2015) (Fig. 12.2). After incubation at 35–37 °C for 2–4 h, the reaction mixture is centrifuged and the supernatant is applied onto a MALDI target, covered by a matrix solution (HCCA or DHB), and measured in a positive ion mode. In the first descrip-

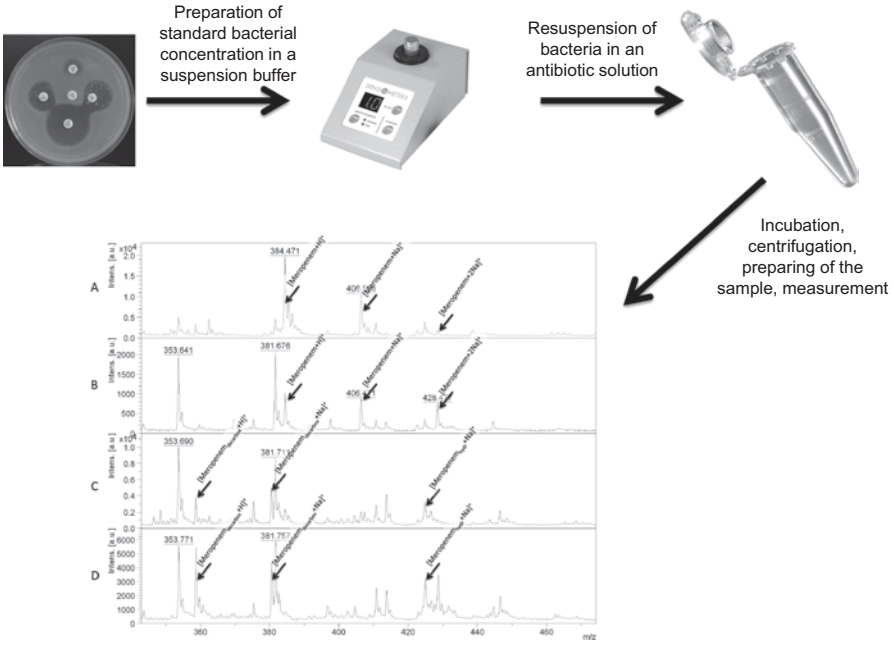


Fig. 12.2 Detection of β-lactamases using a fresh bacterial culture

tion of this assay (Hrabak et al. 2011), the interpretation was based on disappearance of the protonated carbapenem molecule and its sodium salt. Other authors showed that degradation products can also be observed and usually appear as ionic species with a mass shift of +18 Da (hydrolysis of amide bound) or −26 Da (decarboxylated hydrolyzed form) (see Figs. 12.1 and 12.3; Burckhardt and Zimmermann 2011; Sparbier 2012).

Quality of the measurement and thus detection of modified molecules may be enhanced under proper reaction conditions (i.e., concentration of bacteria used and reaction buffer) (Hrabak et al. 2012). These findings improved interpretation criteria and suppressed false-positive or -negative results of the assay.

Detection of CHDL enzymes represents a significant challenge for microbiological laboratories. Using a MALDI-TOF MS carbapenemase assay, different sensitivity values have been reported (Kempf et al. 2012; Chong et al. 2015). Studentova et al. (2015) and Papagiannitsis et al. (2015) both showed that addition of ammonium bicarbonate may enhance re-carboxylation of active-site lysine of OXA-48-like enzymes, resulting in an increased sensitivity of the assay due to enhancement of enzymatic activity. Interestingly, the assay without any modification also gives superior results for a detection of CHDL in *Acinetobacter baumannii* (Kempf et al. 2012).

Carvalhaes et al. (2013) validated a liquid chromatography–mass spectrometry (LC–MS) method for the detection of carbapenemases in CHDL-, IMP-, NDM-,

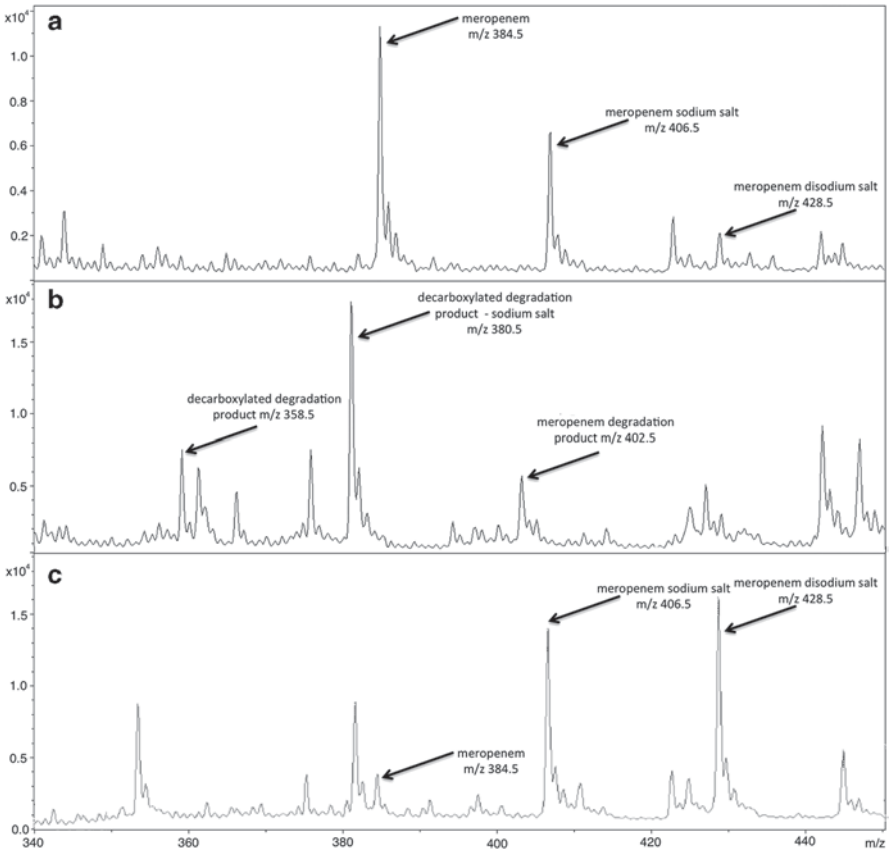


Fig. 12.3 Mass spectra of intact meropenem and its sodium salts (**a**), meropenem degradation products (**b**) after hydrolysis by carbapenemase KPC-2, and mass spectra of carbapenemase non-producing strain (**c**)

VIM-, GIM-, KPC-, and GES-producing bacteria. Sensitivity and specificity of the method after 4-h incubation of the reaction mixture achieved 100 % for both parameters. Similar results were observed using ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) (Carricajo et al. 2014).

Identification of β-Lactamases Using Inhibitors

As described in Table 12.1, β-lactamases are categorized based on their hydrolytic and inhibitory pattern. Therefore, preliminary characterization of a detected enzyme has been tested using specific inhibitors (Sparbier et al. 2012). For ESβL-type enzymes, serine inhibitors such as clavulanic acid and tazobactam can easily be used. However, the inhibitors used for discrimination of different types of carbapene-

mases (e.g., EDTA for MβLs, phenylboronic acid for KPCs) may interact during MALDI-TOF MS measurement. For example, after addition of β-lactamase inhibitors to the reaction mixture, unusable spectra are usually obtained due to the high noise level and inhibitor peaks overlapping with the corresponding β-lactam peaks. Therefore, high-inhibitor concentrations allowing reliable inhibition rate determination cannot be used directly, and further research in this area is necessary.

Comparison with Carba NP and Spectrophotometric Assays

Since 2011, two new direct assays for detection of carbapenemases have been developed (MALDI-TOF MS hydrolysis assay and Carba NP). Both are designed to substitute reference spectrophotometric assays. Contrary to the spectrophotometric assay and the Carba NP test, the MALDI-TOF MS hydrolysis assay allows for direct detection of hydrolysis products, which underpins the specificity of the test. By February 2015, three studies focused on a comparison of the MALDI-TOF MS hydrolysis assay with Carba NP (Knox et al. 2014; Papagiannitsis et al. 2015; Chong et al. 2015). Contrary to Knox et al. who found similar sensitivity for both tests, two others found better sensitivity in the MALDI-TOF MS hydrolysis assay in Enterobacteriaceae. As already mentioned, the MALDI-TOF MS hydrolysis assay can give excellent results for the detection of CHDL in *Acinetobacter baumannii* (Kempf et al. 2012).

Automated Detection

The main disadvantage of the MALDI-TOF MS hydrolysis assay for detection of β-lactamases is the interpretation of mass spectra. As most clinical microbiologists are not familiar with manual MS measurement, there is an urgent need for software providing automated reading and interpretation of the spectra. Bruker Daltonics has developed a tool that is currently under laboratory evaluation and allows for automated measurement and interpretation of β-lactam hydrolysis (MBT STAR-BL; Bruker Daltonik GmbH). Using proper thresholds, the software may provide valuable results (Papagiannitsis et al. 2015). Further evaluation, however, is needed prior to its routine use in diagnostic laboratories.

Detection of β-Lactamase Types

Direct molecular visualization of β-lactamases has been a big challenge for another application of MALDI-TOF MS in microbiology (Hrabak et al. 2013). First, successful detection of a peak corresponding to a β-lactamase was published by Ca-

mara and Hays in 2007. They differentiated wild-type *E. coli* (ATCC 700926) from ampicillin-resistant (Amp^R) plasmid-transformed *E. coli* strains by the direct visualization of a β-lactamase. The detected peak with an *m*/*z* of ca. 29,000 corresponded to the pUC19-encoded TEM-1 β-lactamase (theoretical relative molecular weight 28,949 g/mol). In a recent MALDI-TOF MS study by Schaumann et al. (2012), it was not possible to distinguish clinical isolates of Enterobacteriaceae and *Pseudomonas aeruginosa* producing extended-spectrum β-lactamases or MβLs from non-producers. In that study, the mass measurements were performed between *m*/*z* 2000 and 12,000, which is rather inappropriate for the detection of intact β-lactamases. Consequently, so far the attempts to visualize native β-lactamases by MALDI-TOF MS in spectra of wild-type bacteria have been mostly unsuccessful.

In 2014, Papagiannitsis et al. described a new assay for the identification of CMY-2-like β-lactamases in clinical enterobacterial isolates by MALDI-TOF MS. The new method is based on the extraction of periplasmic proteins and the detection of CMY-2-like β-lactamases by MALDI-TOF MS according to their molecular weight. Successful extraction of β-lactamases from the periplasmic space of bacteria was a crucial step for the performance of the described assay. Therefore, a modified sucrose method for the extraction of periplasmic proteins has been used. Purified β-lactamases were used as positive controls for MALDI-TOF MS measurements, and for calibration and setting up the mass spectrometer's parameters.

The MALDI-TOF MS assay described above was able to detect the presence of an approximately 39,850-*m*/*z* peak, which can be used as an indicator for the presence of the *C. freundii*-derived CMY-2-like group of the acquired AmpC β-lactamases (Papagiannitsis et al. 2014). In addition, the observation of the 39,670 and 38,900-*m*/*z* peaks for ACC-4 and DHA-1 enzymes, respectively, indicated that MALDI-TOF MS may discriminate the diverse groups of acquired AmpC-type cephalosporinases (see Fig. 12.4). In addition, the latter method revealed a peak at *m*/*z* 383, representing the putative acyl–enzyme complex (complex of CMY-2 β-lactamase with the meropenem molecule).

These recent data indicate that MALDI-TOF MS has the potential to directly detect the most clinically important AmpC β-lactamases, such as the CMY-2-like, ACC, and DHA types, in clinical isolates of Enterobacteriaceae. In agreement with other MALDI-TOF MS applications (Hrabak 2013), the described protocol is quick and economical. In addition, detection of β-lactamases by MALDI-TOF MS in a proteomic approach allowing the study of the behavior of the tested strains can complement the already used techniques for characterization of β-lactamases, such as PCR and isoelectric focusing (IEF). MALDI-TOF MS can directly detect the class A (Camara and Hays 2007) and class C β-lactamases, as well as other mechanisms such as methylation of rRNA and cell wall components (Cai et al. 2012; Hrabak et al. 2013). We conclude that establishing a MALDI-TOF supplementary database of resistance mechanisms would promote further research in this field.

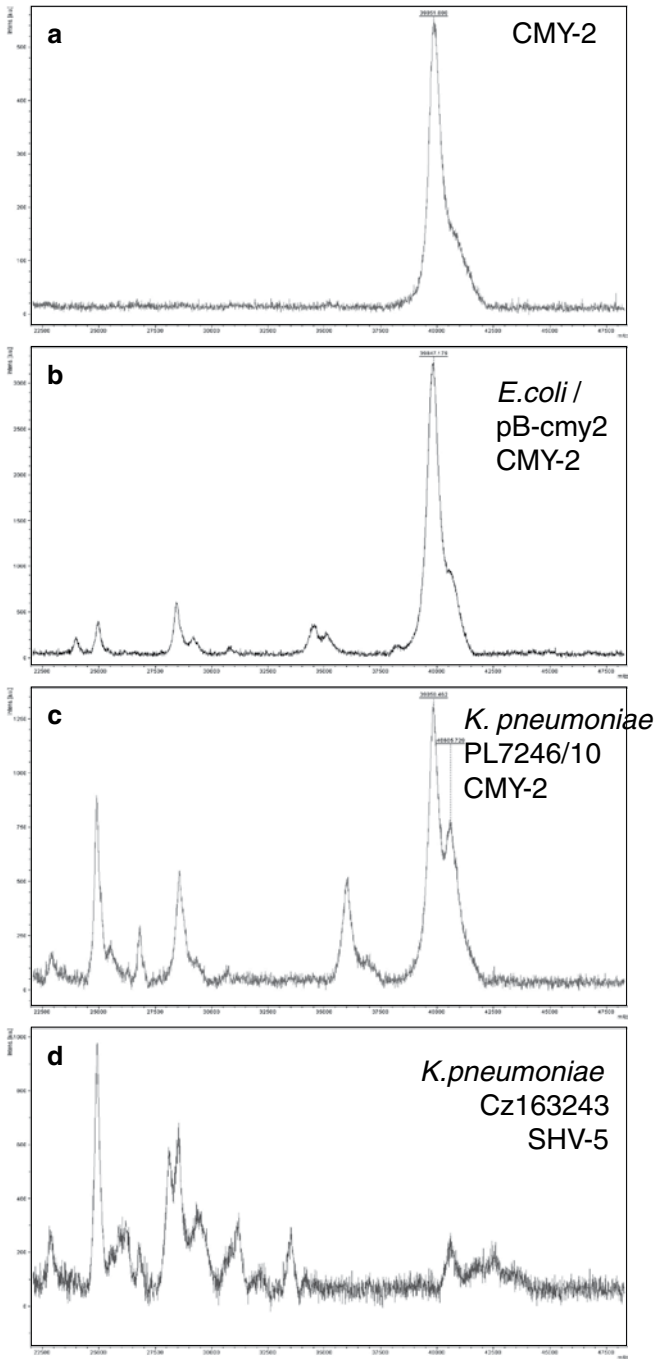


Fig. 12.4 Mass spectra showing CMY-2 β-lactamase. Mass spectra of a purified enzyme (**a**), of the periplasmic extract of laboratory *Escherichia coli* strain with cloned *bla*CMY-2 (**b**), of the periplasmic extract of wild *Klebsiella pneumoniae* isolate (**c**), and of periplasmic extract of wild *Klebsiella pneumoniae* isolate not producing CMY-2 enzyme

Conclusion

Direct detection of β-lactamase activity has been established as a routine method in many microbiological diagnostic laboratories (Hrabak et al. 2014). Especially in the case of a direct detection of carbapenemase activity, the MALDI-TOF MS hydrolysis assay should soon be accepted as a "gold standard method" and may serve as a reference technique together with spectrophotometric assays. The main advantage of MS tools is the ability to detect rapidly both antibiotic hydrolytic products and intact lactamases.

References

Borer A, Saidel-Odes L, Riesenberg K. Attributable mortality rate for carbapenem-resistant *Klebsiella pneumoniae* bacteremia. Infect Control Hosp Epidemiol. 2009;30:972–6.

Borg MA, Zarb P, Ferech M, Goosens H. Antibiotic consumption in southern and eastern Mediterranean hospitals: results from the ARMed project. J Antimicrob Chemother. 2008;62:830–6.

Burckhardt I, Zimmermann S. Using matrix-assisted laser desorption ionization-time of flight mass spectrometry to detect carbapenem resistance within 1-2.5 hours. J Clin Microbiol. 2011;49:3321–4.

Bush K. The ABCD's of β-lactamase nomenclature. J Infect Chemother. 2013;19:549–59.

Bush K, Jacoby GA. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother. 2010;54:969–76.

Bush K, Jacoby GA, Medeiros AA. A functional classification scheme for beta-lactamases and its correlation with molecular structure. Antimicrob Agents Chemother. 1995;39:1211–33.

Cai JC, Hu YY, Zhang R, Zhou HW, Chen G-X. Detection of Ompk36 porin loss in *Klebsiella* spp. by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50:2179–82.

Camara JE, Hays FA. Discrimination between wild-type and ampicillin-resistant *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Bioanal Chem. 2007;389:1633–8.

Cantón R, Akóva M, Carmeli Y, Giske CG, Glupczynski Y, Gniadkowski M, Livermore DM, Miriagou V, Naas T, Rossolini GM, Samuelsen Ø, Seifert H, Woodford N, Nordmann P, European Network on Carbapenemases. (2012) Rapid evolution and spread of carbapenemases among *Enterobacteriaceae* in Europe. Clin Microbiol Infect. 18: 413–31.

Carricajo A, Verhoeven PO, Guezzou S, Fonsale N, Aubert G. Detection of carbapenemase-producing bacteria using the ultra-performance liquid chromatography-tandem mass spectrometry method. Antimicrob Agents Chemother. 2014;58:1231–4.

Carvalhaes CG, Cayô R, Assis DM. Detection of SPM-1-producing *Pseudomonas aeruginosa* and class D β-lactamase-producing *Acinetobacter baumannii* isolates by use of liquid chromatography-mass spectrometry and matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2013;51:287–90.

Chong PM, McCorrister SJ, Unger MS, Boyd DA, Mulvey MR, Westmacott GR. MALDI-TOF MS detection of carbapenemase activity in clinical isolates of *Enterobacteriaceae* spp., *Pseudomonas aeruginosa,* and *Acinetobacter baumannii* compared against the Carba-NP assay. J Microbiol Methods. 2015;111:21–3.

Cornaglia G, Akova M, Amicosante G. Metallo-beta-lactamases as emerging resistance determinants in Gram-negative pathogens: open issues. Int J Antimicrob Agents. 2007;29:380–8.

Datta N, Kontomichalou P. Penicillinase synthesis controlled by infectious R factors in *Enterobacteriaceae*. Nature. 1965;208:239–41.

Diene SM, Rolain JM. Carbapenemase genes and genetic platforms in Gram-negative bacilli: *Enterobacteriaceae, Pseudomonas* and *Acinetobacter* species. Clin Microbiol Infect. 2014;20:831–8.

Dortet L, Poirel L, Nordmann P. Rapid identification of carbapenemase types in *Enterobacteriaceae* and *Pseudomonas* spp. by using a biochemical test. Antimicrob Agents Chemother. 2012;56:6437–40.

European Centre For Disease Prevention and Control (ECDC). (2012) Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). http://ecdc.europa.eu. Accessed 31 March 2015.

Gniadkowski M. Evolution of extended-spectrum beta-lactamases by mutation. Clin Microbiol Infect. 2008;14(Suppl 1):11–32.

Hede K. Antibiotic resistance: an infectious arms race. Nature. 2014;509:S2–S3.

Hirsch EB, Tam VH. Detection and treatment options for *Klebsiella pneumoniae* carbapenemases (KPCs): an emerging cause of multidrug-resistant infection. J Antimicrob Chemother. 2010;65:1119–25.

Hrabak J, Walkova R, Studentova V, Chudackova E, Bergerova T. Carbapenemase activity detection by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2011;49:3222–7.

Hrabak J, Chudackova E, Walkova R. Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. Clin Microbiol Rev. 2013;26:103–14.

Hrabak J, Chudackova E, Papagiannitsis CC. Detection of carbapenemases in *Enterobacteriaceae:* a challenge for diagnostic microbiological laboratories. Clin Microbiol Infect. 2014;20:839–53.

Hughes VM, Datta N. Conjugative plasmids in bacteria of the 'pre-antibiotic' era. Nature. 1983;302:725–6.

Kempf M, Bakour S, Flaudrops C. Rapid detection of carbapenem resistance in *Acinetobacter baumanii* using matrix-assisted laser desorption ionization-time of flight mass spectrometry. PLoS ONE. 2012;7:e31676.

Kirby WMM. Extraction of a highly potent penicillin inactivator from penicillin resistant *Staphylococci*. Science. 1944;99:452–3.

Knox J, Jadhav S, Sevior D, Agyekum A, Whipp M, Waring L, Iredell J, Palombo E. Phenotypic detection of carbapenemase-producing *Enterobacteriaceae* by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry and the Carba NP test. J Clin Microbiol. 2014;52:4075–7.

Levy SB, Marshall B. Antibacterial resistance worldwide: causes, challenges and responses. Nat Med. 2004;10:S122–S9.

Mendez A, Chagastelles P, Palma E, Nardi N, Schapoval E. Thermal and alkaline stability of meropenem: degradation products and cytotoxicity. Int J Pharm. 2008;350:95–102.

Nordmann P, Naas T, Poirel L. Global spread of carbapenemase-producing Enterobacteriaceae. Emerg Infect Dis. 2011;17:1791–8.

Papagiannitsis CC, Kotsakis SD, Tuma Z, Gniadkowski M, Miriagou V, Hrabak J. Identification of CMY-2-type cephalosporinases in clinical isolates of *Enterobacteriaceae* by MALDI-TOF MS. Antimicrob Agents Chemother. 2014;58:2952–7.

Papagiannitsis CC, Študentová V, Izdebski R, Oikonomou O, Pfeifer Y, Petinaki E, Hrabak J. MALDI-TOF MS meropenem hydrolysis assay with NH4HCO3, a reliable tool for the direct detection of carbapenemase activity. J Clin Microbiol. 2015;53:1731–5.

Patel G, Huprikar S, Factor SH, Jenkins SG, Calfee DP. Outcomes of carbapenem-resistant *Klebsiella pneumoniae* infection and the impact of antimicrobial and adjunctive therapies. Infect Control Hosp Epidemiol. 2008;29:1099–106.

Pitout JD. Infections with extended-spectrum beta-lactamase-producing *Enterobacteriaceae:* changing epidemiology and drug treatment choices. Drugs. 2010;70:313–33.

Schaumann R, Knoop N, Gnzel GH, Losensky K, Rosenkranz C, Stingu CS, Schellenberger W, Rodlodd AC, Eschrich K. A step towards the discrimination of beta-lactamase-producing clini-

cal isolates of *Enterobacteriaceae* and *Pseudomonas aeruginosa* by MALDI-TOF mass spectrometry. Med Sci Monit. 2012;18:MT1–7.

Sparbier K, Schubert S, Weller U, Boogen C, Kostrzewa M. Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based functional assay for rapid detection of resistance against beta-lactam antibiotics. J Clin Microbiol. 2012;50:927–37.

Studentova V, Papagiannitsis CC, Izdebski R, Pfeifer Y, Chudackova E, Bergerova T, Gniadkowski M, Hrabak J. Detection of OXA-48-type carbapenemase-producing *Enterobacteriaceae* in diagnostic laboratories can be enhanced by addition of bicarbonates to cultivation media or reaction buffers. Folia Microbiol. 2015;60:119–29.

Chapter 13
Stable-Isotope-Based Strategies for Rapid Determination of Drug Resistance by Mass Spectrometry

Plamen Demirev

Introduction

The emergence of drug-resistant microorganisms ("superbugs"), unaffected by available pharmaceuticals, has been recognized as an acute worldwide health problem (Choffnes et al. 2010) recognized by the World Health Organization (WHO 2015). Rapid determination of the responses of a pathogenic microorganism to antibiotics is very important for taking, for example, timely countermeasures in response to a bioterrorism attack or for efficient curbing of the spread of infections in a hospital setting. A US national strategy for combating drug resistance, formulated in the fall of 2014, envisions complex measures to mitigate the emergence and spread of antibiotic resistance (National Strategy on Combating Antibiotic-Resistant Bacteria 2014). Classical microbiology techniques, including broth dilution or disk diffusion, have been used for decades to determine drug resistance (Coudron et al. 1986; Tang and Stratton 2006). These techniques infer organism proliferation in the presence of a drug (i.e., resistance) by, for example, monitoring changes in optical density (turbidity) of culture suspensions. However, such classical assays for drug resistance are not rapid, typically taking between 24 and 48 h. These delays reduce markedly the efficacy of efforts to curb disease spread or mitigate effects of bioterrorism activities. Time is of the essence in identifying, treating, or eradicating, particularly, virulent and unknown pathogens. In addition, existing classical tests with higher false-positive or false-negative rates can result in additional complications, wasting valuable resources in improper and inefficient treatments.

Mass spectrometry (MS) , a molecular-level biophysical technology, offers several advantages for pathogen detection, including speed, sensitivity, and specificity, and it rapidly revolutionizes the practice of infectious disease diagnostics (Wilkins et al. 2005; Demirev and Fenselau 2008a, b; Shah and Gharbia 2010; Cliff et al.

P. Demirev (✉)
Applied Physics Laboratory Johns Hopkins University, Laurel, MD, USA
e-mail: plamen.demirev@jhuapl.edu

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9_13

2011; Ho and Reddy 2010; Fenselau and Demirev 2011; Havlicek et al. 2013; Sandrin et al. 2013). A number of MS approaches for direct detection of drug-resistant bacterial strains have been introduced recently (Hrabak et al. 2013; Demirev 2014; Kostrzewa et al. 2013). For example, matrix-assisted laser desorption/ionization (MALDI) MS generated biomarker profiles differentiate between methicillin-susceptible and methicillin-resistant *Staphylococcus aureus* (MRSA) strains (Edwards-Jones et al. 2000; Jackson et al. 2005; Du et al. 2002; Wolters et al. 2011; Shah et al. 2011; Burckhardt and Zimmermann 2011; Muroi et al. 2012). Direct detection of drug metabolites (e.g., degradation products) in cultures grown in the presence of drugs is another MS approach for establishing drug resistance (Hrabak et al. 2013). In it, the lower *m/z* (less than 600) range of a mass spectrum is monitored for the presence of peak characteristic of the drug and/or its metabolites. This approach relies on knowledge of the degradation products, arising as a result of the known enzyme-facilitated resistance mechanisms.

Functional Assays for Establishing Drug Resistance

Stable-Isotope Biomarker Labeling During Growth in Isotopically Manipulated Culture Medium (Metabolic Labeling) The labeling of biomolecules with stable (nonradioactive) isotopes during microorganism growth was introduced almost 60 years ago (Meselson and Stahl 1958). In the past 20 years, improved sensitivity and signal-to-noise ratio in MS of biomolecules derived from microorganisms have been achieved by growth media manipulation—enrichment or depletion of minor stable isotopes, such as ^{13}C and ^{15}N—and metabolic biomolecule labeling (Oda et al. 1999; Mann 2006; Marshall et al. 1997; Stump et al. 2003). This has resulted in improved and more accurate microorganism identification (ID; Stump et al. 2003) . A rapid functional method to determine drug resistance in microorganisms has been demonstrated recently (Demirev et al. 2013). It is based on microorganism growth in stable-isotope-labeled media that also contain target drugs (Sparbier et al. 2013; Jung et al. 2014). Mass spectra obtained from microorganisms grown under such conditions are compared to control spectra obtained from microorganisms grown in non-labeled media without the drug. Isotope-labeled nutrients are consumed and metabolized during the continued microorganism growth if the organism is not susceptible to the specific drug(s) present. Drug resistance is inferred by observing characteristic mass shifts of one or more microorganism biomarkers. These characteristic mass shifts are a result of the isotopic label(s) being incorporated into the biomarker molecules. The shifts thus indicate that the microorganism is viable even in the presence of the drug. The shifts can be determined experimentally and can also be derived by bioinformatics algorithms. For example, they can be predicted from the isotope composition and the stable-isotope ratios and the actual or estimated biomarker elemental composition. All C-containing molecules consist of approximately one ^{13}C-atom for every 99 ^{12}C-atoms at natural-isotope abundance.

When a molecule containing *n* C-atoms is enriched 100 % in ^{13}C (i.e., all ^{12}C-atoms are replaced by ^{13}C) its molecular weight will be shifted by approximately *n* Da, compared to the molecular weight upon natural-isotope abundance.

Two approaches for isotope labeling of the growth medium have been demonstrated. One approach utilizes globally labeled medium with all growth medium molecules labeled at a predefined isotope ratio of, for example, ^{13}C–^{12}C atoms (Demirev et al. 2013). A partially labeled ("locally labeled") medium contains a particular component that is labeled, for example, ^{13}C- and/or ^{15}N-labeled-specific amino acids. The second approach has been introduced more than 10 years ago as a major step in MS-based quantitative proteomics approaches. Stable isotope labeling by amino acids in cell cultures (SILAC; Mann 2006) is the such a major protocol, and it has been used extensively for identification and quantitative evaluation of the expression levels of individual cellular proteins under various conditions.

The first approach has been illustrated with intact *Escherichia coli,* grown in control (unlabeled) and ^{13}C-labeled media, and analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Demirev et al. 2013). This approach is nearly universal: it can be applied to both known and unknown microorganisms (e.g., with unsequenced genomes and/or newly emerging organisms). If the organism's genome is available, simultaneously with testing for drug resistance the microorganism can also be identified by bioinformatics tools (Pineda et al. 2000; Demirev et al. 2004). On the other hand, neither prior identification of the microorganism being tested nor prior availability of a reference mass spectrum for that organism is required. Depending on the prior knowledge, one or more algorithms for data analysis can be utilized either individually or in parallel (Fig. 13.1). For example, if the organism's genome is known, mapping of observed microorganism protein biomarkers to their respective amino acid sequences is possible (Jung et al. 2014). Thus, direct "counting" of all C- or N-atoms in an observed protein would provide the expected mass shift upon substitution of ^{12}C with ^{13}C (^{14}N with ^{15}N), respectively.

The local labeling approach based on the incorporation of specific isotopically labeled amino acids has been demonstrated by Kostrewa and coworkers using methicillin-resistant *S. aureus* (Sparbier et al. 2013) and *Pseudomonas aeruginosa* (Jung et al. 2014). In both the studies, either lysine at natural-isotope abundance or ^{13}C- and ^{15}N-labeled lysine has been added to the growth medium. In the second study, the susceptibility of *P. aeruginosa* to meropenem, tobramycin, and ciprofloxacin (antibiotics with differing mechanisms of action) has been evaluated utilizing a semiautomated algorithm. As expected, a shift of organism-specific biomarkers is observed for drug-resistant strains upon growth in labeled medium (Fig. 13.2). The observed shifts in biomarker masses in the case of locally isotope-labeled medium (i.e., individual amino acids) are typically at least an order of magnitude smaller when compared to shifts, observed for globally labeled medium. Thus, local labeling places more stringent requirements for experimental mass resolution than global labeling. In addition, since in many cases the labeled amino acid may be nonessential (i.e., some can be synthesized de novo during cell development) both unlabeled and labeled biomarker peaks are observed (Fig. 13.2). More sophisti-

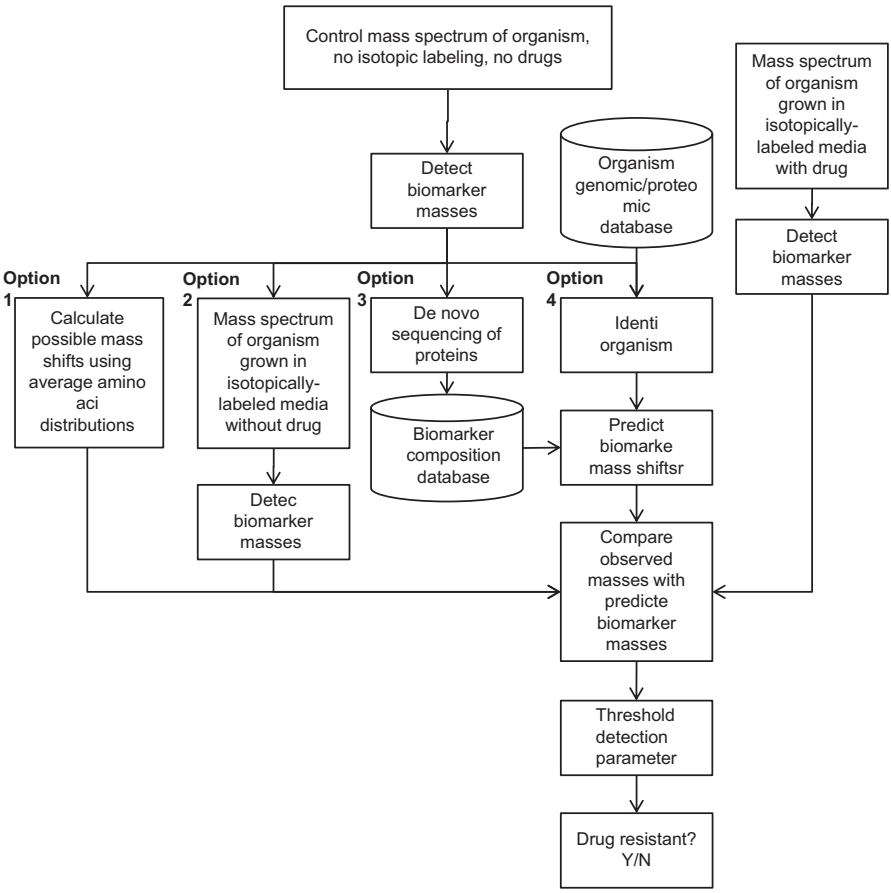


Fig. 13.1 Algorithmic approaches for establishing drug resistance in microorganisms by MS, after growth in an isotopically labeled medium. Each algorithm can be applied individually or in parallel, depending on the available prior information. (Reprinted from Demirev et al. (2013), with permission from Springer)

cated algorithms, preferably combined with knowledge of the amino acid sequence of the protein biomarker, are required for data deconvolution and semiquantitative analysis when using local isotope labeling protocols. On the other hand, considering the availability of isotope-labeled components of growth media [35], local labeling protocols can be developed more readily from currently existing culture protocols than protocols for global growth medium labeling. Both labeling protocols can be combined with direct MALDI-TOF MS analysis of intact cells or their protein extracts as well as with bottom-up or top-down approaches using liquid chromatography/ESI tandem MS instrumentation. The advantages of using MALDI-TOF MS are relative simplicity of the analysis, combined with microorganism ID, speed, wide instrument availability, and cost effectiveness. Almost all sample preparation and analysis stages can be automated and multiplexed. Isotope-labeling strategies

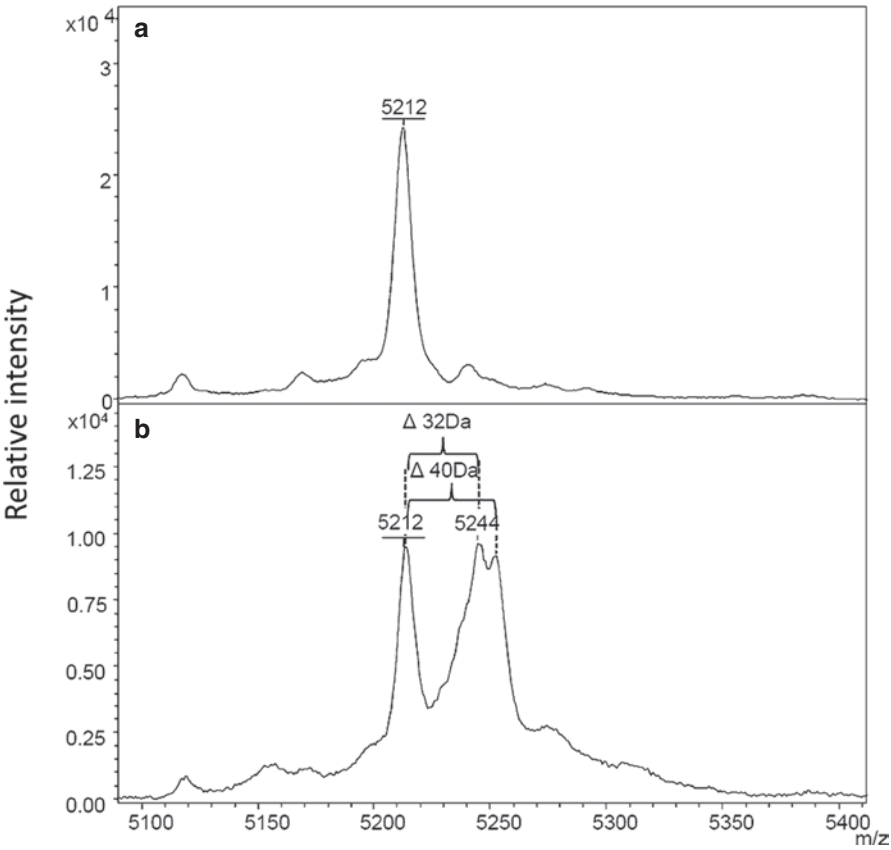


Fig. 13.2 Mass *(m/z)* shifts in MALDI-TOF mass spectra of *Pseudomonas aeruginosa*. **a** Control, spectrum for bacteria grown in a medium containing lysine at natural-isotope abundance. **b** *P. aeruginosa* grown for 2.5 h in a medium containing ^{13}C- and ^{15}N-labeled lysine. Lysine has six C- and two N-atoms, thus ^{13}C- and ^{15}N-labeled lysine will be 8 Da heavier than its unlabeled counterpart. Incorporation of four and five labeled lysines in the (most probably) ribosomal protein biomarker at *m/z* 5212 can be inferred from the mass shifts in spectrum (**b**). (Reprinted from Jung et al. (2014), with permission from Springer)

for establishing drug resistance can be particularly useful for classes of anaerobes, fastidious bacteria, and other slow-growing bacteria (Biswas and Rolain 2013) .

Bacteriophage-Based Amplification for Drug Resistance Detection Phage amplification detection (PAD) of bacteria via MS relies on detecting bacteriophages specific to their target host organism (Rees and Voorhees 2005). Phages self-replicate and proliferate only in metabolically active host cells, have extensive shelf lives, and are inexpensive. Targeted microorganisms are identified through detection by MS of secondary biomarkers originating from organism-specific bacteriophages after their amplification in the target cells. Only proteins indicative of progeny phages, that are bacteria specific, are detected. For example, when both MS2 and MPSS-1 are

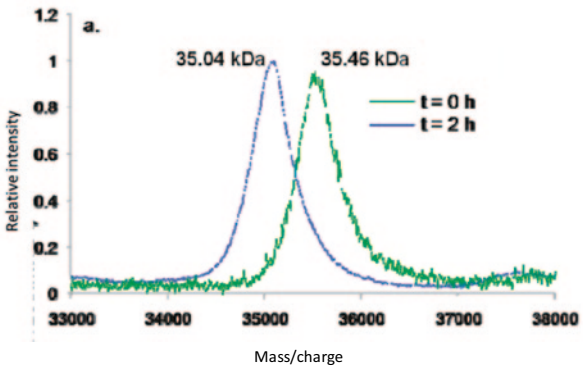


Fig. 13.3 Overlaid MALDI-TOF mass spectra of initial *Staphylococcus* bacteriophage 53 inoculum (2×10^8 PFU mL^{-1}), grown in ^{15}N-labeled medium—*green trace,* and after its propagation for 2 h in *S. aureus* culture (6.7×10^7 CFU mL^{-1})—*blue trace.* The downward shift in mass of the observed biomarkers before and after amplification is due to replacement of the ^{15}N isotope of the approximately 420 N atoms in the phage capsid protein with ^{14}N isotopes. (Adapted with permission from Pierce et al. (2011), copyright 2011, American Chemical Society)

mixed with *E. coli,* only an MS2 biomarker protein is detected by MALDI-TOF MS. Phage-based amplification has been successfully expanded into a functional assay for drug resistance screening of targeted bacteria (Pierce et al. 2011, 2012; Rees et al. 2015). In this process, the shift in characteristic biomarker masses for phages, initially proliferated in stable-isotope-manipulated growth medium, for example, ^{15}N-enriched, is monitored as an indication of successful phage proliferation. High ^{15}N-labeled phage titers (above the MS instrument's detection limits) are used to spike the sample. The isotope-labeled phages can be readily distinguished from phages subsequently proliferating in bacteria in media with natural-isotope abundance by the observed (downward) shifts in biomarker masses (Fig. 13.3). Monitoring phage growth in drug-containing and control culture media in parallel results in an assay for establishing drug susceptibility—detection of phage-specific biomarkers in samples with the drug will signal the presence of live drug-resistant bacterial strains. The initial input inoculum can be readily distinguished from phages successfully proliferating in the bacteria, which eliminates the possibility of false-positive results. Isotope labeling leads to improved selectivity, high initial phage titers, and sensitivity—only organism-specific phages proliferate. The overall time for organism ID as well as drug susceptibility testing is markedly reduced. The method, combined with MALDI-TOF MS, has been demonstrated for rapid detection of MRSA strains (Pierce et al. 2011). In this method, a *S. aureus*-specific phage is initially labeled by proliferating in organisms in ^{15}N-labeled culture medium. Subsequently, the presence of *S. aureus* is confirmed by detecting a ^{14}N-labeled bacteriophage capsid protein signal after 90-min phage incubation in a sample in a growth medium at natural-isotope abundance. The assay has been combined with LC/ESI/tandem MS for multiple reaction monitoring (MRM) of phage-specific tryptic peptides for rapid and accurate quantitation of viable *S. aureus* (Pierce et al. 2012) . After spiking the sample with ^{15}N-labeled phages, and following 2-h incubation, the sample is

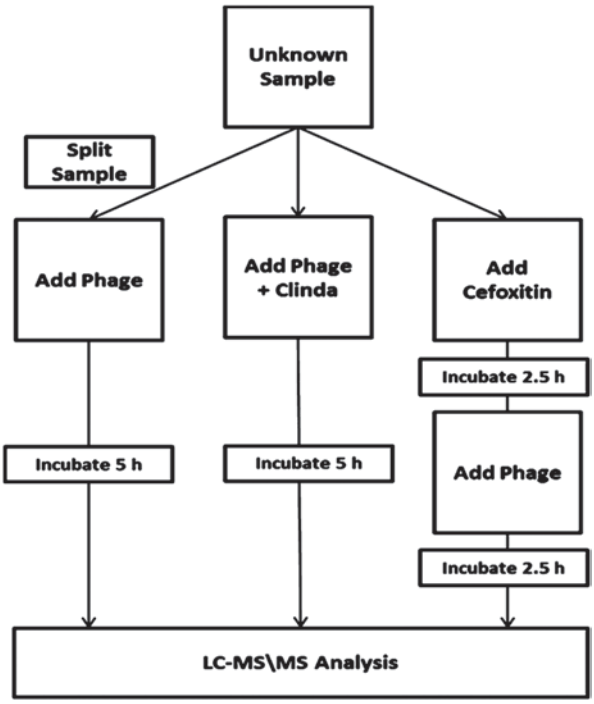


Fig. 13.4 Workflow for a phage-amplification assay using stable-isotope labeled phages to simultaneously determine the presence of *Staphylococcus aureus* in a sample, as well as the susceptibility of microorganism to two antibiotics—clindamycin and cefoxitin. (Reprinted with permission from Rees et al. (2015), copyright 2015, American Chemical Society)

rapidly digested with trypsin. Target tryptic peptides unique to both the ^{15}N-labeled input and ^{14}N progeny capsid proteins are analyzed in MRM. The peptides from a known number of ^{15}N-metabolically labeled phages (10^9 PFU) are used as an internal standard for quantitation. A linear *S. aureus* response is achieved in the range from 5.0×10^4 to 2.0×10^6 CFU/mL. Since different antibiotics have different mechanisms of action, there may be a latency period during which the drug might be effective but there will still be phage amplification. To avoid such false positives of purported microbial resistance against, for example, β-lactam antibiotics, a modified phage amplification protocol for simultaneous ID and susceptibility testing to multiple drugs has been implemented (Rees et al. 2015). Delayed ^{15}N-labeled phage K infection is tested when testing for resistance against slower-acting drugs, including cefoxitin, allowing for effects of the drug to take action (Fig. 13.4). The end-to-end assay can be performed in less than 8 h. This is at least a factor of 3 faster than clinical microbiology assays currently used to detect MRSA.

Conclusion

The emergence of multidrug-resistant microorganism strains requires novel methods for antimicrobial resistance and virulence testing. The functional assays for establishing drug resistance, reviewed here, are based on MS and stable isotopes

for manipulating culture media. Drug resistance can be inferred from characteristic differences in the masses of primary (microorganism-derived) or secondary (phage-derived) biomarkers appearing upon microorganism growth in the presence of a drug. These assays have a number of advantages. They are much more rapid than classical microbiology assays. Also, they are broadly applicable to a variety of drug/ microorganism types since no prior knowledge of the type of drug and its mechanism of action is required. The assays have potential in high-throughput mass screening against known or unknown pathogens. In a clinical sample, pathogenic microorganism ID/confirmation and drug susceptibility testing can be performed in parallel, thus reducing the time for accurate diagnosis and subsequent treatment. These assays can be multiplexed for simultaneous analysis of mixtures of drugs and organisms (samples), potentially reducing cost. Bioinformatics/statistical methods can be implemented for automated data analysis

Acknowledgments Preparation of this review was supported in part by a Stuart Janney Sabbatical Fellowship administered by the Johns Hopkins University Applied Physics Laboratory.

References

Choffnes ER, Relman DA, Mack A, Rapporteurs, Forum on Microbial Threats, Institute of Medicine. Antibiotic resistance: implications for global health and novel intervention strategies. Washington, DC: National Academy of Sciences; 2010.

Biswas S, Rolain J-M. Use of MALDI-TOF mass spectrometry for identification of bacteria that are difficult to culture. J Microbiol Methods. 2013;92:14–24.

Burckhardt I, Zimmermann S. Using matrix-assisted laser desorption ionization-time of flight mass spectrometry to detect carbapenem resistance within 1-2.5 hours. J Clin Microbiol. 2011;49:3321–4.

Cliff JB, Kreuzer HW, Ehrhardt CJ, Wunschel DE, editors. Chemical and physical signatures for microbial forensics. New York: Springer; 2011.

Coudron PE, Jones DL, Dalton HP, Archer GL. Evaluation of laboratory testing for detection of methicillin-resistant *Staphylococcus aureus* and *Staphylococcus epidermidis*. J Clin Microbiol. 1986;24:764–9.

Demirev P. Emerging instrumental methods for antimicrobial resistance and virulence testing. In: Havlicek V, Spizek J, editors. Natural products analysis: instrumentation, methods, and applications. New Jersey: Wiley; 2014. p. 25–42.

Demirev P, Fenselau C. Mass spectrometry for rapid characterization of microorganisms. Annu Rev Anal Chem. 2008a;1:71–94.

Demirev P, Fenselau C. Mass spectrometry in biodefense. J Mass Spectrom. 2008b;43:1441–57.

Demirev PA, Feldman AB, Lin JS. Bioinformatics-based strategies for rapid microorganism identification by mass spectrometry. Johns Hopkins APL Tech Dig. 2004;25:27–37.

Demirev PA, Hagan NS, Antoine MD, Lin JS, Feldman AB. Establishing drug resistance in microorganisms by mass spectrometry. J Am Soc Mass Spectrom. 2013;24:1194–201.

Du Z, Yahng AR, Guo Z, Song Y, Wang J. Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix assisted laser desorption ionization time-of-flight mass spectrometry. Anal Chem. 2002;74:487–91.

Edwards-Jones V, Claydon M, Evason D, Walker J, Fox A, Gordon D. Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. J Med Microbiol. 2000;49:295–300.

Fenselau C, Demirev PA, editors. Rapid characterization of microorganisms by mass spectrometry. Washington, DC: ACS; 2011.

Havlicek V, Lemr K, Schug KA. Current trends in microbial diagnostics based on mass spectrometry. Anal Chem. 2013;85:790–7.

Ho YP, Reddy PM. Advances in mass spectrometry for the identification of pathogens. Mass Spectrom Rev. 2010;30:1203–24.

Hrabak J, Chudackova E, Walkova R. Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. Clin Microbiol Rev. 2013;26:103–14.

Jackson KA, Edwards-Jones V, Sutton CW, Fox AJ. Optimisation of intact cell MALDI method for fingerprinting of methicillin-resistant *Staphylococcus aureus*. J Microbiol Methods. 2005;62:273–84.

Jung JS, Eberl T, Sparbier K, Lange C, Kostrzewa M, Schubert S, Wieser A. Rapid detection of antibiotic resistance based on mass spectrometry and stable isotopes. Eur J Clin Microbiol Infect Dis. 2014;33:949–55.

Kostrzewa M, Sparbier K, Maier Th, Schubert S. MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. Proteomics Clin Appl. 2013;7:767–78.

Mann M. Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol. 2006;7:952–8.

Marshall AG, Senko MW, Li WQ, Li M, Dillon S, Guan SH, Logan TM. Protein molecular mass to 1 Da by C-13, N-15 double depletion and FT-ICR mass spectrometry. J Am Chem Soc. 1997;119:433–4.

Meselson M, Stahl FW. The replication of DNA in *Escherichia coli*. Proc Natl Acad Sci U S A. 1958;44:671–82.

Muroi M, Shima K, Igarashi M, Nakagawa Y, Tanamoto K. Application of matrix-assisted laser desorption ionization-time of flight mass spectrometry for discrimination of laboratory-derived antibiotic-resistant bacteria. Biol Pharm Bull. 2012;35:1841–5.

National strategy on combating antibiotic-resistant bacteria, White House Executive Order 13676, Washington, DC, 2014.

Oda Y, Huang K, Cross FR, Crowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A. 1999;96:6591–6.

One source of globally isotope-labeled bacterial growth media, as well as individual stable-isotope labeled compounds for locally-labeled or minimal growth media formulations, is Cambridge Isotope Laboratories, Inc. (Andover, MA, USA).

Pierce CL, Rees JC, Fernandez FM, Barr JR. Detection of *Staphylococcus aureus* using N-15-labeled bacteriophage amplification coupled with matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. Anal Chem. 2011;83:2286–93.

Pierce CL, Rees JC, Fernandez FM, Barr JR. Viable *Staphylococcus aureus* quantitation using N-15 metabolically labeled bacteriophage amplification coupled with a multiple reaction monitoring proteomic workflow. Mol Cell Proteomics. 2012;11:M111.012849. doi:10.1074/mcp.M111.012849-5.

Pineda FJ, Lin JS, Fenselau C, Demirev P. Testing the significance of microorganism identification by mass spectrometry and proteome database search. Anal Chem. 2000;72:3739–44.

Rees JC, Voorhees KJ. Simultaneous detection of two bacterial pathogens using bacteriophage amplification coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2005;19:2757–61.

Rees JC, Pierce CL, Schieltz DM, Barr JR. Simultaneous identification and susceptibility determination to multiple antibiotics of *Staphylococcus aureus* by bacteriophage amplification detection combined with mass spectrometry. Anal Chem. 2015;87:6769–77.

Sandrin TR, Goldstein JE, Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom Rev. 2013;32:188–217.

Shah HN, Gharbia SE, editors. Mass spectrometry for microbial proteomic. New York: Wiley-Interscience; 2010.

Shah HN, Rajakaruna L, Ball G, Misra R, Al-Shahib A, Fang M, Gharbia SE. Tracing the transition of methicillin resistance in subpopulations of *Staphylococcus aureus,* using SELDI-TOF mass spectrometry and artificial neural network analysis. Syst Appl Microbiol. 2011;34:81–6.

Sparbier K, Lange Ch, Jung J, Wieser A, Schubert S, Kostrzewa M. MALDI biotyper-based rapid resistance detection by stable-isotope labeling. J Clin Microbiol. 2013;51:3741–8.

Stump MJ, Jones JJ, Fleming RC, Lay JO, Wilkins CL. Use of double-depleted C-13 and N-15 culture media for analysis of whole cell bacteria by MALDI time-of-flight and fourier transform mass spectrometry. J Am Soc Mass Spectrom. 2003;14:1306–14.

Tang Y-W, Stratton C, editor. Advanced techniques in diagnostic microbiology. New York: Springer; 2006.

Wilkins CL, Jackson O, Lay JO, editors. Identification of microorganisms by mass spectrometry. New York: Wiley; 2005.

Wolters M, Rohde H, Maier T, Belmar-Campos C, Frank G, Scherp S, Aepfelbacher M, Christner M. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. Int J Med Microbiol. 2011;301:64–8.

World Health Organization. Antimicrobial resistance: draft global action plan on antimicrobial resistance, A68/20, March 27, 2015.

Index

P. Demirev, T. R. Sandrin (eds.), *Applications of Mass Spectrometry in Microbiology,*
DOI 10.1007/978-3-319-26070-9