

1

A Paradigm Shift from Research to Front-Line Microbial Diagnostics in MALDI-TOF and LC-MS/MS: A Laboratory's Vision and Relentless Resolve to Help Develop and Implement This New Technology amidst Formidable Obstacles

Haroun N. Shah¹ and Saheer E. Gharbia²

¹ Proteomics Research and ² Genomic Research Unit, Public Health England, London, UK

¹ Current address: Department of Natural Sciences, Middlesex University, London, UK

1.1 Introduction

The outlooks of a microbial systematics research laboratory and a clinical diagnostic laboratory are scientifically and logistically polarized, due almost entirely to the nature of their approach to generating knowledge and translating technological tools. Research laboratories are generally sustained through scientific grants, and their focus is driven by development and applications of novel and emerging technologies; their direction is more fluid and innovative. By contrast, the clinical diagnostic laboratory is more circumspect and notoriously resistant to change in methodology and mindful of the meticulous level of validation and accreditation needed to implement new workflows. Thus, over the years, whereas researchers exploited biochemical and chemotaxonomical advances, the traditional clinical diagnostic laboratory remained largely disinclined.

The arrival of comparative 16S rRNA sequence analysis heralded a new era of microbiology and allowed for the first time the classification of microorganisms along phylogenetic lines. This began in the 1970s with the introduction of rRNA cataloguing, then reverse transcriptase sequencing of the 5S, 16S and 23S rRNA subunits and, finally, with the arrival of PCR and sequencing of genes in the 1990s. The impact of comparative 16S rRNA sequence analysis enabled the biggest change witnessed in microbial systematics for over a century and led to the assembly of the largest database in the history of life sciences. Although at the higher taxonomic level, the topography of many of the lineages remained stable, substantial changes in classification and nomenclature were being recorded at the genus and species levels. The need to study intra-generic and intra-species diversity became essential. Clinical microbiology was at a crossroads; it could not ignore the increasing volume of literature, and began slowly to embrace these changes.

1.1.1 Personal Experience at the Interface of Systematics and Diagnostics

The Public Health Laboratory Service (PHLS) established its first dedicated Molecular Identification Services Unit (MISU) in late 1997 with a remit to find novel methods to identify atypical, rarely isolated and emerging taxa that clinical laboratories failed to identify. PHLS at that time comprised a network of some 50 specialist laboratories in England and Wales that served as a national referral centre for the characterization of human pathogens that could not be delineated by clinical laboratories. In spite of the specialist nature of PHLS laboratories, a significant proportion of samples remained *incertae sedis* and were left unreported. MISU's remit was to address this problem, and one of the authors (HNS), joined PHLS to spearhead reshaping diagnostics along a systematic and technology framework. Despite years of experience (from the early 1970s) in the field of microbial systematics, particularly with the *Bacteroidaceae* using biochemical, chemical and molecular methods, the implementation of 16S rRNA from a research laboratory to a daily service function was a daunting challenge and only became a routine procedure by establishing a workflow from sample to sequence and data in hours rather than in days, the timeline associated with multiple steps in amplification and DNA sequencing. This was facilitated by the new Applied and Functional Genomics Unit (headed by author SEG in 2002) and provided the momentum for the service to be accredited. Over the years, a large number of novel genomics and proteomic technologies were explored by both laboratories and served as a platform to introduce new technologies and their potential applications. Matrix-assisted laser desorption/ionization (MALDI)-time-of-flight mass spectrometry (MALDI-TOF MS) was the first of these and dates back to its inception in October 1997.

1.1.2 MALDI-TOF MS: The Early Years

The arrival of MALDI-TOF MS coincided with attempts to introduce 16S rRNA as a diagnostic method in PHLS. The almost simultaneously publication of work from several laboratories – for example, Claydon *et al.* (1996); Cain *et al.* (1994); Holland *et al.* (1996) and Krishnamurthy and Ross (1996) – highlighted its potential to create a pattern derived from bacterial cells that discriminated between fairly disparate species. However, it was given little attention by clinical microbiologists, largely because of the success, ease and by now, the availability of public databases of 16S rRNA sequences. This was in direct contrast to MALDI-TOF MS, for which there was only a general bench-top instrument, no standard protocol nor more than a few dozen MALDI-TOF MS spectra that were published mainly in MS journals.

The motivation to explore the application of this technology by the author stems from the need to find new methods to identify the unusual range of taxa that were received for identification for which even 16S rRNA results were equivocal. Having previously reported extensive work on characterization of lipids by MS (see, e.g. Shah and Collins, 1980; Shah and Collins, 1983; Shah and Gharbia, 2011), the potential use of MALDI-TOF MS, which had the capability to ionize molecules orders of magnitude greater in mass, was an intriguing challenge and was explored initially using species of the genus *Porphyromonas*. A key character of this group is their inability to ferment carbohydrates, the primary means of identifying bacterial species at that time. DNA-DNA reassociation and lipid analyses by MS were the only

reliable methods to delineate members of this genus, but this was undertaken by only a few specialist laboratories and was lengthy, tedious, required a large biomass (>50 mg of cells) and was limited in scope (see, e.g. Shah *et al.*, 1982). By contrast, when *Porphyromonas* species were subjected to analysis by MALDI-TOF MS in early 1998, they were readily delineated using just a few cells directly from an agar plate (see Figure 1.1; see Shah *et al.*, 2000; Shah *et al.*, 2002). This watershed moment not only provided proof of concept but gave such confidence in the potential of this technology that MISU would go on to painstakingly pursue its development for nearly two decades (Shah, 2005).

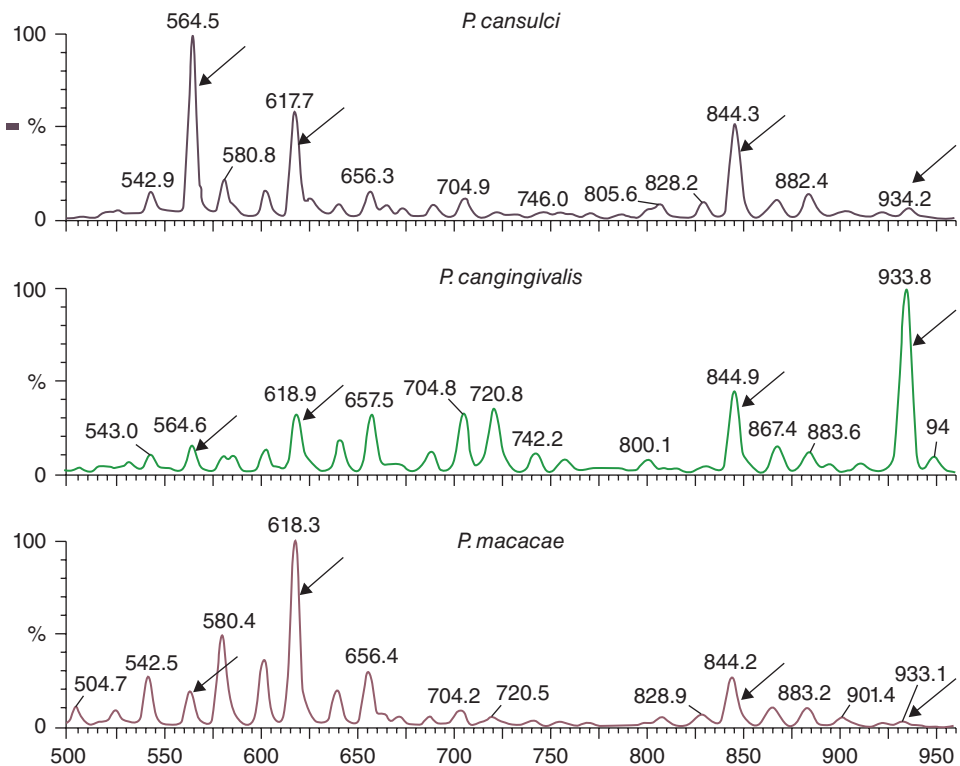


Figure 1.1 Examples of the distinctive MALDI-TOF-MS profiles of intact cells of *Porphyromonas* sp. containing both genus-specific (e.g. 618 and 844 Da mass ions) and also a significant number of species-specific mass ions (examples indicated by arrows). Members of the genus *Porphyromonas* now comprise 18 species, in addition to several others that have not yet been validated (Bergey's manual, 2011). However, with the exception of DNA/DNA reassociation, they could not be reliably delineated at the time. The three representative MALDI-TOF-MS spectra shown and those reported earlier (Shah *et al.*, 2002) revealed that each species could be unambiguously distinguished. It was this poorly characterized group of anaerobes that became one of the compelling forces for the development of this technique for microbial identification. Early meetings to demonstrate an appreciation of the technology were mostly presented at meetings on anaerobic taxa and helped rejuvenate interest in this area of microbiology again (see Chapter 5 and Figure 1.4).

1.1.3 The Formidable Challenge to Gain the Confidence of the Clinical Microbiologist in MALDI-TOF MS

At its commencement, many microbiologists pointed to the dismal failure of pyrolysis MS (PyMS, circa 1970–1990) as a clinical diagnostic tool and simply branded MALDI-TOF MS as a newer version of PyMS that generated a mass spectral profile of unidentified proteins of intact cells. However, a minority believed that MALDI-TOF MS was inherently superior, had more scope for development and took up the mantle to unlock its potential. Our laboratory was one of the forerunners, which embraced the drive to develop and implement MALDI-TOF MS for microbial identification.

The proposal to introduce MALDI-TOF MS for diagnosis in a molecular-based service for human infectious diseases seemed preposterous to many. Efforts to develop MALDI-TOF MS and proteomics, in particular, were met with fierce resistance both from within and outside PHLs. Consequently, no core funding was allocated for its development throughout the years; support came from scientific grants, industrial funding and collaboration with various MS companies. Initially, Kratos Analytical, Manchester, UK, which designed and built the first bench-top linear MALDI-TOF mass spectrometer, the Kompact Alpha™ mass spectrometer, placed this instrument in MISU between 1998 and 1999 to help develop the methodology and explore the potential applications of the technique (Figure 1.2). To bring it to the attention of the wider scientific community, the first conference in the field entitled ‘Intact Cell MALDI: A Novel Technique for the Rapid Identification of Microorganisms’ was held at PHLs, London on 27 October 1998 and was attended by some 150 scientists (see Figure 1.3).

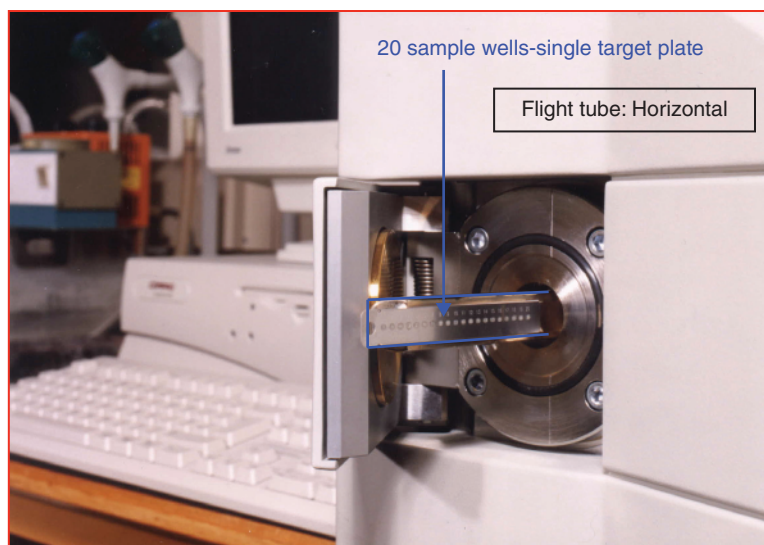


Figure 1.2 The first bench-top linear MALDI-TOF mass spectrometer, the Kratos Kompact Alpha (Kratos Analytical, Manchester, UK), that inspired the development of the technology for clinical microbiology. The target plate, with a capacity for 20 samples, is shown and was used for generating the spectra shown in Figure 1.1. The instrument revealed the potential of the technology, but it was manual and unsuitable for a microbial diagnostic laboratory.

INTACT CELL MALDI

Microbiologists are constantly searching for techniques to improve current identification systems. Traditionally biochemical tests have provided the basis for assigning unknown isolates to known taxa.

These were later complemented and refined by chemotaxonomic criteria such as peptidoglycan composition, polar and non-polar lipid analyses and more recently nucleic acid techniques.

Ideally microbial identification should be rapid, reliable, require minimal sample preparation and utilise the least number of bacterial cells. The MALDI technique encompasses many of these attributes in that analysis can be achieved in a few minutes using only a single bacterial colony of intact cells and with very little sample preparation being required. The sample is mixed with a matrix solution which enables a gentle ionisation process to occur. These ions are then analysed with a time-of-flight mass spectrometer. For microbial samples, these ions are most probably produced from desorbed components of the cell envelope. Studies have so far shown that the spectral patterns obtained provide characteristic fingerprints of known species and can go beyond the species level.

This meeting aims to increase microbiologists' awareness of advances in this field, the potential of the technique, current applications and future developments. Demonstration of the procedure and analysis will also be undertaken.

MEETING CONTACTS

For further information contact one of the following:

K Rolphson, (Kratos Analytical)	H N Shah, (PHLS)
tel: +44(0)161 888 4400 ext: 229	tel: +44(0)181 200 4400 ext: 3749
fax: +44(0)161 888 4402	fax: +44(0)181 205 7483

or e-mail any of the following:

K Rolphson: kathryn.rolfson@kratos.co.uk
 H N Shah: hshah@phls.co.uk
 T Cheasty: tcheasty@phls.co.uk
 M Claydon: m.claydon@mmu.ac.uk

A ONE DAY MEETING

organised by

Kratos Analytical, Manchester
 Bioanalytical Research Centre, Manchester Metropolitan University
 International Diagnostic Group, Bury, Lancs
 in collaboration with
 PHLS Central Public Health Laboratory, London

**INTACT
CELL
MALDI**

**A Novel Technique for the Rapid
Identification of Micro-Organisms**

to be held at
**PHLS Central Public Health Laboratory,
61 Colindale Avenue, London NW9 5HT**

27TH OCTOBER 1998

Figure 1.3 The conference leaflet of the first meeting held on 27 October 1998 to explore the potential application of MALDI-TOF mass spectrometry for microbial identification. These meetings were held annually to showcase developments in the field. However, it still took a decade of research and development for this technology to gain widespread acceptance in the clinical microbiology laboratory.

The authors' presentation titled 'A Review of the Current Methods of Bacterial Identification: MALDI-TOF MS in the Characterisation of the Obligate Anaerobes *Fusobacterium* and *Porphyromonas*' outlined the paucity of reliable characters for delineating many taxa within the *Bacteroidaceae* and presented early data using MALDI-TOF MS as proof of principle of the technique.

Although there were intermittent reports in the scientific community on the analysis of specific taxa, there was neither a generally accepted method nor a coherent plan to establish a universal approach to using MALDI-TOF MS as a diagnostic tool. Furthermore, major concerns were raised at the meeting about introducing a new technique to analyze human pathogens involving use of intact, viable cells. We carried out thorough investigations on the safety aspects of the procedures and produced a detailed application note for Kratos Analytical in 1999. To verify this, tough resistant spore-formers, *Bacillus stearothermophilus* (NCTC 10003) and *Bacillus subtilis* (NCTC 10073), which are industry standard controls for heat and

chemical sterilization, were processed through the instrument at various concentrations and samples collected at potential leakage points within the instrument and vacuum pumps. No growth was observed in any sample, and this, together with other successful safety tests, led to the Health and Safety Executive granting approval for the use of MALDI-TOF MS in a clinical laboratory and permission for engineers to carry out repairs *in situ* when necessary. MISU organized several hands-on workshops for service engineers from MS companies to alert them to the hazards of undertaking fieldwork.

However, in spite of careful assessment of the risk, one major accident occurred in the laboratory in 2000 that nearly terminated all future work at PHLS. The standard method used for cleaning target plates at the time was sonicating for half-hour intervals in methanol followed by washing in 33% (w/v) aqueous nitric acid, and final rinsing in distilled water (e.g. Evason *et al.*, 2001). If the methanol and nitric acid are mixed and kept in a sealed container, the reaction is explosive. This error occurred, and a huge accident resulted in the laboratory being closed for several months and the project nearly being terminated.

However apart from safety issues, the mere separation of a few disparate species by differences in MALDI-TOF MS spectra was insufficient to introduce a new technology for clinical diagnostics. Instrument design, robust sample preparation among the varied bacterial chemotypes, reproducible mass spectra on a single instrument as well as other instruments and development of a database were among the major challenges faced at the time. From these initial stages, we set out to explore the various forms of MALDI-TOF MS and later on, tandem MS/MS proteomics for the characterization of human pathogens. Annual international conferences were organized to showcase developments at each stage together with horizon scanning for future work (see Table 1.1).

1.2 Overcoming the Variable Parameters of MALDI-TOF MS Analysis: Publication of the First Database in 2004

Between October 1997 and December 1999, we began to explore the diagnostic potential of MALDI-TOF MS with Manchester Metropolitan University using the Kratos Kompact AlphaTM mass spectrometer. Samples were taken from broth and plate cultures with cells from various growth phases. We analyzed intact and broken cells preparations, fractions from gradient centrifugation steps, and so on. Several matrix solutions, at a range of concentrations, were used to derive a reproducible MALDI-TOF MS spectrum. These included 5-chloro-2-mercaptobenzothiazole (CMBT), α -cyano-4-hydroxycinnamic acid (α -Cyano), sinapinic acid and ferulic acid. Several matrices were known at the time to improve signal reproducibility (e.g. Gusev *et al.*, 1995), and it soon became common practice to use CMBT for gram-positive and α -Cyano for gram-negative bacteria. We tested instrumental parameters, some of which were inflexible; for example, the laser was set at 337 nm, and the pulse width was fixed at 3 ns. However, other parameters such as the voltage could be altered, and this was rigorously tested and eventually set at 20 kV accelerating potential. Initially, data was collected over a wide range of m/z values, but the highest density of mass ions was

Table 1.1 International conferences organized by the Molecular Identification Services Unit (MISU) and Applied and Functional Genomics Unit (AFGU) to showcase work achieved and future directions in proteomics using MALDI-TOF MS and tandem MS/MS and genomics. (MISU and AFGU were amalgamated in September 2009 in the new Department of Bioanalysis and Horizon Technologies.)

Date	Title
27 October 1998	1st : Intact Cell MALDI – A Novel Technique for the Rapid Identification of Microorganisms
14–15 June 1999	2nd : The Impact of the Environment on Human Infections through Molecular and Mass Spectrometric Analyses
17–18 April 2000	3rd : Microbial Characterisation, Diversity and Function through Genome and Proteome Analysis
25–27 June 2001	4th : Decoding the Microbe Using Advanced Tool of Genomics, Transcriptomics and Proteomics
1–2 July 2002	5th : Applications of Biomics (Genomics, Proteomics and Bioinformatics) in the Research and Diagnostic Laboratory
Workshops: 23–24 June 2004	(Workshops in genomics, proteomics and bioinformatics)
16–17 June 2003	6th : Disease Biomarkers and Polymorphisms in Microbes
21–22 June 2004.	7th : Meeting the Challenges of Infectious Diseases through advances in Developing Technologies
Workshops: 23–24 June 2004	(Workshops in genomics, proteomics and bioinformatics).
13–14 June 2005	8th : Sequenced-Based Approaches to Diagnosis of Infectious Disease Agents
25–26 September 2006	9th : Development and Application of High Throughput Systems in Diagnostic Microbiology
18–19 September 2007	10th : The Changing Landscape of Diagnostic Microbiology; from Decades of Traditional Methods to Applied Genomics and Proteomics
26–27 June 2008	11th Use of New Technologies to Further Understand the Biology of Transient and Host-Derived Human Pathogens
25–26 June 2009	12th : Target Molecules and Biomarkers in the Characterisation of Microbes in Disease and the Environment
24–25 June 2010	13th : Microbial Infections: Novel Approaches to Looking at Old Problems
23–24 June 2011	14th : Exploration of Novel Technologies for Biomarker Discovery and Point of Care Diagnostics
21–22 June 2012	15th : Unlocking Genomic and Proteomic Signatures for Functional Characterisation of Human Pathogens
Supplementary meeting: 4–5 April 2012	Microbial Diagnostic Applications of Mass Spectrometry (Jointly between PHE and the University of Minho, Portugal)
27–28 June 2013	16th : Microbial Subtyping in Disease and the Environment; the Pivotal Role of Reference Collections
26–27 June 2014	17th : The Power of the Genome and Proteome in Public Health Interventions (Joint meeting between PHE and The Royal College of Pathologists)
25–26 June 2015	18th : Applications of High Throughput Genomics and Proteomics in Infection
23–24 June 2016	19th : Next Generation Genomics and Proteomics, Advances in Microscale Analysis (Joint meeting between PHE, Animal and Plant Health Agency and Middlesex University, London)

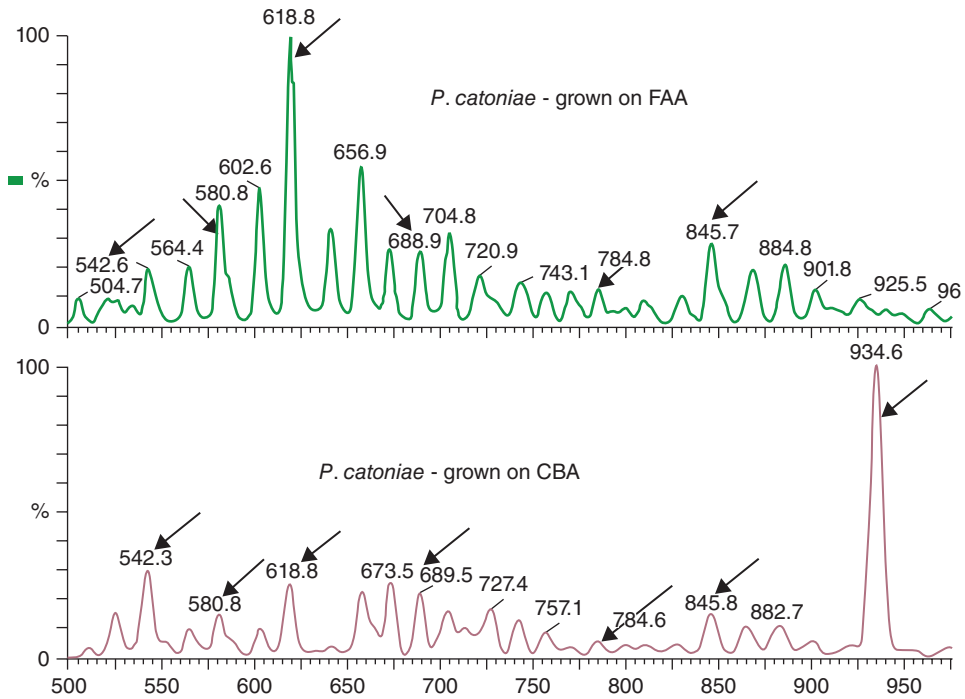


Figure 1.4 Changes in the MALDI-TOF-MS profile of the same strain of *Porphyromonas catoniae* (NCTC 12856) grown on Fastidious Anaerobic Agar (FAA) and Columbia Blood Agar (CBA). Many of the significant mass ions, e.g. 542, 580, 618, 689, 784 and 845 Da, are retained. However, significant mass ions such as 935 Da are present only in cells grown on CBA.

found to lie within the m/z range of 500 to 10,000kDa. Thus, most samples were analyzed within this range. This was subsequently extended to 20,000kDa. Once these constraints were set, various parameters that affect the stability of the mass spectrum were meticulously investigated. Media types and suppliers were tested (see Figure 1.4; Shah *et al.*, 2000). The effect of spores on ionization, impact of temperature, pH, growth phase and cell density were all investigated and used to derive a standard protocol. Data analysis was achieved by using a modified Jaccard coefficient and UPGMA to analyze and interpret interrelationships between strains. At each stage, the progress made was incorporated into an annual report presented at a series of international conferences (Table 1.1).

Following considerable preliminary work using this instrument, recommendations for improvement were made in 1999. The base of the instrument was quite large because of the need to accommodate the horizontally positioned flight tube. A suggestion to reposition the flight tube vertically would shrink its width and so enable the instrument to fit into the normally cramped space of a clinical laboratory. Other recommendations included the need to automate the many operational steps and redesign the fragile 20-well target plate (Figure 1.2), two of which were used as standards for lock mass corrections while those at the extreme ends of the target plate yielded unreliable results.



Figure 1.5 The first dedicated linear MALDI-TOF mass spectrometer for microbial identification used for creating the first MALDI-TOF MS database in 2004 (Keys *et al.*, 2004). The instrument, manufactured by Micromass (Manchester, UK), remedied many of the shortcomings of the Kratos Kompact Alpha and had the capacity to analyze 96 samples automatically.

A new instrument with the designed features described above was built by Micromass (Manchester, UK) in late 1999 and became the first upright bench-top linear and reflectron MALDI-TOF MS instruments (see Figure 1.5). The linear TOF instrument designated M@LDI was delivered to our laboratory on 12 April 2000, on a long-term loan to develop its applications.

This was two working days before the third annual conference on 17–18 April 2000 (see Table 1.1) and just prior to the Congress of the Confederation of Anaerobic Societies, Manchester, 10–12 July 2000 (Figure 1.6A). Work on poorly characterized anaerobic species was presented by author HNS in 2001 at the symposium ‘An Anaerobe Odyssey’ in UCLA, Los Angeles, CA, to demonstrate the optimism that was already developing (Figure 1.6B). PHLS underscored this momentous event in its 2001 year-book, during which these early results were highlighted. This instrument, reconfigured with a vertical flight tube, utilized a solid stainless steel 96-well target plate which comprised rows of 12 sample wells with a central position between each of four wells for lock mass correction. Author HNS incorporated it into the clinical services of the Identification Service Unit, which published its first promotional flyer in 2000, a decision that emanated from its confidence in the use of MALDI-TOF MS for microbial identification (Figure 1.7).

(A)

ANAEROBE 2000; the first Congress of the Confederation of Anaerobic Societies: 10–12 July, 2000 in Manchester, England

Society for Anaerobic Microbiology (UK); Anaerobe Society of the Americas, Japanese Association for Anaerobic Infection Research



**Micromass Satellite Symposium
New Techniques in Mass
Spectrometry
Tuesday, July 11, 2000**

Micromass UK Ltd. will offer a seminar on “New Techniques in Mass Spectrometry” during the Lunch Period on Tuesday, July 11, 2000. The session will be presented by Prof. Haroun N. Shah and will address how mass spectrometry techniques (MALDI-TOF MS) enable bacteria to be identified with unprecedented selectivity and speed. This workshop is intended both to illustrate the potential of these emerging typing strategies and stimulate debate on their practicability.

(B)

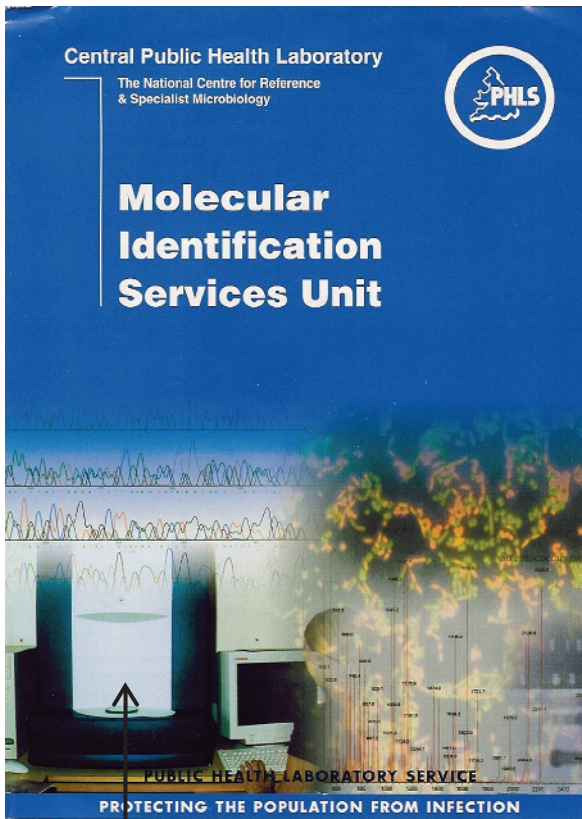
**2001: An Anaerobe Odyssey
August 12, 2001
UCLA Faculty Center, Los Angeles, CA**

The **Anaerobe Society of the Americas** and the **Infectious Disease Association of California** present **2001: An Anaerobe Odyssey** at the UCLA Faculty Center in Los Angeles, CA. The symposium celebrates the 80th birthday of **Dr. Sydney M. Finegold**, noted anaerobe researcher and Founding President of the Anaerobe Society.

Program:

- **Ellen Baron:** *Speculations on the Microbiology Laboratory of the Future*
- **Diane Citron:** *Changes in the Genus Fusobacterium*
- **Sydney Finegold:** *The Possible Role of Intestinal Flora in Autism*
- **Ellie Goldstein:** *Therapy of Anaerobic Infections*
- **David Hecht:** *Evolution of Anaerobic Susceptibility Testing*
- **Hannele Jousimies-Somer:** *The Taxonomic Evolution of Gram-negative Anaerobes*
- **Carl Nord:** *The Status of Research on Anaerobes in Europe*
- **Haroun N. Shah:** *MALDI-TOF Mass Spectrometry and Proteomics: A New Era in Anaerobic Microbiology*
- **Dennis Stevens:** *Clostridial Toxins*
- **Kazue Ueno:** *The Status of Research on Anaerobes in Japan*
- **Hannah Wexler:** *Outer Membrane Proteins of Gram-negative Anaerobes*

Figure 1.6 Early meetings on MALDI-TOF MS were focused on poorly defined anaerobic species to emphasize, even at that time, the resolution of the method. Many of these species are non-fermentative so that methods involving the use of API or various biochemical tests, which were the primary methods then, were negative. New species were proposed mainly on the basis on DNA-DNA reassociation, which could not be applied in a clinical laboratory. MALDI-TOF MS when introduced contributed significantly to a resurgence of interest in anaerobic microbiology because of its capacity to resolve such complex taxonomic problems. (A) As early as the year 2000, a specific symposium was held to demonstrate the high resolving power of MALDI-TOF MS in delineating poorly defined anaerobic species in the United Kingdom and (B) in 2001, in the United States.



MALDI-TOF-Mass Spectrometer– (2000)

Figure 1.7 The first promotional service flyer for the Molecular Identification Services Unit (MISU, PHLS), which was printed and advertised from 2000. It signalled a strong declaration of MISU's vision to elevate MALDI-TOF MS as its principal method of microbial identification by placing a M@LDI-TOF mass spectrometer (Micromass, Manchester, UK) in the central position on the flyer.

With the basic parameters for a microbial database established, our major goal over the following five years was to develop a database of mass spectral profiles. Funding for this came through a collaborative multicentre venture with Manchester Metropolitan University and Micromass (Floats Road, Manchester) for a five-year intensive programme. The first year of the project was used to meticulously optimize protocols and interrogate the software and search engine. The National Collection of Type Cultures (NCTC) is part of the PHLS, and its vast range of type and reference strains were used as the resource for creating the database. For development of the database, NCTC's strains were sent out independently and blindly to each of the three collaborating laboratories for analyses. Twenty strains per week were analyzed over 50 weeks annually with each strain being analyzed 12 times at each site. Thus, during the first year, 36,000 spectra were collected and analyzed. Operation of the mass spectrometer was

performed using a specially designed software designated MassLynx™. Automated calibration of the TOF tube, followed by automated acquisition of the bacterial spectra, was then performed using the real-time data selection (RTDS) function in the MassLynx™ software. Spectral profiles were collected in the mass range 500–10,000 kDa, acquiring 10 shots per spectrum at a laser firing rate of 20 Hz. Fifteen spectra per sample well and 10 spectra per lock mass well were collected using the RTDS option to optimize the collection of quality data. For database inclusion, the spectral reproducibility between the 12 replicates per sample was tested using a root mean square (RMS) calculation to identify and reject outliers at a value greater than 3.0. The RMS is the normalized deviation of the median of test spectra from the spectral average and therefore was used to compare each replicate spectrum in turn to the composite spectra of the remaining replicates. All verified spectra were combined to produce a composite spectral entry for each bacterium included in the database. The system was challenged with unknown and clinical isolates. Database searching was based on an estimation of the probability that the mass spectral peaks in the test spectrum are comparable with the database spectrum. A list of the top matches was provided together with RMS values. A high probability and low RMS value indicated a good match. The correlation between results at each laboratory was so high that after the first year, the compilation of the database was done independently and added to a composite database. The results were reported in 2004 when the database comprised 3500 spectra (see Keys *et al.*, 2004).

While this work was in progress, our laboratory, MISU, began using the instrument in tandem with 16S rRNA sequencing as part of its service function, and numerous unusual isolates that would hitherto be left unreported were now being identified (Figure 1.7). For example, the receipt of an unusual isolate from a patient who developed a severe wound infection after a visit to the Dead Sea was erroneously identified on primary culture as *S. aureus* because of its striking morphological resemblance to this species. However, its MALDI-TOF MS spectrum was significantly different, triggering further analysis, and it was subsequently identified as *Exiguobacterium aurantiacum* using 16S rRNA (Mohanty and Mukherji, 2008). The inclusion then of *E. aurantiacum* into the database enabled rapid identification of this species and prompted clinicians to send in samples from patients with similar symptoms. Within six months, 18 patients with bacteraemia were shown to have *E. aurantiacum*. A subsequent study using MALDI-TOF MS rapidly identified strains of this alkaliphilic, halotolerant bacteria from six patients with bacteraemia, three of whom had myeloma (Pitt *et al.*, 2007). Some of most difficult species received for identification by MISU belonged to the Acinetobacter, Kingella and Moraxella complex. These were extensively studied and delineated very early in this study and reported at the American Society for Mass Spectrometry 48th meeting in 2000.

Work continued on microbial identification using MALDI-TOF MS, with the database now containing nearly all the NCTC's type and reference strains. The next phase of the work involved the trial of the instrument in a hospital laboratory. Through The Royal London Hospital (University of London), it was possible to gain access to primary routine cultures over several weeks. Hospital staff were trained in sample preparation and analysis, and over 600 samples were analyzed. Because of the problems associated with MRSA and *C. difficile* infections at the hospital, priority was given to presumptive

isolates of these species. Although it was not possible to separate MRSA from sensitive strains, all samples were correctly identified to the species level (Rajakaruna *et al.*, 2009). However, the work suffered a major setback by the failure to confidently identify isolates of *C. difficile*. This was remedied through a collaboration which began in 2007 with a small company, AnagnosTec GmbH (Potsdam OT Golm, Germany), and led to a change of matrix solutions and formic acid extraction of samples prior to MALDI-TOF MS analysis. Meetings organized by AnagnosTec were held annually for four years in Potsdam Golm and comprised a small group of about 20–30 participants from various parts of Europe who had a vested interest in implementing MALDI-TOF MS as a diagnostic tool (Figure 1.8). Most of the motivation to establish MALDI-TOF MS in clinical laboratories was concentrated in Europe, as the United State's FDA approval seemed a distant goal. This is reflected in the authorship of the book *Mass Spectrometry for Microbial Proteomics* (Wiley, 2010).

Initially, the instrument was set up to profile the surface molecules of cells, the rationale being that differences between virulent and avirulent strains, where the pathogenic potential is due to surface-associated molecules, could be mapped and used for detection of pathogenic variants. For some species such as *Peptostreptococcus micros*, this was highly successful, where resolution of the two pathotypes ('smooth' and 'rough' variants) were readily distinguishable through characteristic mass ions (see Rajendram, 2003). However, to obtain such mass spectra, it was necessary to use rigorously standardized parameters that may alter the morphology of cells grown on agar plates. In our experience, a more fruitful way to approach this is the use of MALDI-TOF MS with ProteinChip Arrays (designated surface-enhanced laser desorption/ionization TOF MS, or SELDI-TOF MS).

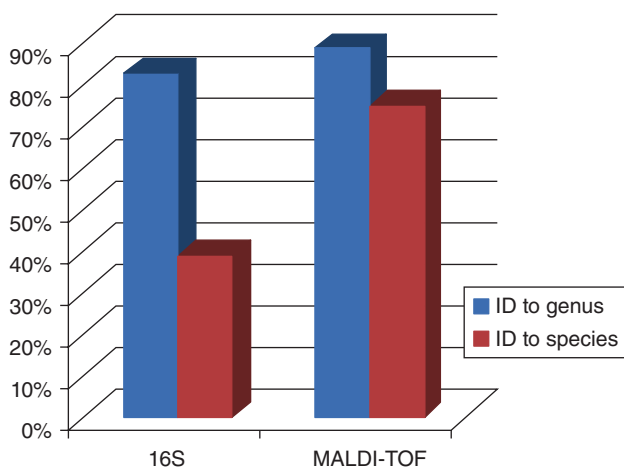


Figure 1.8 Identification of clinical isolates received by MISU using 16S rRNA and MALDI-TOF MS (Micromass, UK) over a 10-year period. MISU receives atypical, rarely isolated and emerging pathogens, and for such unusual isolates, the results shows that MALDI-TOF MS was significantly more useful in identifying these isolates to the species level.

1.3 SELDI-TOF MS: A Powerful but Largely Unrecognized Microbiological MALDI-TOF MS Platform

Microorganisms constantly undergo microevolution, and the expression of these changes is reflected in proteins; consequently, we envisaged that these biomolecules would be a rich source of biomarkers. Because proteins dictate virtually all the biological functions of a cell, we proposed to systematically catalogue cellular proteins of key pathogens and investigate how they change in response to disease and environmental factors including exposure to antibiotics. Initially, some investigators attempted to deduce components of the bacterial cell envelope using linear MALDI-TOF MS (see, e.g. Claydon *et al.*, 1996). We were acutely aware at the commencement of this programme that this instrument would only provide diagnostic profiles for microbial identification. To unravel the microbial proteome, various forms of higher-resolution MS would be necessary and, at that time, several gel-based methods were also indispensable prior to MS analysis.

Through a collaborative programme with Ciphergen Biosystems, we acquired the ProteinChip® System, designated SELDI-TOF MS. The ProteinChip technology has the capacity to rapidly perform separation, detection and analysis of proteins (at the femtomole level) directly from biological samples. Key components of the technology are ProteinChip arrays that comprise minute chromatographic wells that contained either a chemical (anionic, cationic, metal ion, hydrophobic, hydrophilic, etc.) or biochemical (antibody, receptor, DNA, etc.) surfaces to capture specific classes of proteins from a crude sample (see Figure 1.9). In practice, a few microlitres of a cell extract may be dispensed onto the ProteinChip surfaces, a quick wash is performed to remove unbound proteins and interfering substances and the sample allowed to air-dry for a few minutes. The matrix solution (sinapinic acid) is then added and samples analyzed by MALDI-TOF MS using the Ciphergen Biosystems instrument. The resulting molecular ions of the proteins, which remained bound to the ProteinChip surface, are deduced in minutes from its mass spectrum (see Figure 1.9). Because of the chemistry of the ProteinChip, the system has the capacity to capture a range of biomolecules. Our first applications of the technology for characterization of microbes were presented at the fourth of the annual conferences (Table 1.1) together with other potential applications. For example, antibodies may be covalently immobilized onto the ProteinChip Array surface by an initial incubation and washing, after which antigens may be specifically captured and analyzed directly to determine their intact masses. The potential to replace many conventional ELISA systems therefore existed. Furthermore, for epitope mapping, the bound proteins may be enzymatically digested by on-chip incubation with endoproteases. Following a wash to remove the unbound fragments, SELDI analysis may be used to identify the retained fragments and so enable characterization of the epitope. Through various grant-funded projects, MISU characterized a number of pathogenic determinants such as the botulinum toxins to find alternatives to animal experiments, determined pathotypes of *Peptostreptococcus micros* (Rajendram 2003), adhesins of *Enterococcus faecalis* (Reynaud *et al.*, 2007), etc. (see review by Shah *et al.*, 2010).

The potential of SELDI-TOF MS has been recognized by only a few microbiologists and consequently was never fully exploited. As early as 2001, the SELDI-TOF MS technology was interfaced with the Applied Biosystems QSTAR Hybrid LC/MS/MS to

Overview of the SELDI process

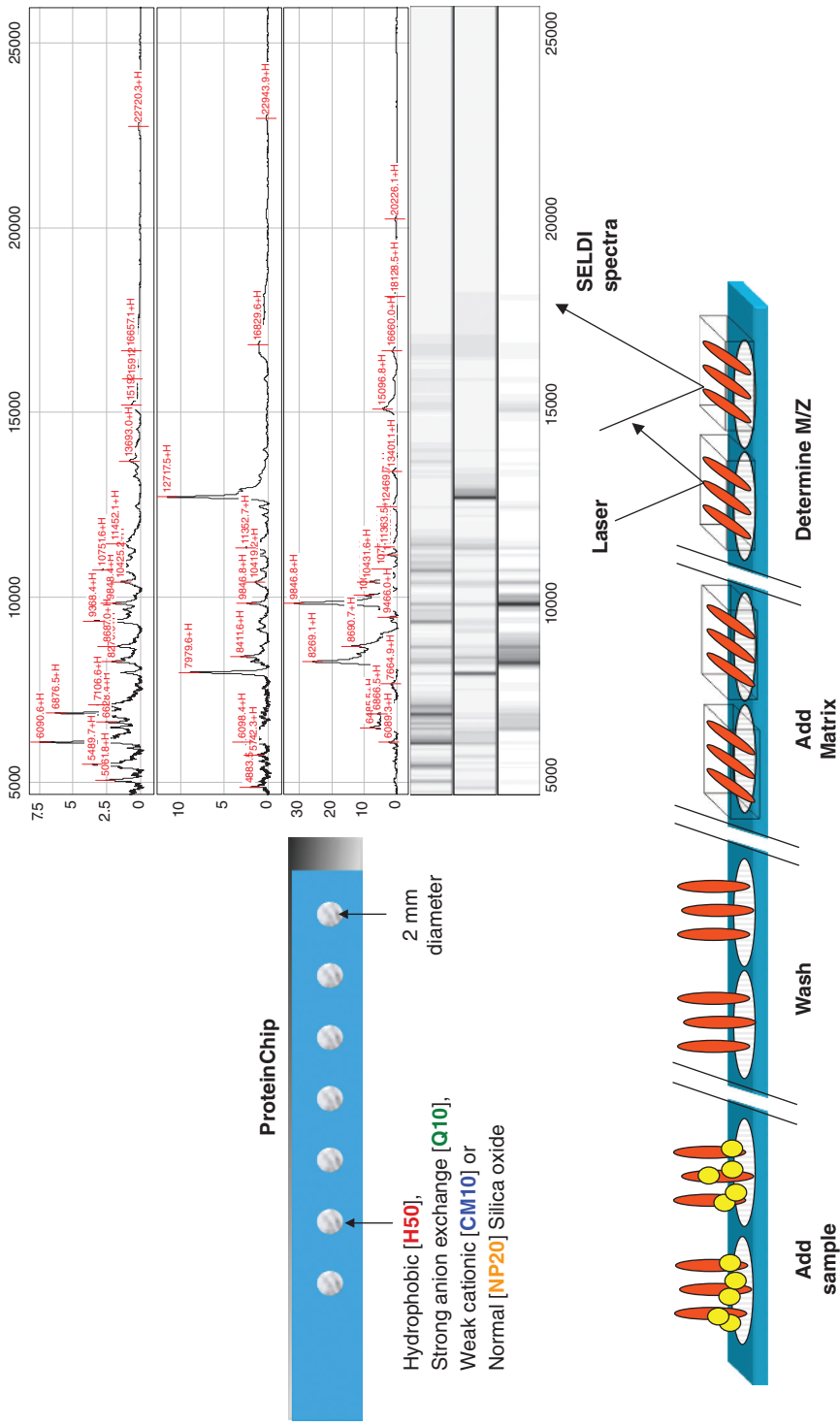


Figure 1.9 Some of the most popular types of ProteinChip arrays (H50, Q10, CM10 and NP20) used for analysis of microbial cells extracts and an overview of the process to obtain a mass spectral profile. From left to right, the sample is added to the ProteinChip array, the wells washed and air-dried for a few minutes, the matrix (Sinapinic acid) added, followed by mass spectral analysis to yield the spectra shown.

identify biomarkers with an upper limit of <4kDa. This significantly expanded the scope of SELDI to include characterization of proteins, protein interactions and structural analysis. Had this developed further, SELDI-TOF would have had a greater presence in microbiology. Ciphergen Biosystems was acquired by Bio-Rad in 2006 and with their long-standing interest in characterizing microbes using SDS-PAGE methods, this again should have stimulated interest among microbiologists. The company phased out the early instruments, and customers were left with mass spectrometers that could not be serviced; interest rapidly declined. In July 2009, a similar attempt was made to combine Bio-Rad's ProteinChip (SELDI)-based technology with Bruker's ultrafleXtreme MALDI-TOF/TOF mass spectrometer, but as before, this had little impact among microbiologists and interest rapidly declined. As a typing tool, we believe that systems such as SELDI-TOF MS offer considerable advantage over traditional MALDI-TOF MS (see review, Shah *et al.*, 2005) and used this approach recently to delineate the heterogeneity of *Propionibacterium acnes*. For the first time, mass ions derived from both MALDI- and SELDI-TOF MS were used as new criteria to propose two new subspecies (see Chapter 5; Dekio *et al.*, 2015) and supports the confidence MALDI-TOF MS inspires today.

1.4 MALDI-TOF MS as a Platform for DNA Sequencing

Although MALDI-TOF MS is widely used for proteomic and peptide mass fingerprinting, very few laboratories worldwide have applied this powerful technology for microbial genomics. A barrier to its direct analysis stems from the instability of DNA when subjected to MALDI-TOF analysis because, during desorption, base protonation causes rapid destabilization of the *N*-glycosidic bond, causing base loss and fragmentation at many positions along the DNA (Figure 1.10). This is in contrast to RNA, whose 2'-hydroxyl group enables greater stability. In practice, primers tagged with a T7 promoter enable transcription of DNA amplicons to the more 'stable' RNA molecules. These are cleaved at nucleotide bases, and their molecular weights readily identified by MS and compared with the simulated spectra of reference sequences. Not only is the method rapid and very accurate but base modifications such as methylation and acetylation are readily detected. Sequenom GmbH was the first company to exploit this phenomenon and initially focused on single-nucleotide polymorphisms for which viruses were common targets (see Chapter 8). As early as 2002, the authors (HNS and SEG) began to investigate the potential of this technology for microbial genotyping, and early data were reported by Dirk van den Boom at the fifth annual meeting (Table 1.1). Data obtained was used to support a successful joint scientific grant application in 2005 to expand the depth of the technology for microbial typing (Honisch *et al.*, 2007, 2010). Among the applications developed using the Sequenom MassARRAY® System was the transfer of the traditional serotyping system of Kauffmann and White for typing of *Salmonella*, which because of identifiable mutations in the genes encoding the 'O' and 'H' antigens, was readily transferable (Bishop *et al.*, 2010).

The long awaited first visit of Franz Hillenkamp, the pioneer and inventor of the technique he called matrix-assisted laser desorption/ionization TOF-MS, took place at the Gordon Museum, Guy's Campus, University of London, on 13 December 2007, through an invitation of the London Biological Mass Spectrometry Discussion

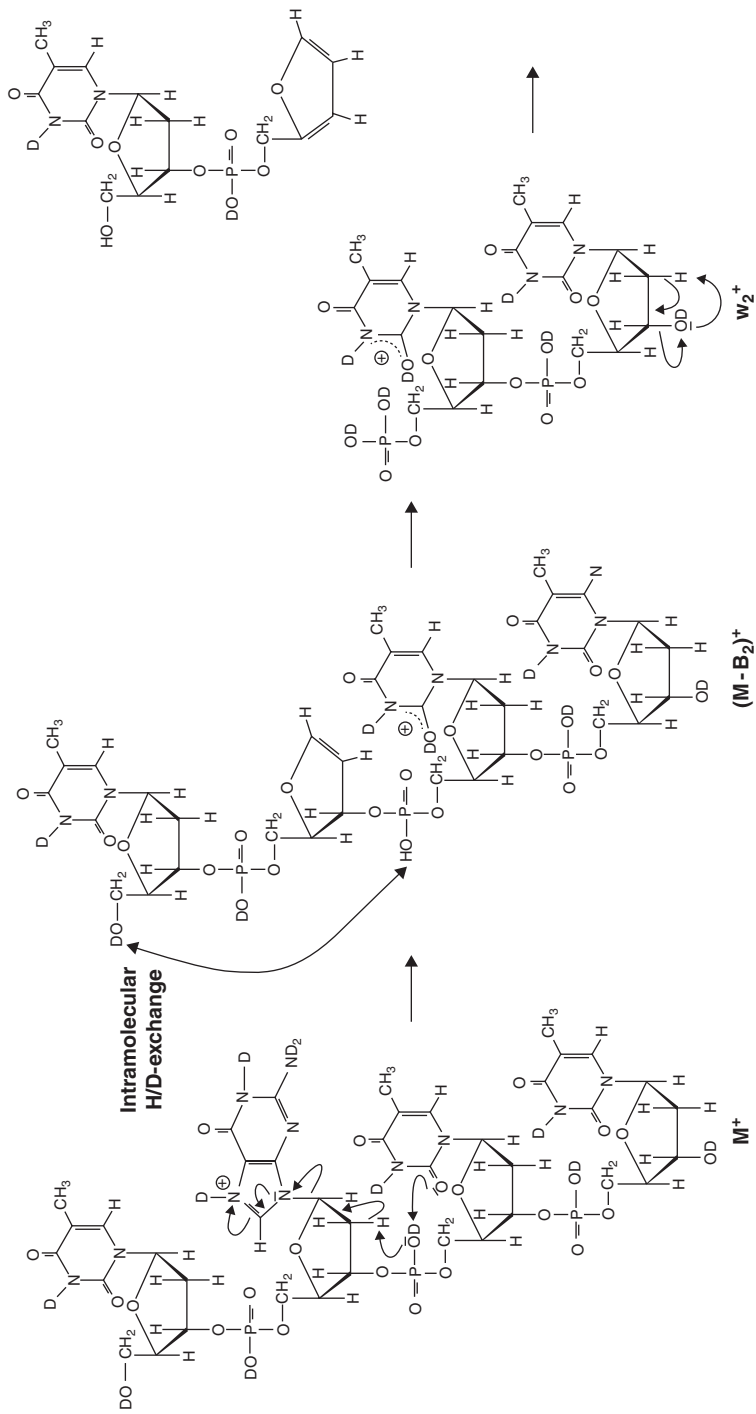


Figure 1.10 Proposed mechanism for the degradation of DNA in a MALDI-TOF mass spectrometer. This was taken from Franz Hillenkamp's notebook during his visit to MISU and AFGU in 2008. See text for details.

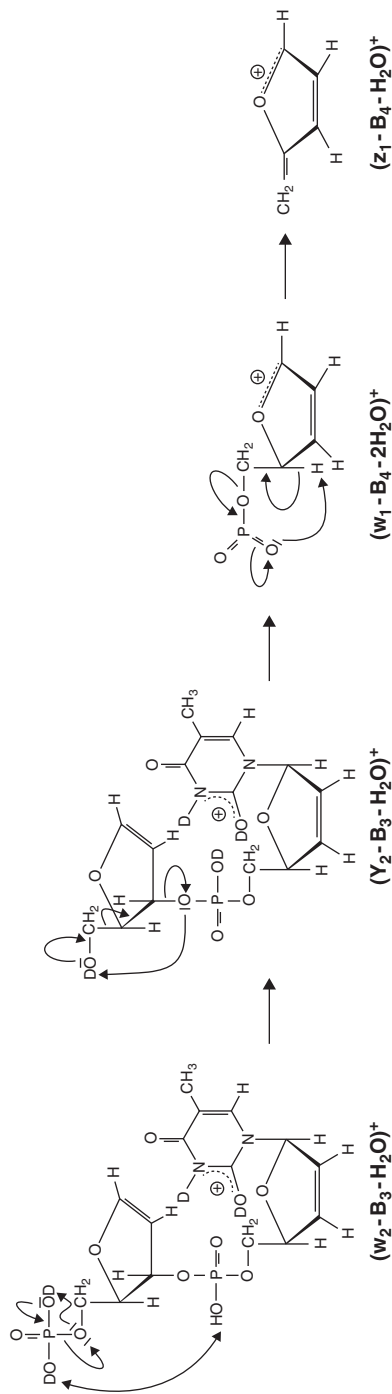


Figure 1.10 (Continued)

Group (LBMSDG). In a room packed to capacity with enthusiasts, Hillenkamp delivered a personal, riveting and moving lecture on the history of the technology; generously acknowledging his competitors and expressing his pride in seeing the applications of the technology today ‘in areas he had not foreseen.’ He generously complimented the work of our laboratory, and in the midst of his presentation invited author HNS to present some of the biological applications of the technology that were in use at PHE. He then resumed his presentation and speculated on future applications.

This visit led directly to his second trip to London the following year where he was invited to deliver the Plenary Lecture ‘MALDI-TOF MS: Development of an Analytical Tool for Biological Sciences’ at our 11th annual meeting (Table 1.1). He spent four days in the laboratory working with staff and PhD students and left a lasting impression on all who met him as a humble, gifted scientist who had a lifelong passion for science. The scientific community strongly believed he should have shared the Noble Prize in 2002 for Chemistry with Koichi Tanaka for the development of MALDI-TOF MS. He dismissed this as inconsequential and basked in the practical applications of his pioneering work. He took up an advisory post with Sequenom (which moved from Germany to San Diego, CA) and co-authored a chapter titled ‘DNA Resequencing by MALDI-TOF MS and Its Application to Traditional Microbiological Problems’ (Honisch *et al.*, 2010).

1.5 Insights into the Proteome of Major Pathogens 2005–2009: Field Testing of MALDI-TOF MS

The award of a major five-year scientific grant in 2005 to the authors for integrating genomics and proteomics for biomarker discovery (entitled ‘Detection of virulence and species biomarkers of deliberate release pathogens using an integrated genomic-proteomic high resolution platform’ (S.E. Gharbia and H. N. Shah for about £2 million) served as a springboard for the development of proteomics in PHLS (by then known as the Health Protection Agency [HPA]). Prior to this, our proteomics research was undertaken by PhD students. This major programme enabled the appointment of several dedicated staff and the purchase of a plethora of proteomic equipment including, a MALDI Reflectron, Sequenom’s MassARRAY® System, various gel-based equipment, imaging systems for DIGE, robotic stations for trypsin digestion, etc. However, it was the acquisition of the new and novel Thermo Electron Corporation’s nano-LC LTQ Orbitrap mass spectrometer system which propelled the organisation into advanced proteomics and had a dramatic impact in the field. At the eighth annual conference in 2005, the need for a London-based biological MS group took root and was eventually established by Anthony Sullivan as the LBMSDG, which up to now holds quarterly meetings in central London and attracts some of the pioneers in the field.

Alexander Makarov invented the Orbitrap, which subsequently led to a whole series of Thermo Fisher Scientific instruments that are used for both bottom-up and top-down proteomics. Michaela Scigelova, who worked closely with Makarov (Scigelova and Makarov, 2006), expounded the virtues of the Orbitrap in her presentation “Advances in the analysis of biomolecules using high resolution hybrid mass spectrometers” at our ninth annual conference in 2006. Although Makarov had presented

proof-of-principle results of the Orbitrap analyzer at the ASMS conference as early as 1999, commercial introduction of this analyzer by Thermo Fisher Scientific, as part of the hybrid LTQ Orbitrap instrument, did not materialize until 2005. With the timely success of this grant, we were therefore the recipient of one of the first instruments, and work began almost immediately. Proteome analysis of serovars Typhimurium and Pullorum of *Salmonella enterica* subspecies (Encheva *et al.*, 2005, 2007) and *Neisseria gonorrhoeae* were reported (Schmid *et al.*, 2005). A basic proteome reference map of *Streptococcus pneumoniae* was published in the journal *Proteomics* (Encheva *et al.*, 2006) while the Sequenom Mass Cleave technology was by then firmly established as a microbial typing tool in our laboratory (Honisch *et al.*, 2007). The bridging of genomics and proteomics was now established (see, e.g. Al-Shahib *et al.*, 2010; Misra *et al.*, 2012).

1.6 2010–2011: The Triumph of MALDI-TOF MS and Emerging Interest in Tandem MS for Clinical Microbiology

During the first decade of the millennium, ideas that were being explored at the turn of the century began to take root. There was a period of considerable development both in hardware, software and bioinformatics tools. The latter was already an integral part of genomics and was by then beginning to underpin developments in proteomics. Whole genome sequencing was becoming a reality for the clinical research laboratory, and with the vast amount of microbial sequence data accumulating, protein identification by MS/MS analysis, which is dependent on sequence data, accelerated significantly. The burgeoning field of proteogenomics was now inseparable and began expanding rapidly. The organization restructured its laboratories accordingly to facilitate this interaction. Thus, in 2010, Proteomics and Genomics Services and Research were amalgamated into one large specialty designated Department of Bioanalysis and Horizon Technologies, bringing genomics and proteomics laboratories in juxtaposition. Even though developments were gradual, a decade of work seemed to have reached a new milestone in 2010. New bioinformatics approaches were being used to characterize microbial biomarkers (see, e.g. Shah *et al.*, 2011). Bottom-up analysis of MS/MS data was permitting in-depth analysis of the microbial proteome, and new insights into their structure, metabolism and pathogenicity were being reported (see reviews in Shah and Gharbia, 2010). A notable example was the elucidation of the outer membrane proteome of *Salmonella enterica* serovar Typhimurium utilizing a lipid-based protein immobilization technique which was undertaken in collaboration with Roger Karlsson's laboratory in Sweden (Chooneea *et al.*, 2010). This work has now advanced as a means of proteotyping microbial species, which significantly expands the use of proteomics for microbial characterizations (see Chapter 16).

Bacterial pathogens considered category A bioterror agents such as *Bacillus anthracis*, *Francisella tularensis*, *Clostridium botulinum* and *Yersinia pestis* or category B agents such as *Burkholderia pseudomallei* and *Burkholderia mallei* require work in Biosafety level 3 cabinets (Rotz *et al.*, 2002). Because of the need to undertake all preparative work at such a high containment level, work on such pathogens was restricted to few laboratories and focused mainly on genomics. With the arrival of

nano-LC systems coupled to the LTQ Orbitrap, it was possible to work with microlitre volumes of cell extracts and undertake analysis of such pathogens. The number of strains and species could now be scaled up for MS/MS analysis. Consequently, we were able to develop a pipeline for biomarker discovery using MS and bioinformatics to include category A pathogens (Al-Shahib *et al.*, 2010). Among species where differentiation from closely allied taxa were contentious by genomic methods, unique strain-specific peptides were found that delineated each clearly (see Chapter 13) and was reported at the 20th meeting of the European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) in Vienna in 2010.

ECCMID, because of its strong focus on microbial diagnostics has, over the years, become a vehicle for reporting developments in the application of MALDI-TOF MS for microbial identification, typing and antimicrobial resistance detection. In previous years, lecture theatres would be sparsely occupied during presentations on MALDI-TOF MS. At ECCMID in Glasgow, 10–13 May 2003, the author's presentation on MALDI-TOF MS in microbial identification drew 15 people in a room with a capacity for 150. However, at ECCMID 2010, the largest lecture halls could not accommodate all who wished to attend, signalling a significant turning point in the acceptance of this technology. There was a dedicated symposium of invited speakers titled 'MALDI-TOF in Clinical Microbiology' (11 April 2010) in which our presentation titled 'MALDI-TOF MS of Surface-Associated and Stable Intracellular Proteins for Identification and Resistance Profiling of Human Pathogens' covered the work of the last decade. Many groups of species in which we previously encountered difficulties in obtaining confident identification, such as *Clostridium difficile*, *Bacillus* species or mycobacteria, were now being resolved and reported with improved confidence scores (e.g. Figures 1.11 and 1.12). The three other speakers of this session echoed this confidence to a large audience. A Poster Session on 12 April titled simply 'MALDI-TOF' reinforced this: there were 27 poster presentations on MALDI-TOF MS. The following day ended with an oral session titled 'What Can We Expect from MALDI-TOF?' in which there were 10 presentations. This immense endorsement of MALDI-TOF MS at one meeting led to a feeling of reflective optimism following a decade of persistent work and coincided with the publication of the book *Mass Spectrometry for Microbial Proteomics* (Shah and Gharbia, 2010). Displayed at the publisher's, Wiley's, desk at ECCMID 2010, copies were sold out to the new wave of transformed microbiologists.

Perhaps the biggest barrier to acceptance of this radical change to MALDI-TOF MS was for microbiologists to leave behind the deeply rooted role that morphological and biochemical tests played in bacterial identification since its inception in the late nineteenth century. The first chapter of the book, titled 'Changing Concepts in the Characterisation of Microbes and the Influence of Mass Spectrometry' (Shah *et al.*, 2010), charts the course of determinative microbiology and makes the case for this radical change to MALDI-TOF MS. Fundamental changes had already taken place in microbiology with the accessibility of DNA sequencing as described in the following chapter, 'Microbial Phylogeny and Evolution Based on Protein Sequences – The Change from Targeted Genes to Proteins' (Gupta, 2010), but clinical microbiologists still alluded to this as the domain of the research laboratory. Thus, nearly all the presentations at ECCMID 2010 used biochemical tests as a comparator to assess the performance of MALDI-TOF MS. It is interesting that bioMérieux, one of the global leaders of biochemical tests, revealed their acclamation of MALDI-TOF MS at ECCMID 2010 by the

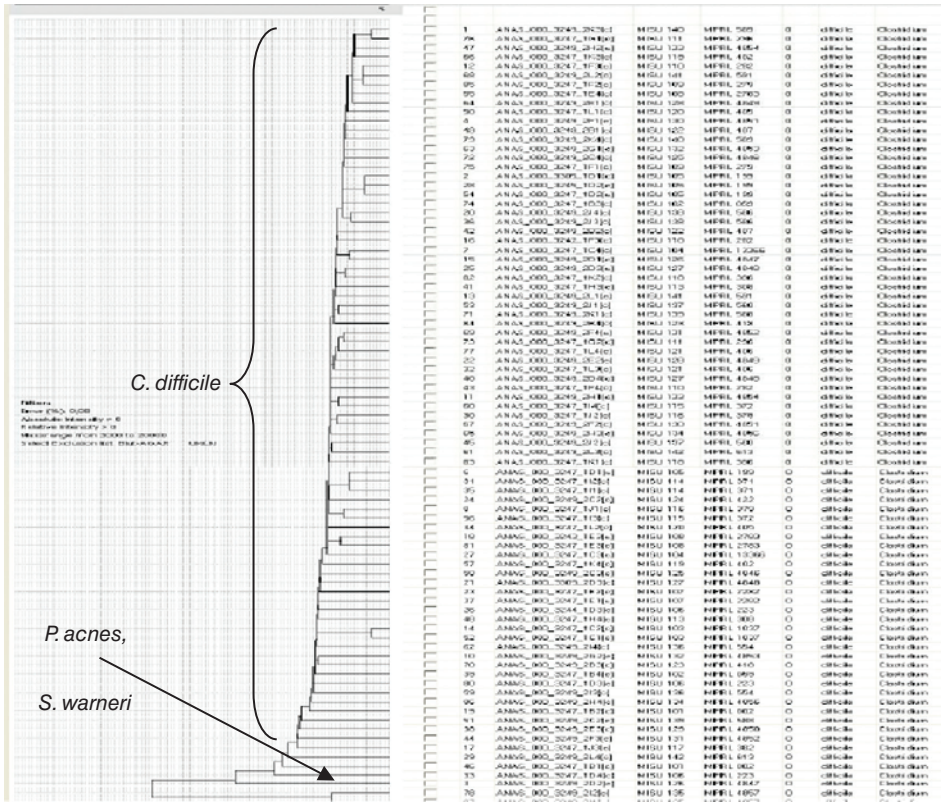


Figure 1.11 Identification of *Clostridium difficile*. Initially this posed a major problem for accurate and reproducible identification. The dendrogram shows some 100 clinical isolates of *C. difficile* clustering in a single phenon and distant from other species such as *Propionibacterium acnes* and *Staphylococcus warneri*, with which it initially formed a common cluster.

acquisition of AnagnosTec in a sweeping public display in which they substituted all AnagnosTec marketing literature with bioMérieux’s promotional material during the Trade Exhibition. bioMérieux states on their web site: ‘Right from the moment it launched, API® completely revolutionized the field of bacteriology. API® brings together high quality and ease of use with standardized, miniaturized strips of biochemical tests to use with comprehensive identification databases. With API®, bacterial and fungal identification is simple, rapid and reliable’ (<http://www.biomerieux-diagnostics.com/apir-id-strip-range>). To many, this change seemed extraordinary because although a few bioMérieux’s staff visited and worked at PHE for short intervals on MALDI-TOF MS, the company had no experience in MS, while Shimadzu, which manufactured the instruments, remained anonymous.

Thermo Fisher Scientific, which has an enormous global portfolio in MS, also had a strong presence at ECCMID 2010 but from their microbiological products perspective. They too were strongly influenced at ECCMID 2010, and a meeting was set up with the authors on the first evening to explore ways in which they might enter microbial MS. We already had five years experience with Thermo Fisher’s LTQ Orbitrap and presented

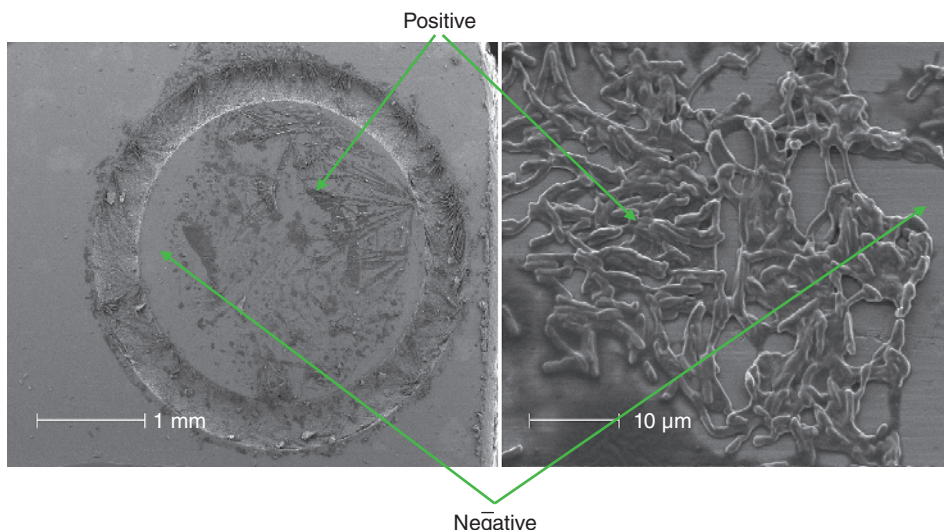
Scanning Electron Microscopy of cells of *C. difficile* mixed with Matrix Solution

Figure 1.12 Scanning electron micrographs of *Clostridium difficile* cells mixed with the matrix solution 2,5-dihydroxy benzoic acid in acetonitrile: ethanol: water (1:1:1) with 0.3% TFA. The electron micrographs show the clumps of cells which give a MALDI-TOF MS spectrum (indicated by positive) if the laser strikes, whereas in areas where there are no cells, no spectra are obtained (negative).

a significant amount of microbial proteomics at this meeting using this instrument. The outcome of this meeting was to either develop the bottom-up approach using nano-LC MS/MS further or explore the use of a new MALDI-TOF mass spectrometer.

The euphoric end to ECCMID 2010 was followed by several hundred e-mails and communications for more information immediately following the meeting. This optimism was channelled into our 13th annual conference in June 2010 (Table 1.1) in which both MALDI-TOF MS and LC-MS/MS work on human pathogens was reported. Prominent among these was a presentation titled 'MALDI-TOF MS Detection of Low Abundance and Low Molecular Weight Proteins Using Nanoparticles' which was based on the preliminary work we had embarked upon with the National Physical Laboratory; the hypothesis being to incorporate nanoparticles to capture low-abundance proteins that normally evade mass spectral analysis.

1.7 Preparations for MALDI-TOF MS Analysis on a Grand Scale: The Looming London 2012 Olympics

Following the summer of 2011 in the which a major outbreak of food-borne illness was caused by *E. coli* (see below), attention began to focus on the summer of 2012 in which the London Olympics and Paralympic Games were due to take place. Mass gatherings both in winter and summer pose a health threat, the most common being viral respiratory tract infections, including influenza, measles and multi-resistant TB, STDs and low-dose enteric infections that show person-to-person spread. Through the consumption

and sharing of food and drink, the list of bacterial and fungal pathogens can be greatly extended. The HPA (later titled PHE) had a central role in infectious disease surveillance for these events, setting up a suite of robust and multisource surveillance systems (see, e.g. Severi *et al.*, 2012).

At the laboratory level, cases were being made for simple, rapid, cost-effective, accurate, high-throughput methods. Up to this point, all MS work in PHE from 1998 onwards developed through successful research grants by the authors. For implementation to service function, it was necessary for the organization to purchase its own equipment, and between 2005 and 2010, request was made annually by the authors but declined. With the London Olympics looming, we made a fresh case again for a MALDI-TOF MS for its service function and was this time successful. An instrument was quickly installed, and we began reporting data on patients' samples by October 2011, three days after its arrival. The immediate success of this led to instruments being purchased for other PHE laboratories in Cambridge, Southampton, Birmingham and Bristol. With such a network developing, the authors established a PHE User Group in January 2012 which held meetings and symposia. With the University of Minho, Portugal, we immediately organized a two-day conference titled 'Microbial Diagnostic Applications of Mass Spectrometry' on 4–5 April 2012 at PHE, London. Just prior to this, ECCMID held its 22nd meeting between 31 March and 3 April 2012 in London, within the vicinity of the London 2012 main Olympics venue. Bruker Daltonik sponsored a session titled 'Microbial Identification for the 21st Century – and Beyond' which was chaired by Markus Kostrzewa. A presentation by Matthew Ellington (HPA, Cambridge, Addenbrookes, UK) titled 'HPA MS Implementation Group: Current Results and Future Plans' highlighted our hope and aspirations for MALDI-TOF MS within PHE. Soon after this meeting, PHE purchased a second instrument for our site in London to meet the demands of many laboratories that were now processing clinical samples daily using the microflex Biotyper. Over the next few months, this would expand to about 20 laboratories within PHE- and NHS-affiliated laboratories and now represents the largest global network of MALDI-TOF MS instruments in a single organization.

1.8 Investigating the Detection and Pathogenic Potential of *E. coli* O104:H4 during Outbreak of 2011

On 1 May 2011 an *E. coli* outbreak began in Germany with a trickle of patients presenting bloody diarrhoea (Askar *et al.*, 2011). By the end of the month, the number of reported cases surged to 1240, including cases reported in eight other European countries (see Shah and Gharbia, 2012). The whole outbreak resulted in more than 4000 reported cases and 50 deaths. A number of patients suffered from haemolytic-uraemic syndrome (HUS), a devastating and rare disease characterized by disintegration of red blood cells, acute kidney failure, and impaired ability to clot blood. The outbreak strain was positively identified on 25 May 2011. Workers at the University of Münster and the Robert Koch Institute identified the strain through serotyping and PCR assays. Multilocus sequencing was used to confirm that the outbreak was caused by a single clone, HUSEC041, and that it had the rare serotype of O104:H4. This serotype is normally associated with enteroaggregative *E. coli* (EAEC) that are known to cause

persistent diarrhoea, but not haemorrhaging or HUS (Frank *et al.*, 2011). The strain lacked features characteristic of O157:H7, such as an enterocyte effacement pathogenicity island and an intimin-positive gene, but produced aggregative factors typical of EAEC. However, the strain did exhibit high resistance to third-generation cephalosporins, trimethoprim/sulfamethoxazole and tetracycline, which is typical of O157:H7. The strain also possessed the rarer and more potent Shiga toxin 2 gene.

DNA sequencing would provide the initial blueprint for understanding the pathogen's novel set of characteristics. In June, two independent groups completed DNA sequencing of the outbreak isolate's 5.2 Mb genome and two large plasmids using short-read DNA sequencers. Both groups released the sequencing data to the scientific community, which rapidly performed bioinformatics to explain the strain's pathogenicity and evolutionary origin. It was also suggested that the strain may harbour genes unique from those in other strains.

By early June, the first cases were being reported in the United Kingdom, and samples were sent to our laboratories for whole genome sequence analysis. Cultures were grown overnight in 5 × 100 LB broth and nutrient agar plates in a Class 3 cabinet. All extracts were subcultured into enriched broths and retested for loss of viability before any further work was allowed. Author SEG had just acquired a Roche Junior and began 454 sequencing on 7 June 2011. By 9 June, the group began to assemble the scaffolds, which was successfully completed the following day and the full sequence deposited in real time in the National Center for Biotechnology Information (NCBI) library for crowd-sourcing bioinformatics analysis. This work, which was acknowledged in *Nature Biotechnology* (see editorial, 2011), provided the blueprint for a unique opportunity to explore whether it was possible to undertake proteome analysis of the cell extracts in real time using nano-LC-MS/MS.

Proteomic analysis was performed using an established workflow (Figure 1.13) on five *E. coli* strains of serotype O104: three clinical isolates from patients affected by the German outbreak and two other isolates that were previously characterized as serotype O104, but have EAEC, EHEC/STEC genetic composition, respectively. The genomes of the three German outbreak isolates were sequenced to confirm they were from the same strain. Strains were cultured on LB broth and agar as described above for DNA extracts and then harvested prior to employing two parallel approaches for reducing complexity of the mixture for MS analysis. In the first approach, lysates were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis, and 1 cm gel slices were digested with trypsin. Peptides were analyzed using nano-LC-MS/MS. In the second approach, the entire cell lysate was digested directly with trypsin in solution and injected onto two LC-MS/MS systems (LTQ Orbitrap, DBHT and LTQ Orbitrap Velos (Thermo Fisher, Hamel Hampstead), each with a front-end Ultimate 3000 Dionex nano/capillary liquid chromatography system, Thermo Fisher Scientific) that provided ultra-high-resolution and accurate masses for differentiating closely related peptides.

The recorded peptide MS/MS spectra were matched to both protein and in silico genome-translated databases to identify expressed proteins. The peptides were then fed into a bioinformatics pipeline (Al-Shahib *et al.*, 2010) to acquire unique signatures at the genus and species levels. An extensive list of identified peptides was then searched, using Blast and Scaffold, for virulence determinants, *E. coli* virulence factors and putative EHEC/STEC and EAEC-specific virulence markers. The peptide lists identified proteins

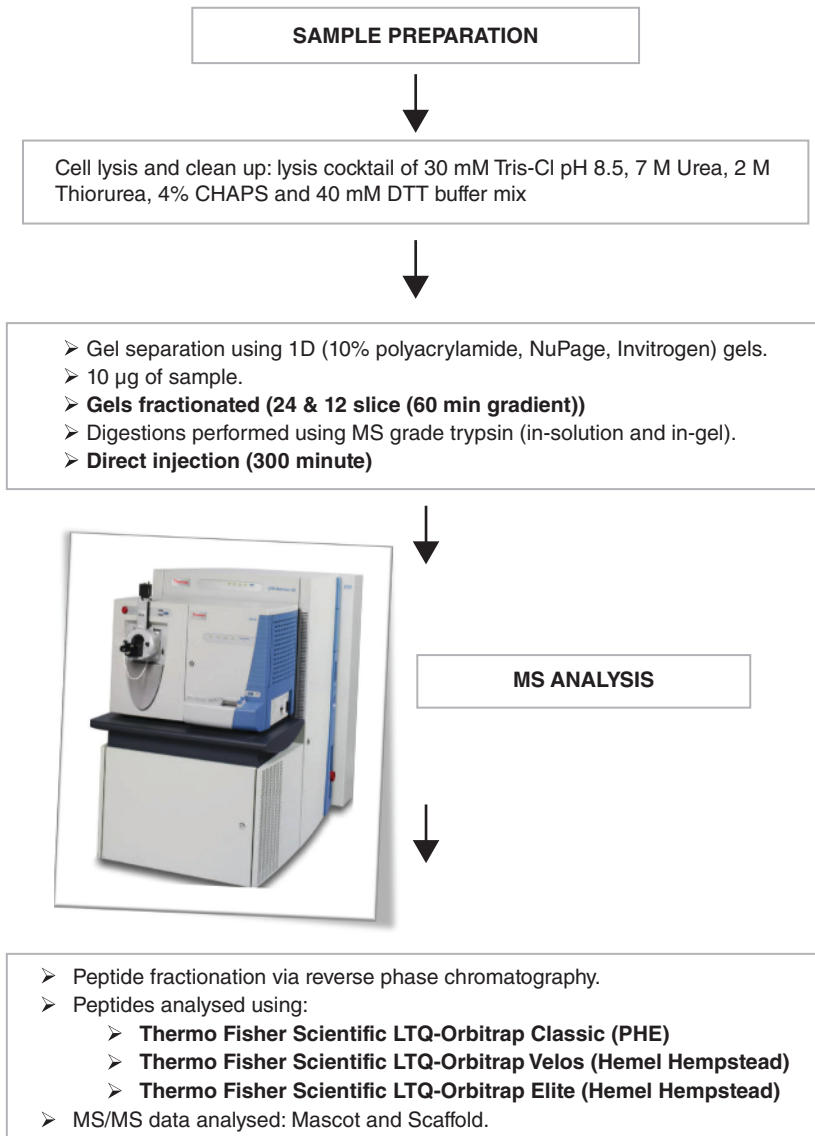


Figure 1.13 Bottom-up workflow used to deduce strain-specific peptides and virulence determinants of *E. coli* O104:H4 during the outbreak of 2011. The classical 1-D SDS-PAGE followed by in-gel trypsin digestion or directly loaded onto LC-MS/MS were also used. Three different Orbitrap mass spectrometers, with varying resolutions, were used to obtain comprehensive profiles of the proteome of strains.

that covered a significant percentage of the predicted open reading frames of the sequenced outbreak strain genome, indicating the sensitivity and reliability of the nano-LC-MS/MS method in yielding protein profiles using selective or enriched culture preparation. Peptides resulting from high-abundance proteins were then analyzed for markers and signatures that uniquely identified genus, species or virulence characteristics.

The mosaic outbreak strain's virulence signatures were compared to both EAEC and EHEC protein signatures, all obtained using the same proteome approach.

Data for isolate *E. coli* O104:H4 strain 280 was the first genome to be assembled into near complete topology. Approximately 2500 proteins from the outbreak isolates were identified. A collection of 68 peptide signatures were unique to the outbreak *E. coli* isolates and not shared by the parent source, separating the outbreak strain from other closely related Enterobacteriaceae. Species-level peptide signatures were also detected, including those for the AggR transcription factor, haemolysin protein, Aaf fimbriae protein and Iha adhesion protein. In all, 3031 peptides were identified as unique to the outbreak strains when compared against control isolates.

The technique detected features that were expected on the basis of prior laboratory tests and genomic data, including the production of Shiga toxin, Pic serine protease (autotransporter toxin) and tellurium resistance. The list of peptides was then filtered to exclude physiological and regulatory proteins. Search of the simplified list for *E. coli* pathotype virulence determinants and virulence factors resulted in a definitive list of expressed virulence determinants of the outbreak strain. The results supported the view that the background genome came from an EAEC progenitor that acquired plasmids and prophages, and exchanged chromosomal loci, leading to the emergence of an aggressive strain with a distinctive profile. All strains shared 89% of the expressed proteins. The two large plasmids encoded 31 proteins. Peptide signatures for adhesion and multidrug resistance (including β -lactamase, CTX-M extended spectrum β -lactamase and metallo- β -lactamase enzymes) were observed.

These experimental results demonstrated that a proteomic approach, based on nano-LC-MS/MS and comparison against a database of known pathogenic markers, accelerates the identification and characterization of the sources of *E. coli*-related illnesses and diseases. This study revealed for the first time that nano-LC-MS/MS was able to identify a significant number of pathogenic markers with no requirement for enrichment, selective media or antibiotic incorporation that can otherwise delay analysis. The protein signatures detected provide definitive characterization at the genus, species and often strain level, as well as detection of expressed pathogenic determinants and antibiotic resistance mechanisms. This mass-spectrometry-based approach enables clinical laboratories investigating outbreak strains to design screening and verification tests directly and in an unbiased manner, rather than performing multiple, potentially futile detection approaches while the outbreak is under way. The level of resolution achieved in this study could not be done using a linear MALDI-TOF MS and paves the way for MS/MS-based analysis for simultaneous microbial strain typing and pathotyping during outbreak investigations.

1.8.1 The Transition from MALDI-TOF MS to High-Resolution LC-MS/MS: Merits of Bottom-Up and Top-Down Proteomics for Microbial Characterization

By 2010, our laboratory had worked with all available forms of MALDI-TOF MS and now had five years experience with nano-LC-electrospray-MS/MS (see overview in Figure 1.14). Between 2011 and 2015, we investigated high-resolution MS to rapidly analyze minute volumes of complex cell extracts for strain identification and typing of expression markers. Supported by Thermo Fisher Scientific, we used three different models of the Q-Exact instruments during the *E. coli* O104:H4 outbreak of 2011 that

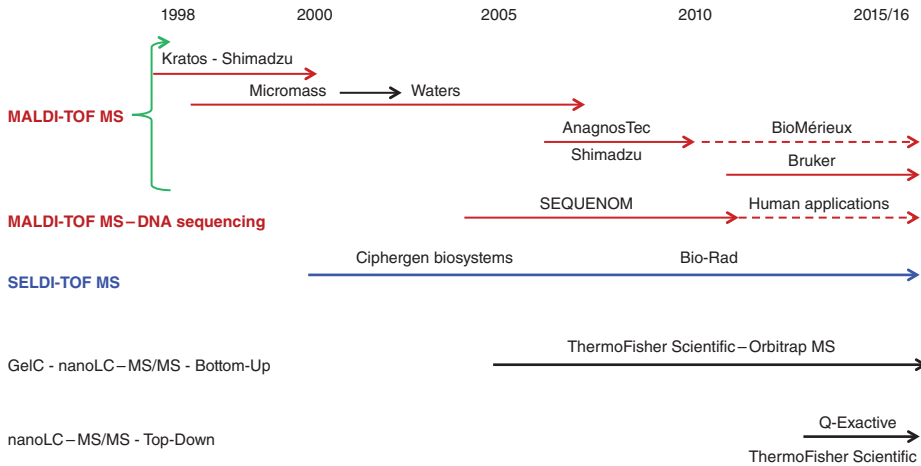


Figure 1.14 Phasing in of MS-based-methods in MISU-AFGU into clinical microbiology from 1998. These methods were used to characterize clinical isolates by our laboratories. The time lines and providers of the technology are shown in the figures. In 2010 bioMérieux acquired AnagnosTec, while Sequenom's interest shifted away from microbiology to solely human applications.

provided proof of principle that LC-MS/MS could be used to type and reveal the pathogenic potential of strains simultaneously (Shah and Gharbia, 2012).

Unlike MALDI-TOF MS, scrupulous attention needs to be given to bacterial protein extraction prior to MS/MS analysis. The challenge for the clinical microbiologists is to devise a universal protocol that encompasses the most fragile gram-negative cells to the most robust structures such as spores and the complex cell envelopes of mycobacteria. Furthermore, although effective reagents are needed, lysis methods need to avoid chemicals/detergents that cause interference with chromatography or MS. This was investigated, and the lysis mixtures that best met these criteria were designated as follows:

DIGE	7 M Urea, 2 M thiourea, 4% CHAPS, 30 mM Tris and 70 mM DTT
CHAPS	4% CHAPS, 30 mM Tris and 70 mM DTT
ACN/TFA	Acetonitrile : Water : Trifluoroacetic acid (33%:66.9%:0.1%)
ACN/FA	Acetonitrile : Water : Formic acid (50%:45%:5%)

ACN/TFA and ACN/FA were introduced to mitigate the potential problems associated with urea and were being used for top-down proteomics. As an example, two clinically relevant species representing opposite ends of the spectrum, *Clostridium difficile* and *Escherichia coli*, were tested. *E. coli* is a gram-negative, non-spore-forming, facultative anaerobe, whereas *C. difficile* is a gram-positive spore-forming anaerobe; they represent many of the physiological and phenotypic features that are likely to be encountered in the clinical laboratory and would provide a challenging panel for mass-spectrometry-based workflows.

A standard shotgun protocol for *C. difficile* and *E. coli* was, briefly, to harvest cells grown for 24 h on Columbia blood agar plates and suspend the cells in 300 μ l of one of

the lysis solutions. Samples in triplicate were then lysed in one of two ways: (1) either incubating the bacteria in the lysis solution for 30 min at room temperature before centrifugation and protein extraction or (2) with mechanical lysis using the FastPrep system (MP Biomedicals). The FastPrep samples contained 100–150 μ l of glass beads added to the sample and were beaten at 4 m/s for three rounds of 20 s. With lysis complete, the samples were centrifuged at $21,000 \times g$ for 30 min at 4 °C, and the supernatant stored at –20 °C until LC-MS/MS analysis. If lysis is efficient, samples were separated on a 12% SDS-PAGE gel for a brief period (\approx 15 min) to exclude impurities and salts. Gel slices of 1 cm were excised, destained and digested overnight with trypsin. Chromatographic separation of the complex mixture of peptides derived from tryptic digest was achieved on C_{18} column over a 60 min gradient and immediately analyzed with the LTQ Orbitrap mass spectrometer. Protein identification was performed by searching databases with search engines such as Mascot, Sequest or Phenyx (IS 2011). The resultant datasets were processed using a pipeline that we established as an in-house marker discovery workflow (Al-Shahib *et al.*, 2010). For *C. difficile* there were 2383 different peptide markers specific for *C. difficile* and conserved between the three biological replicates. For *E. coli* there were 104 different peptide specific markers conserved between the three replicates. The disparity between the number of markers is due in part to the difficulties in species resolution. *E. coli* is closely related to species of the genus *Shigella*; hence, many peptides are shared, whereas *C. difficile* is phylogenetically more distinct and therefore has a greater number of unique peptide biomarkers.

Despite the marked differences between these species, the results demonstrate unequivocally that among bacterial species several hundred unique peptides are present that may be used to characterize bacterial isolates at the species and strain levels. Furthermore, markers of antibiotic resistance are evident that correlate with known antimicrobial profiles of strains (unpublished work; see abstract ECCMID, 2010). A drawback of the method is that the workflow is still too complex and lengthy for routine processing of clinical samples. However, methodologies are rapidly changing, and trypsin digestion, for example, which hitherto required overnight incubation is now done in minutes in most laboratories (Hustoft *et al.*, 2012). With processing rapidly decreasing, bottom-up proteomics will soon be achievable in a few hours. It is expected that the entire workflow will be automated, which will bring this approach into the realms of the clinical laboratory.

An alternative approach would be to utilize top-down proteomics, which is infinitely faster, capable of quantitating unique proteoforms including post-translational modifications (PTM) and avoids the need for enzymic digestion of samples (Roth *et al.* 2008). In the top-down protocol, proteoforms are delivered to the mass spectrometer intact and then sequenced by fragmentation inside the instrument, thereby retaining their critical linkage information (see overview of the method compared to the earlier bottom-up approach in Figure 1.15). This is technically more demanding in that the intact proteins are more difficult to fractionate and fragment than peptides, and more challenging to separate by liquid chromatography. Given the need to distinguish proteins varying by only small chemical differences, high-end, high-resolution instruments are required to resolve such large molecules when they are so similar in size. Dedicated software (e.g. ProSightPC 2.0) is used for analysis. A range of instruments are now amenable to top-down analysis. Prior to 2010, Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometers were the main instruments used for top-down analysis.

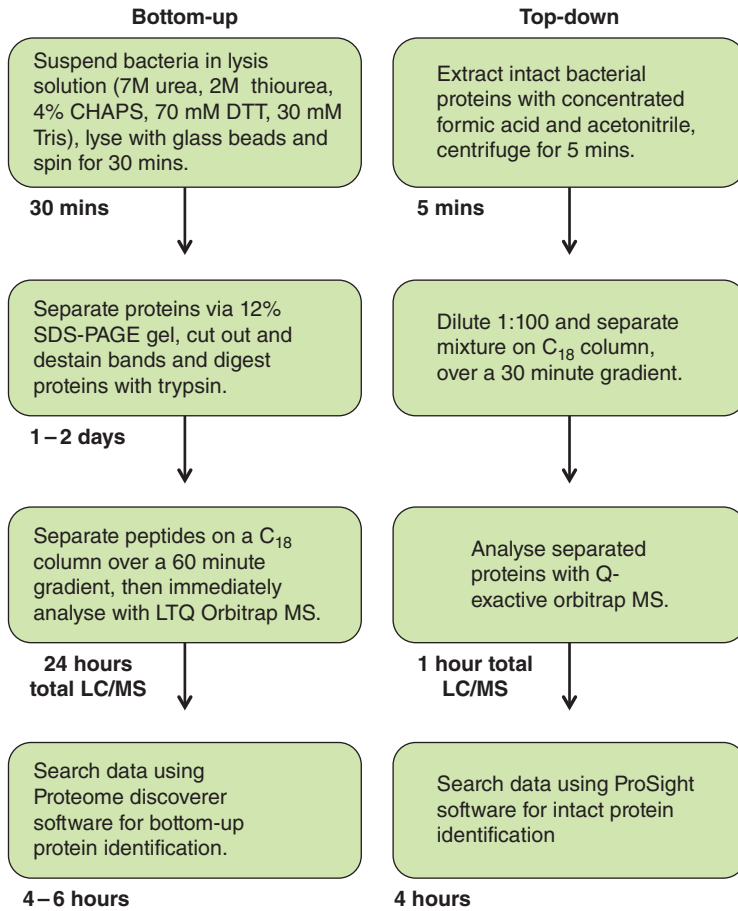


Figure 1.15 A comparative overview of bottom-up versus top-down proteomics. Although the former has been used extensively for comprehensive analysis of the proteome of several species, it is currently too cumbersome for a clinical laboratory. It seems likely that top-down based approaches will supersede MALDI-TOF MS as the next-generation approach to microbial identification and typing.

Today, more affordable quadrupole-time-of-flight (qTOF) instruments, such as the Waters SYNAPT G2-Si, the Bruker maXis II or the new range of Orbitrap instruments (Thermo Fisher Scientific) have made the technology more accessible (Shah *et al.*, 2015).

Unlike the human proteome bacterial extracts are not subjected to the same level of complexity, and hence pre-separation methods such as those employed by Ahif *et al.* (2013) are superfluous. However, the multiplicity of the bacterial proteome is known to be significantly enhanced through PTMs (see review, Mijakovic *et al.*, 2014). These include methylation, phosphorylation, acetylation, glycosylation, pupylation, sirtuin acetylation, lipidation, carboxylation, bacillithiolation, etc. This rapidly expanding catalogue, made possible by advanced detection methods in MS, will provide a plethora of new proteoforms. Because many of these have regulatory functions, the importance of their detection in microbial pathogenicity cannot be overstated. A significant advantage of top-down analysis is its ability to detect these PTMs, sequence variants

and degradation products, markedly increasing the potential number of species/strain protein signatures.

In terms of elucidating molecular mechanisms of microbial pathogenicity, the ability to now deduce protein variations that arise from alternative splicing, allelic variation, or PTMs opens a new chapter in this field. One elegant example by Julia Chamot-Rooke's group used a combination of bottom-up and top-down MS to characterize a PTM on the major pilin (PilE) of *Neisseria meningitidis* Type IV pili. It was then shown that this modification (glycerophosphorylation), which is induced in vivo after several hours of host cell contact, is a prerequisite for the dissemination of the bacterium, a crucial step that precedes invasive infection (Chamot-Rooke *et al.*, 2011; see Chapter 18). Such applications are the real beneficiaries of these new technologies and will have a profound impact on the next phase of elucidating the diverse range of pathogenic mechanisms exhibited by microorganisms.

1.9 Conclusions

In the vast landscape of proteomics, the greatest success achieved to date in the clinical laboratory is the rapid, accurate, low-cost and simple method of identification of microbial pathogens using a linear MALDI-TOF mass spectrometer. A statement of our confidence in the method was evident since 2000 through various publications (Shah *et al.*, 2010). Having analyzed tens of thousands of clinical samples in parallel with 16S rRNA sequencing, chemotaxonomic and biochemical tests, we believe that MALDI-TOF MS has significantly surpassed all as the method of choice today, and the implementation of a network of over 20 instruments in the Health Services laboratories at present is a testament to this confidence.

There is currently no comparable achievement in clinical biomarker application for complex human and animal diseases, but many believe that top-down proteomics will achieve equivalent success in the near future (see review; Gregorich and Ying, 2014). Because the proteome of microorganisms are orders of magnitude smaller than eukaryotes, success in clinical microbiology may be more forthcoming using top-down MS. Although top-down proteomic analysis can be operated at relatively high speed in order to successfully map the bacterial proteome, the rate of identification of new proteins after initial rounds will need to be significantly enhanced for application in the clinical laboratory. Direct infusion is possible but may not be practical for automation; instead, an integrated single or dual LC with tandem MS that is completely 'hands-free' is likely to be the way forward.

There is little doubt that highly automated sophisticated instruments will soon be available. For MS/MS to reach a similar level of success as MALDI-TOF MS, the quality and reproducibility of cell extracts from the same strain grown in different media, various time intervals and analyzed by different models of the same instrument are critical parameters that need validation in the transition to top-down proteomics from the research to the clinical laboratory. Unlike MALDI-TOF MS, protein concentrations need to be predetermined and standardized and up to the present time, available methods for quantifying protein concentrations do not give similar results. For example, newer, simple methods using NanoDrop, which performs efficiently for DNA, does not give comparable results with the standard 40-year-old Bradford assay (1976). A simple, rapid method needs to be incorporated as a baseline for strain comparison

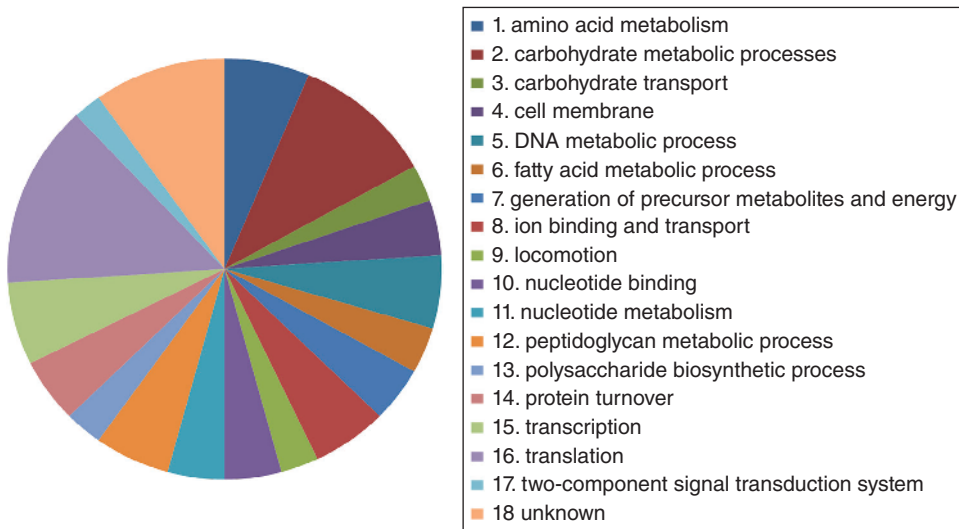


Figure 1.16 Distribution of cellular proteins of a cell extract to assess the efficacy of an extraction method. The figure shows the protein distribution of a standard *E. coli* strain using mechanical breakage and centrifugation steps. Methods such as sonication released mainly ribosomal proteins and yields profiles similar to MALDI-TOF MS but more complex. A method that captures a broad cross-section of protein is, in our view, far more beneficial for a future database.

because the key to success is the quality of the sample and the standardization of cell preparations.

The success of MALDI-TOF MS has been underpinned by the ability to retain the basic platform on which to assemble the database. Thus, although current commercial systems are capable of higher resolution and protein identification in reflectron mode, microbial identification is performed in the linear mode. In our experience, analysis of the proteome of cells cultured on agar plates is more reproducible than broth cultures but in assessing the pathogenic potential of a strain such as *C. difficile* for toxin expression, it may be necessary to culture in both broth and agar plates. The distribution of cellular proteins is markedly influenced by the type of extraction method. Some methods involving harsh detergents can yield very high levels of the more stable ribosomal proteins (>70%), which can mask the cellular proteins (unpublished). We believe that comprehensive representation of the proteome (e.g. in the form of a simple pie chart; Figure 1.16) enables visualization of the sample and is necessary prior to assembly of an MS/MS database. Current work in several laboratories is now enabling this baseline to be established, and it is expected in the near future that MS/MS-based methods will become a reality for strain typing, antibiotic sensitivity profiling and pathotyping in real time.

References

- Ahif, D. R., Thomas, P. M. and Kelleher, N. L. (2013). Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues. *Curr. Opin. Chem. Biol.* 17: 787–794.

- Al-Shahib, A., Misra, R., Ahmod, N., Fang, M., Shah, H. N. and Gharbia, S. E. (2010). Coherent pipeline for biomarker discovery using mass spectrometry and bioinformatics. *BMC Bioinformatics*. 11: 437.
- Askar, M., Faber, M. S., Frank, C., Bernard, H., Gilsdorf, A., Fruth, A., Prager, R., Höhle, M., Suess, T., Wadl, M., Krause, G., Stark, K. and Werber, D. (2011) Update on the ongoing outbreak of haemolytic uraemic syndrome due to Shiga toxin-producing *Escherichia coli* (STEC) serotype O104, Germany, May 2011. *Euro. Surveill.* 16: 19883.
- Bishop, C., Arnold, C. and Gharbia, S. E. (2010). Transfer of a traditional serotyping system (Kauffmann-White) onto a MALDI-TOF MS Platform for the rapid typing of *Salmonella* isolates. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 463–496). Wiley, Chichester.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248–254.
- Cain, T. C., Lubman, D. M. and Weber, W. J. Jr. (1994). Differentiation of bacteria using protein profiles from matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectro.* 8: 1026–1030.
- Chamot-Rooke, J., Mikaty, G., Malosse, C., Soyer, M., Dumont, A., Gault, J., Imhaus, A. F., Martin, P., Trellet, M., Clary, G., Chafey, P., Camoin, L., Nilges, M., Nassif, X. and Duménil, G. (2011) Posttranslational modification of pili upon cell contact triggers *N. meningitidis* dissemination. *Science*. 331: 778–782.
- Choonea, D., Karlsson, R., Encheva, V., Appleton, H. and Shah, H. N. (2010). Elucidation of the outer membrane proteome of *Salmonella enterica* serovar Typhimurium utilising a lipid-based protein immobilization technique. *BMC Microbiol.* 10: 44.
- Claydon, M. A., Davey, S. N., Edwards-Jones, V. and Gordon, D. B. (1996). The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotech.* 14: 1584–1586.
- Dekio, I., Culak, R., Misra, R., Gaulton, T., Fang, M., Sakamoto, M., Oshima, K., Hattori, M., Klenk, H.-P., Rajendram, D., Gharbia, S. E. and Shah, H. N. (2015). Dissecting the taxonomic heterogeneity within *Propionibacterium acnes*: Proposal for *Propionibacterium acnes* subsp. *acnes* subsp. nov. and *Propionibacterium acnes* subsp. *elongatum* subsp. nov. *Int. J. Syst. Evol. Microbiol.* 65: 4776–4787.
- EDITORIAL (2011). Whole-genome sequencing and crowd sourced analyses proved a powerful adjunct to traditional typing in the recent *Escherichia coli* outbreak. *Nature Biotech.* 29: 769.
- Encheva, V., Wait, R., Gharbia, S. E., Begum, S. and Shah, H. N. (2005). Proteome analysis of serovars Typhimurium and Pullorum of *Salmonella enterica* subspecies. *BMC Microbiol.* 5: 42
- Encheva, V., Wait, R., Gharbia, S. E., Begum, S. and Shah, H. N. (2006). Comparison of extraction procedures for proteome analysis of *Streptococcus pneumoniae* and a basic reference map. *Proteomics*. 6: 3306–3317.
- Encheva, V., Wait, R., Begum, S., Gharbia, S. E. and Shah, H. N. (2007). Protein expression diversity amongst serovars of *Salmonella enterica*. *Microbiology*. 153: 4183–4193.
- Evason, D. J., Claydon, M. A. and Gordon, D. B. (2001). Exploring the limits of bacterial identification by intact cell-mass spectrometry. *J. Amer. Soc. Mass Spectro.* 12: 49–54.
- Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., an der Heiden, M., Bernard, H., Fruth, A., Prager, R., Spode, A., Wadl, M., Zoufaly, A., Jordan, S., Kemper, M. J., Follin, P.,

- Müller, L., King, L. A., Rosner, B., Buchholz, U., Stark, K. and Krause, G. (2011). Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N. Engl. J. Med.* 365: 1771–1780.
- Gregorich, Z. R. and Ying, G. (2014). Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics.* 14: 1195–1210.
- Gupta, R. S. (2010). Microbial phylogeny and evolution based on protein sequences – the change from targeted genes to proteins. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 35–53). Wiley, Chichester.
- Gusev, A. I., Wilkinson, W. R., Proctor, A. and Hercules, D. M. (1995). Improvement of signal reproducibility and matrix/co-matrix effects in MALDI analysis. *Anal. Chem.* 67: 1034–1041.
- Holland, R. D., Wilkes, J. G., Rafii, F., Lay, J. O. (1996). Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Comm. Mass Spectro.* 10: 1227–1232.
- Honisch, C., Chen, Y. and Hillenkamp, F. (2010). DNA resequencing by MALDI-TOF MS and its application to traditional microbiological problems. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 443–462). Wiley, Chichester.
- Honisch, C., Chen, Y., Mortimer, C., Arnold, C., Schmidt, O., van den Boom, D., Cantor, C. R., Shah, H. N. and Gharbia, S. E. (2007). Automated comparative sequence analysis by base-specific cleavage and mass spectrometry for nucleic acid-based microbial typing. *Proc. Natl. Acad. Sci. U S A.* 104: 10649–10654.
- Hustoft, H. K., Malerod, H., Wilson, S. R., Reubsæet, L., Lundanes, E. and Greibrokk, T. (2012). A critical review of trypsin digestion for LC-MS based proteomics. In Dr. Hon-Chiu Leung (Ed.), *Integrative Proteomics*. ISBN: 978-953-51-0070-6, InTech.
- Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M. and Shah, H. N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect. Genet. Evol.* 4: 221–242.
- Krishnamurthy, T. and Ross, P.L. (1996). Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectro.* 10: 883–888.
- Mijakovic, I., Grangeasse, C. and Stülke, J. (Eds.). In: Regulatory potential of post-translational modifications in bacteria. Frontiers Media SA. ISBN: 978-2-88919-610-4. Product Name: *Frontiers Research Topic Ebook*. From *1st International Conference on Post-Translational Modifications in Bacteria* (September 9–10, 2014, Göttingen, Germany).
- Misra, R. V., Ahmod, M. Z., Parker, P., Fang, M., Shah, H. N. and Gharbia, S. E. (2012). Developing an integrated proteo-genomic approach for the characterisation of biomarkers for the identification of *Bacillus anthracis*. *J. Microbiol. Meth.* 88: 237–247.
- Mohanty, G. and Mukherji, S. (2008). Biodegradation rate of diesel range *n*-alkanes by bacterial cultures *Exiguobacterium aurantiacum* and *Burkholderia cepacia*. *Intern. Biodeter. Biodegrad.* 61: 240–250.
- Pitt, T. L., Malnick, H., Shah, J., Chattaway, M. A., Keys, C. J., Cooke, F. J. and Shah, H. N. (2007). Characterisation of *Exiguobacterium aurantiacum* isolates from blood cultures of six patients. *Clin. Microbiol. Infect.* 13: 946–948.

- Rajakaruna, L., Hallas, G., Dare, D., Sutton, H., Encheva, V., Culak, R., Innes, I., Ball, G., Sefton, A. M., Eydmann, M., Kearns, A. M. and Shah, H. N. (2009). High throughput identification of clinical isolates of *Staphylococcus aureus* using MALDI-TOF-MS of intact cells. *Infect. Genet. Evol.* 9: 507–513.
- Rajendram, D. (2003). PhD Thesis: 'Dissecting the Diversity of the Genus *Peptostreptococcus* using Genomics and Proteomic Analyses' (University of East London). Supervisors: H. N. Shah and S. E. Gharbia.
- Reynaud af Geijersstam, A., Culak, R., Molenaar, L., Chattaway, M., Røslie, E., Peciuliene, V., Haapasalo, M. and Shah, H. N. (2007). Comparative analysis of virulence determinants and mass spectral profiles of Finnish and Lithuanian endodontic *Enterococcus faecalis* isolates. *Oral Microbiol. Immunol.* 22: 87–94.
- Roth, M. J., Parks, B. A., Ferguson, J. T., Boyne, M. T. and Kelleher, N. L. (2008). Proteotyping: Population proteomics of human leukocytes using top down mass spectrometry. *Anal Chem.* 80: 2857–2866.
- Rotz, L. D., Khan, A. S., Lillibrige, S. R., Ostroff, S. M. and Hughes, J. M. (2002). Public health assessment of potential biological terrorism agents. *Emerg. Infect. Dis.* 2: 225–230.
- Schmid, O., Ball, G., Lancashire, L., Culak, R. and Shah, H. N. (2005). New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionisation time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. *J. Med. Microbiol.* 54: 1205–1211.
- Scigelova, M. and Makarov, A. (2006) Orbitrap mass analyzer – overview and applications in proteomics. *Proteomics.* 6 Suppl 2 (S2): 16–21.
- Severi, E., Heinsbroek, E., Watson, C. and Catchpole, M. (2012). HPA Olympics Surveillance Work Group. Infectious disease surveillance for the London 2012 Olympic and Paralympic Games. *Euro Surveill.* Aug 2; 17(31): 20232.
- Shah, H. N. (2005). MALDI-TOF-Mass Spectrometry: Hypothesis to proof of concept for diagnostic microbiology. *Clin. Lab. Intern.* 29: 35–38.
- Shah, H. N., Chilton, C., Rajakaruna, L., Gaulton, T., Hallas, G., Atanassov, H., Khoder, G., Rakowska, P. D., Eleonora Cerasoli, E. and Gharbia, S. E. (2010). Changing concepts in the characterisation of microbes and the influence of mass spectrometry. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 3–34).. Wiley, Chichester.
- Shah, H. N. and Collins, M. D. (1980). Fatty acid and isoprenoid quinone composition in the classification of *Bacteroides melaninogenicus* and related taxa. *J. Appl. Bacteriol.* 48: 75–84.
- Shah, H. N. and Collins, M. D. (1983). Genus *Bacteroides* a chemotaxonomical perspective. *J. Appl. Bacteriol.* 55: 403–416.
- Shah, H. N., Encheva, V., Schmid, O., Nasir, P. et al. (2005). Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS): A potentially powerful tool for rapid characterisation of microorganisms. In M. J. Miller (Ed.), *Encyclopedia of Rapid Microbiological Methods* (Vol. 3, pp. 57–96). DHI Publishing, IL, USA.
- Shah, H. N. and Gharbia, S. E. (2010). *Mass Spectrometry for Microbial Proteomics*, H. N. Shah and S. E. Gharbia (Eds.). Wiley, Chichester.
- Shah, H. N. and Gharbia, S. E. (2011). A century of systematics of the genus *Bacteroides*: From a single genus up to the 1980s to an explosion of assemblages and the dawn of MALDI-TOF-Mass Spectrometry. *The Bulletin of BISMis.* 2: 87–106.

- Shah, H. N. and Gharbia, S. E. (2012). Using nano-LC-MS/MS to investigate the toxicity of outbreak *E. coli* 0104:h4 strain. *Culture*. 33, ISSN 0965-0989.
- Shah, H. N., Keys, C., Gharbia, S. E., Ralphson, K., Trundle, F., Brookhouse, I. and Claydon, M. (2000). The application of MALDI-TOF Mass Spectrometry to profile the surface of intact bacterial cells. *Micro. Ecol. Health. Dis.* 12: 241–246.
- Shah, H. N., Keys, C. J., Schmid, O. and Gharbia, S. E. (2002). Matrix-assisted laser desorption/ionisation time of flight mass spectrometry and proteomics; a new era in anaerobic microbiology. *Clin. Infect. Dis.* 35: 58–64.
- Shah, H. N., Rajakaruna, L., Ball, G., Misra, R., Al-Shahib, A., Fang, M. and Gharbia, S.E. (2011). Tracing the transition of methicillin resistance in subpopulations of *Staphylococcus aureus* using SELDI-TOF Mass Spectrometry and Artificial Neural Network Analysis. *System. Appl. Microbiol.* 34: 81–86.
- Shah, H. N., van Steenberg, T. J. M., Hardie, J. M. and de Graaff, J. (1982). DNA base composition, DNA-DNA reassociation and isoelectric focusing of proteins of strains designated *Bacteroides oralis*. *FEMS Microbiol. Letts.* 13: 125–130.
- Shah, H. N., Stephenson, J., Cardasis, H., Neil, J., Yip, P., Ravela, S., Ritamo, I., Damsbo, M., Grist, R., Freeke, J., Stielow, B., Gvozdyak, O., Dukik, K., de Hoog, S., Damoc, E., Valmu, L., Cherkassky, A., Fang, M., Gaulton, T., Misra, R. and Gharbia, S. E. (2015). A Global Diagnostic Approach for Microbial Identification: Accurate characterization of difficult to differentiate pathogens – using Top-down proteomics. *Clinical Microbiology and Infectious Diseases (ECCMID) – 25th European Congress*. Copenhagen, Denmark-April 25–28, 2015.