This minireview discusses the use of mass spectrometry in biomarker discovery, the current utility of these markers for bacterial identification after culture, and the potential for non-culture-based diagnosis of infectious diseases. The bases of these thoughts are the independent revolutions that have occurred in the fields of molecular biology and analytical chemistry, leading to the current interrelatedness of genomics, proteomics, and bioinformatics.

The particular focus is on protein markers and proteomics, which are today essentially synonymous with biomedical mass spectrometry. The methods and approaches, while discussed here in the context of bacterial identification, are equally applicable to viruses, fungi, and parasites. To gain a better understanding of the current state of the art, it is important to understand what mass spectrometry has achieved, what are its current capabilities, and what might be expected in the not-too-distant future.

The molecular biology revolution included the development of the PCR and use of restriction enzymes for recognition of sequence differences among organisms by employing “known” genetic markers. The process of marker discovery has been greatly aided in recent years by whole-genome sequencing. In turn, there has been a revolution in mass spectrometry, leading to sequencing of the expressed protein products of these genomes (proteomics). Matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) and electrospray ionization (ESI) mass spectrometry (MS) and tandem mass spectrometry (MS-MS) have been at the core of these developments.

Microbiologists, whose research does not focus on mass spectrometers, are often thinking of an older technology (gas chromatography-mass spectrometry [GC-MS]). In the clinical microbiology field GC is now routinely used for whole-cell fatty acid profiling after prior growth in culture medium (10, 19); included in reference 10 is a nice chapter by Wayne Moss (one of the pioneers of this technique, now known as GC analysis of FAMEs [fatty acid methyl esters]). GC-MS provides additional structure information on these profiles. Therefore, a short section on small-molecule profiling using GC-MS has been included here to provide some perspective (“past developments”). The genomic revolution has given us a vast array of molecular biology tools for discrimination of well-known pathogens as well as emerging infections by the presence or absence of genes or, for closely related organisms, small changes in DNA sequence. It is anticipated that protein sequence-based discrimination will be as important for the next generation of clinical microbiologists. Thus, a section on large-molecule analysis for bacterial identification after culture is presented (8, 20) (“the present”). Finally, the potential for rapid diagnosis of diseases without culture is briefly discussed (“the future”).

SMALL MOLECULES AND CLINICAL MICROBIOLOGY: BACTERIAL IDENTIFICATION AFTER CULTURE (PAST DEVELOPMENTS)

As noted above, MS technology falls into two distinct areas and from the perspective of clinical microbiology relates primarily to the type of molecule that can be analyzed. In the older GC-MS technology (introduced in the 1970s and 1980s) polymers or oligomers (e.g., phospholipids) are first broken into their constituent monomers (e.g., fatty acids) by saponification or hydrolysis. The monomers are then chemically converted (e.g., through derivatization by methylation), which inhibits ionic or hydrogen-bonding interactions, allowing volatility in a gas chromatograph. The component monomers (e.g., a mixture of methylated fatty acids) are then separated into components by passage through the GC prior to passage into a detector. The most common GC detector universally detects all fatty acids, and the only information provided is that a fatty acid has been detected; identification is based on how long the fatty took to reach the detector as it passed through the GC column (retention time). Additionally, the exact structure of each fatty acid in the profile can be determined using a mass spectrometer (GC-MS).

When MS is used to confirm the identity of a microbial constituent (e.g., a fatty acid), molecules are bombarded with electrons in the gas phase (electron impact) or with chemicals (chemical ionization); this adds a single charge to the molecule needed for separation by mass, as positive or negative ions, within the mass spectrometer. As the name GC implies, the molecule must be volatile; large molecules such as peptides or oligonucleotides will not pass through a GC and do not ionize adequately in the electron impact or chemical ionization MS process. Structural monomers (fatty acids or sugars, both of
which are used in microbial profiling) generally serve as the subject of these analyses.

Microbiologists are often not well versed in performing organic chemical reaction schemes, and thus fatty acid profiling is limited to laboratories with an emphasis on microbial biochemistry. However, fatty acid profiling is still considered a gold standard in taxonomy and classification and is widely used in reference laboratories. There are also several companies that will provide a fatty acid profile. Sample preparation for fatty acid analysis takes several hours, and each GC run generally takes around 20 to 30 min. The analysis of sugars requires several more chemical derivatization steps and is more complicated. For example, the spore of Bacillus anthracis can be discriminated from that of Bacillus cereus by its carbohydrate profile (11). Both species contain rhamnose, 3-O-methyl rhamnose, and galactosamine, but *B. cereus* additionally contains 2-O-methyl rhamnose and fucose. The multistep derivatization procedure takes over 50 h from beginning to end. In the next section, a one-step procedure for peptide marker identification to achieve a similar species discrimination, which takes around 10 min, is discussed for comparison. The difference in the times taken for the two analyses (hours versus minutes) gives a perspective on how things have changed. Additionally, the presence or absence of a sugar (or fatty acid monomer) provides considerably less specificity than a peptide sequence.

**LARGE MOLECULES AND CLINICAL MICROBIOLOGY: BACTERIAL IDENTIFICATION AFTER CULTURE (THE PRESENT)**

In the newer, so-called soft ionization MS technology, introduced in the 1980s and 1990s, biomolecules are analyzed without any separation of components or after separation employing high-performance liquid chromatography (LC) or electrophoresis. This is performed in the liquid phase, which is often aqueous in nature. Small molecules can be analyzed, but the real power of the technique is in its being amenable to analysis of large molecules (e.g., oligonucleotides and PCR products as well as peptides and proteins) without chemical pretreatment.

For the non-mass spectrometrist, it should be pointed out that nowadays the analysis of these large molecules is based primarily on MALDI-TOF MS or ESI MS. In the former case, the sample is spotted, with a matrix, on a metal plate and allowed to air dry. When struck with a laser beam, after the plate is inserted into the MS, the matrix absorbs the light, transferring it to the molecule of interest (e.g., DNA or proteins). Generally, only a singly ionized species having a single charge is produced. In contrast, ESI MS is performed in solution, and the sample is sprayed into the MS by using a syringe pump. As the droplets evaporate, charges are transferred to molecules present within the droplet. Ions that can have multiple charge states are produced. Since mass analyzers generally separate by the mass-to-charge ratio, simple spectra (for molecules having only one charge) are generated for MALDI-TOF, but ESI spectra (reflecting mixtures of molecules each having one, few, or multiple charges) are more complex. Thus, MALDI-TOF MS has been more popular with microbiologists because of the simplicity of the spectra. However, ESI MS often allows the analysis of larger molecules. An extensive knowledge of chemistry is not required for performing MALDI-TOF or ESI MS, since the molecule is analyzed in its native form without chemical treatment. Indeed, as mentioned above, in certain applications it is not necessary to employ a separation stage (i.e., LC or electrophoresis), and the sample can be analyzed directly in the MS with minimal sample pretreatment.

Another independent but equally important instrumental advance has been the commercial introduction of the tandem mass spectrometer (in the 1990s and continuing into the 2000s), allowing routine sequencing of peptides. An MS-MS instrument takes oligonucleotides or peptides (volatilized in the MS as intact molecules) and breaks them into their constituent ions. For example, in the following simplistic illustration, the peptide A-C-D (alanine-cysteine-aspartate) might produce A and CD on MS-MS analysis. The observation of a mass of 71 suggests that position 1 is A. This is confirmed by the difference in mass between ACD and CD, suggesting that CD makes up amino acids 2 and 3 of the peptide. The finding of a dipeptide of the mass of AC suggests that C is linked to A, i.e., is at position 2. The sequence is thus A-C-D. This is illustrated in Fig. 1. In practice, the peptides are much longer in sequences (e.g., 20-mers), and there are automated programs that perform the sequence interpretations. The sequences of these peptides are then compared to protein databases (generated from the genomic sequence), and the protein is then identified and the sequence of the entire, usually much larger, protein determined.

Direct extraction of bacterial vegetative cells or spores followed by MALDI-TOF MS analysis has become popular for bacterial identification, since it is simple to perform and mass spectra are readily interpreted. However, only high-abundance peptides that are of low mass and ionize readily are observed (e.g., those in the 2,000- to 10,000-Da mass range). Generally the spectra are plotted as the amount of each protein present (as defined by its molecular weight [MW]). Unfortunately, MW alone is not sufficient to identify a characteristic biomarker, and one must rely on the entire spectrum; this is often referred to as mass profiling or fingerprinting. These spectral comparisons can be made by eyeballing, but generally pattern recognition-based computer programs are employed; unfortunately, there is often considerable variability in the spectra from run to run or between samples, complicating data interpretation (1, 4, 13, 16).

Alternatively, the sequences of individual proteins can be determined using MS-MS (5). The presence of an individual
marker can be determined with great confidence, and one does not have to depend on the consistency of the mass profile, which can sometimes be problematic. For example, in our recent work, the MWs of small acid-soluble proteins (SASPs) were measured using MALDI-TOF MS and confirmed by ESI MS. ESI MS-MS analysis was employed for the generation of sequence-specific information. The analysis consists of simply extracting the samples and introducing the extract directly into the MS-MS instrument without further sample treatment (2). ESI MS revealed a prominent doublet of SASPs for all strains in these studies. The first SASP varied in mass and sequence between \textit{B. anthracis} and \textit{B. cereus}/\textit{B. thuringiensis}. The second SASP had the same MW for all strains regardless of species and served as an internal standard allowing comparison between mass spectra in this study and previous ones. The molecular mass determined by MALDI-TOF MS of the first SASP generally differed by 32 mass units (mass of 6,679 versus 6,711) between \textit{B. anthracis} and \textit{B. cereus}/\textit{B. thuringiensis}, and the difference was identified as an alanine-to-serine switch (16 mass units) and a phenylalanine-to-tyrosine switch (also 16 mass units) (2). As noted above, the entire sample extraction and analysis take less than 10 min.

It should be emphasized that protein profiling is distinct from classical proteomics-based approaches, which involve more time-consuming sample processing. Proteomics often employs two-dimensional gel electrophoresis to isolate individual protein spots, which are then digested in situ, usually with trypsin, to generate peptides of characteristic masses that are subsequently analyzed using MALDI-TOF MS (or TOF-TOF MS-MS) analysis. Alternatively, after tryptic digestion of whole cells, the mixture of peptides is subjected to on-line liquid LC-ESI MS-MS analysis. In either case, separation (electrophoresis or chromatography, respectively) is important in reducing the complexity of mixtures for analysis by the mass spectrometer but increases the learning curve in implementing the MS technology for routine applications (17). It is noteworthy that the LC separation is more difficult to perform than the MS or MS-MS analysis or electrophoresis.

Proteomics is quite time-consuming and technically demanding and is best used for comparing the relatedness of two strains or species (6). Bioinformatics can be used to relate identified peptides to those predicted to be present in proteins coded by whole genomes. In theory, a novel strain could be categorized in this fashion. This requires bioinformatics analysis of multiple strains of each pair or group of organisms to be discriminated, which is complex and labor-intensive. Alternatively, LC-MS-MS or two-dimensional gel electrophoresis–MS-MS could be used for the process of marker discovery. Once the markers have been discovered, simple MS or MS-MS assays could be employed for routine analysis. The analogy is the discovery of DNA markers by whole genomic comparison followed by real-time PCR for diagnostic applications.

**CLINICAL MICROBIOLOGY: BACTERIAL SPECIATION, USING MS-MS WITHOUT PRIOR CULTURE (THE FUTURE)**

Sensitivity and specificity are both of particular importance in trace detection of microbial markers in complex biological matrices such as infected body fluids or tissues. Indeed, there is usually a separation (e.g., by GC or LC) or PCR amplification of the target marker in clinical diagnosis. In all instances this serves to increase the concentration of the marker relative to the background derived from other components of the matrix, and this simplifies the analysis.

Indeed, there has been notable success in detecting small molecules by using GC–MS-MS for trace analysis in clinical samples (9, 12, 15). In this form of MS-MS (GC–MS-MS), the mass spectrometer is used to focus on molecules of interest, and the detection limit is approximately 100-fold lower than that observed for GC-MS. For example muramic acid, found in the cell walls of most bacteria, is readily detected in the spinal fluids of patients with pneumococcal meningitis (15) or synovial fluids from septic arthritis (15) at levels of $<$12 ng/ml.

3-Hydroxy fatty acids, found in gram-negative bacteria, have been successfully used in the diagnosis of periodontitis (9). However, like for its counterpart GC-MS for bacterial profiling, sample processing is time-consuming and technically demanding. Furthermore, detection is not at the species level, which requires DNA or protein-sequence information. However, the success of GC–MS-MS for analysis of clinical samples provides a prototype approach for what might be done in more advanced (and more rapid) detection of peptide markers.

Real-time PCR is the current leading non-culture-based technology for determination of infection. More discriminating PCR-MS for bacterial DNA markers was developed in the United States through collaboration between the University of South Carolina and Pacific Northwest National Laboratory (14). An automated commercial PCR-MS instrument was subsequently introduced by others (7). PCR-MS has several additional stages versus PCR, including post-PCR sample cleanup and robotic transfer from PCR to an MS module. Thus, PCR-MS is currently performed as a reference laboratory technique. For example, it has been successfully used for determining nucleotide compositions for strain typing in epidemiological studies of outbreaks of respiratory infections with \textit{Streptococcus}, \textit{Haemophilus}, or \textit{Neisseria}.

Similar information can be obtained, at the protein and DNA levels, by targeting appropriate biomarkers. Indeed, efforts have been made to simplify complex samples to allow trace detection of marker peptides (using monoclonal antibodies) to purify the marker of interest prior to MS (or MS-MS) analysis. However, it is difficult to apply such approaches for universal detection, since the appropriate antibody must be available for each organism of interest. One of the notable successes in moving protein characterization from identification to detection for protein markers employs phages which amplify on binding to bacterial pathogens prior to MS. Phage proteins that are below the detection limit prior to amplification are detected (18). The amplification provides results similar to those with PCR but is much simpler to perform, since the only reagent is the phage. This method shows great promise for targeting currently known pathogens but will be harder to implement for emerging infections for which phages may not have yet been defined.

An alternative approach might involve the use of microfluidics, which has been successfully used in commercial instruments for determination of the MWs of both proteins and DNA. Separations on commercial microfluidic chips (a form of capillary electrophoresis) generally take only a few minutes,
and sample processing is carried out automatically within the chip (3). These instruments are much simpler to use than conventional gel electrophoresis, and there is a short learning curve. Whether microfluidic devices can be successfully coupled with MS-MS for dedicated trace detection of protein markers for bacterial pathogens in clinical samples only the future will tell.

CONCLUDING COMMENTS

After biological amplification by culture, bacterial identification by mass profiling (MALDI-TOF MS) or biomarker sequences (more advanced MALDI-TOF MS-MS or ESI MS-MS) involves minimal sample preparation and is competitive with widely used molecular biology approaches that often involve several sample-processing steps (e.g., PCR). However, the use of mass spectrometers is still daunting to many in the microbiology community. It is hoped that this review will contribute to removing some of the mystery behind what is ultimately a simple tool that is highly amenable to automated sample preparation and computer-based decision making. Direct detection of protein biomarkers in clinical matrices, without culture, remains a challenging but solvable task.

REFERENCES