A century of mass spectrometry: from atoms to proteomes

John R Yates III

Long before mass spectrometry became an important tool for cell biology, it was yielding scientific insights in physics and chemistry. Here is a brief history of how the technology has expanded from a tool for studying atomic structure and characterizing small molecules to its current incarnation as the most powerful technique for analyzing proteomes.

The constant stream of impressive and important advances in biological mass spectrometry over the last two decades may give the impression that the technology is a recent innovation. In fact, mass spectrometry has had a long and interesting history and a pivotal role in many important scientific advances since 1900 (**Fig. 1** and **Supplementary Fig. 1**) (http://masspec.scripps.edu/mshistory/mshistory.php).

Fundamentally, a mass spectrometer is used to measure the mass-to-charge ratios (m/z) of ions, a metric from which we can determine molecular weight. This process involves three steps. First, molecules must be converted to gas-phase ions, which poses a considerable challenge for molecules in a solid or liquid phase. Next, ions are separated by their m/z values via magnetic or electric fields in a component called a mass analyzer. Finally, the separated ions and the abundance of each species with a particular m/z value are detected.

The earliest mass spectrometers

In its early history at the beginning of the 20th century, mass spectrometry was used to probe fundamental aspects of atomic and molecular structure, driving the determination of atomic weights of elements and the discovery of stable isotopes. The birth of mass spectrometry is commonly attributed to the physicist J.J. Thomson with his discovery of the electron, using an electric field inside a cathode ray tube. His success led him to develop a crude 'mass spectrograph' to measure atomic weights of elements[1]. Thomson's students F.W. Aston and A.J. Dempster went on to develop more advanced versions of this instrument; Aston's discovery of ^{20}Ne and ^{22}Ne (ref. 2) proved the existence of stable isotopes. Physicists designed increasingly sophisticated instruments, ushering in many scientific advances. For example, Alfred Nier's improvements to the design of the mass spectrometer allowed more sensitive and precise measurements of isotopes and their ratios, and permitted the discovery of the third and very rare isotope of uranium, ^{234}U (ref. 3). By cleverly using the mass spectrometer as a preparative tool, Nier and colleagues showed that the fissionable isotope of uranium was ^{235}U. Nier also discovered the ^{13}C isotope and subsequently purified the isotope for use in tracer experiments to understand metabolic pathways.

The onset of World War II saw more applied uses of mass spectrometry. It was used in the Manhattan project to purify and assess the enrichment of the fissionable isotope of uranium. Nier constructed a device based on a mass spectrometer to detect leaks in the gas centrifuges used to enrich ^{235}U. Such an instrument was sorely needed as the uranium hexafluoride used in the purification process was extremely corrosive. As the demands for higher-octane fuel to improve performance of fighter aircraft increased during the war, mass spectrometers were used to monitor petroleum processing to increase fuel quality. These needs drove a better understanding of molecular ionization and fragmentation to create methods for producing more reproducible and diagnostic mass spectra. After the war, the petroleum industry increasingly used mass spectrometry for the molecular characterization of oil and the development of oil-based products, boosting fundamental research studies of gas-phase ionization and fragmentation. Researchers in academic laboratories continued to develop new types of instruments, and the creation of the double-focusing magnetic sector instrument that used an electric sector to correct for kinetic energy spread in ions before separation in the magnetic field ushered in high-resolution mass spectrometry, leading to better mass accuracy and peak capacity.

The first biomolecules

From the late 1950s and into the early 1960s, Klaus Biemann led efforts to use mass spectrometry to measure the molecular weight of small molecules to verify their structure[4]. Biemann developed a method to use mass spectrometry to compare mass spectra of natural products to those of synthetic products to confirm structural assignments. Previously, structural confirmation was performed either by the arduous and imprecise method of melting-point analysis—in which the structure was deemed incorrect if a mixture of authentic and synthetic material had a lower melting point than the authen-

John R. Yates III is in the Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California, USA.
e-mail: jyates@scripps.edu

tic material—or through synthetic deconstruction of the molecule to something of known structure. These processes could take years, and Biemann's new approach took only weeks. The success of Biemann's approach kicked off the use of mass spectrometers for the characterization of small molecules and natural products, a use that continues to this day.

The technology used by Biemann, electron impact ionization, created fragment ions by using a beam of energetic electrons to bombard volatilized molecules. Though the technology had been developed during World War II, it was not well understood how ions fragmented during this process. With the new-found interest in the use of mass spectrometry for molecular characterization came the realization that a detailed understanding of the mechanisms of fragmentation could be used to determine structures *de novo*. Through painstaking studies incorporating stable-isotope labels into specific positions in molecules to follow the stable-isotope label during fragmentation, the mechanisms of ion fragmentation were slowly teased apart. As an understanding of how ions fragmented grew, a group at Stanford University led by Joshua Lederberg deemed the interpretation of mass spectra an ideal problem to drive the development of algorithms based on artificial intelligence in the Dendral Project. The goal was to mimic the human thought process used for interpretation of mass spectra. To this day, however, reliable automated *de novo* interpretation of small molecule mass spectra has not been achieved, but this project was an early realization of the synergies between mass spectrometry and computers.

Biemann's efforts in the late 1950s quickly turned to biomolecules[5]. With electron-impact ionization, even molecules of limited volatility exhibited enough vapor pressure to obtain a mass spectrum. Biomolecules, however, posed a greater challenge: they are polar and often charged, and thus have no vapor pressure even at 10^{-6} torr. Making biomolecules compatible with mass spectrometry required derivatization to reduce the polarity or charge of the molecules. By derivatizing free amino groups and carboxylic acids in peptides to reduce polarity and charge, sufficient volatility could be created to acquire a mass spectrum of short peptides (two to three amino acids).

Figure 1 | A timeline showing important advances in mass spectrometry over the last century. Highlighted are instrument developments (blue), applications of mass spectrometry (gray) and general scientific achievements (green). LC-MS/MS, liquid chromatography, tandem mass spectrometry.

Peptide sequencing

By the late 1960s two important technological advances emerged to enable sequencing of peptides using mass spectrometry. The first was the invention of chemical ionization, a method of ionization that induced very little fragmentation in peptide ions[6]. Chemical ionization uses gas-phase ion-molecule reactions between a reagent ion such as a CH_4^+ and the volatilized analyte. A proton is transferred to the molecule, creating a protonated ion, and very little energy is imparted to the new ion in this 'soft ionization' process, so there is little fragmentation. This was a dramatic improvement over electron-impact ionization, which induced extensive ion fragmentation. For some ions with labile structures, ionization by electron impact would leave little molecular ion from which to infer molecular weight. The second innovation was the development of methods to interface the gas chromatograph to the mass spectrometer (GC-MS), which allowed the separation of molecules online to the mass spectrometer[7,8]. A means to both separate and characterize the structure of molecules in tandem was an important milestone in analytical chemistry as no longer did a pure sample have to be introduced into the mass spectrometer.

In the late 1960s and early 1970s mass spectrometry techniques for peptide and protein sequencing began developing at a rapid pace. Biemann developed a strategy using polyamino alcohol derivatives of peptides and detection in a GC-MS–based strategy[9], allowing *de novo* sequencing of peptides and checking DNA-derived sequences. By the end of the 1970s this strategy was highly automated, including computer interpretation of mass spectra. The drawbacks of the approach were the extensive derivatization procedure, and the limited sensitivity and the involatility of larger peptides (tetramers and greater), which restricted the approach to di- and tripeptides. In a competing strategy developed in 1969, Howard Morris derivatized peptides by permethylation[10]. Although not amenable to separation by gas chromatography, permethylated peptides could be fractionally distilled from the sample probe to produce easily interpretable mass spectra of larger peptides.

In mid-1970s, the triple quadrupole mass spectrometer was developed, and Donald Hunt began developing a peptide sequencing strategy based on tandem mass spectrometry[11]. A tandem mass spectrometer uses a first-stage mass analyzer to select an ion that is then passed into a collision cell where it collides with gas molecules such as argon (via a technique called collision-induced dissociation, CID) causing the ion to fragment. The fragment ions are then analyzed in the second-stage mass analyzer. The tandem mass spectrometry sequencing strategy overcame the limited fragmentation of peptide ions created by soft ionization techniques and improved the ability to analyze peptide mixtures by precise selection of co-ionizing peptides. Despite these advances, however, the central weakness of peptide sequencing strategies in this era was the need to derivatize peptides to increase volatility for mass spectrometry analysis.

New ionization techniques

Paradigm shifts in mass spectrometry have been driven by the development of new ionization methods, and the start of the 1980s saw a big one. In 1981 Michael Barber and colleagues introduced fast atom bombardment (FAB), a soft ionization method that required no derivatization[12]. FAB used a fast beam of atoms (or ions) to desorb molecules from a glycerol matrix. The fast moving atoms (~8,000 electron volts) struck the glycerol surface and sputtered the glycerol and sample molecules into the gas phase. This new capability spurred furious efforts in peptide-sequencing methods. Still, despite the excitement this new method generated, the size of molecules that could be ionized was limited to thousands of daltons (a huge advance at the time nevertheless) and the method required the use of offline separation techniques such as high-performance liquid chromatography. Additionally, because FAB created very little ion fragmentation, tandem mass spectrometry was required to achieve the peptide fragmentation necessary for accurate sequencing and to allow the analysis of individual peptides present in a mixture[13]. Despite these limitations, the introduction of FAB drove the generation of mass spectrometers with expanded mass range and led to the creation of new tandem instruments and types of tandem hybrids that combined different types of mass analyzers. This time period saw the development of many of the ideas and methods for sequence analysis of peptides and modified peptides that laid the groundwork for the development of proteomics. By the end of the decade the dominant tandem mass spectrometers were the tandem double focusing magnetic sector instruments and the triple quadrupoles, but looming on the horizon, two new ionization techniques would soon emerge that would rearrange the peptide sequencing landscape[14].

Even with improvements in the ability to analyze biomolecules with the invention of FAB, applications were still limited and far from robust. Two new ionization techniques for biomolecules emerged at the end of the 1980s that transformed the field: matrix-assisted laser desorption ionization (MALDI) developed by Michael Karas and Franz Hilllenkamp, and electrospray ionization (ESI) developed by John Fenn[15,16]. MALDI is similar to FAB in that it requires a sample be placed on a stage, but rather than being bombarded by fast particles, samples are 'bombarded' by photons from a UV-light laser. Its unique capability to ionize very large proteins, carbohydrates and even DNA as singly charged ions, allowed accurate mass measurement of large molecules never before possible. At the time of development, the only instrument that could accommodate such a large mass range and the pulsed nature of the ionization technique was the time-of-flight mass spectrometer, a mass analyzer that separates ions based on their flight times over a defined distance.

ESI provided a very different set of capabilities than MALDI. By placing a high voltage on a liquid flowing through a capillary, a spray containing tiny droplets forms at the exit, and the spray is directed to the entrance of the mass spectrometer. The liquid is removed from the droplet by heat and results in the formation of an ion beam. ESI also results in the formation of multiply charged ions, which ensures that m/z of large molecules are in the range of 800–1,800 atomic mass units, well within the detection capabilities of many mass spectrometers used at the time. Perhaps even more exciting was that, with ESI, the user could now achieve the long-desired goal of interfacing liquid-phase separations to a mass spectrometer[17]. This meant that polar and charged molecules could be separated online allowing robust, direct analysis of biological mixtures. ESI interfaced quite easily to triple quadrupole mass spectrometers, but because the four-sector mass spectrometer's ion source operated at

high voltage (8,000–10,000 electron volts) it was difficult to interface ESI on this type of instrument.

Within a few years, the four-sector mass spectrometer virtually disappeared from the market as the field moved increasingly toward the use of ESI. The excitement of finally being able to perform robust and highly sensitive analysis of biomolecules quickly began to spill over into other fields. At about this same time discussions to sequence the human genome and the genomes of model organisms were at a fever pitch—projects that would have a great impact on the field of mass spectrometry. Ironically, none of the discussions or rationales for sequencing genomes suggested that protein biochemistry might benefit from the data created.

The rise of proteomics

This new and robust capability to analyze biomolecules, in the form of ESI, drove improved processes for peptide and protein sequence analysis. But although the mass spectrometry technology was in place, the means to analyze the data generated lagged. For the mass spectrometry community, data interpretation was perhaps one of the pleasures of the process—a bit like the Sunday *New York Times* crossword puzzle is for some—but the complexity of the data-interpretation process kept all but the most fearless nonspecialists at bay. In a stroke of synergy between mass spectrometry data and the sequence data produced in the genome-sequencing projects, researchers discovered that the molecular weights of peptides from protein digests and tandem mass spectra of peptides could be matched to sequences in the genome databases. By using MALDI–time-of-flight to analyze a proteolytically digested protein obtained from a two-dimensional gel electrophoresis separation, a peptide mass map or fingerprint could be used to search a sequence database and identify a protein based on matching the experimental and predicted peptide molecular weights[18]. Proteins need to be digested, as post-translational modifications and proteolytic processing sufficiently modify a protein's molecular weight from that predicted by the gene as to make matching by molecular weight unreliable. This capability to obtain peptide fingerprints allowed the identification of proteins obtained by a powerful separation tool and marked the beginning of the proteomics era.

The ability to identify proteins in mixtures and their post-translational modifications, using tandem mass spectra and sequence databases[19], altered the existing protein-sequencing paradigm and ushered in an approach known as shotgun proteomics—a strategy that makes possible large-scale measurements of proteins in protein complexes, organelles, cells and tissues as well as their post-translational modifications, without the need to isolate individual proteins[19–22]. Rapid and accurate identification of proteins greatly increased the pace of biological discovery by simplifying what used to be the most time-consuming part of discovery—protein sequencing. These new capabilities and applications in turn drove rapid developments in tandem mass spectrometry technology to create new types of mass spectrometers such as linear ion traps, quadrupole/time-of-flights, time-of-flight/time-of-flight, ion mobility and Orbitrap instruments with greater sensitivity, faster scan speeds, better resolution and higher mass accuracy. Many of these instrumental developments also proved useful for the analysis of small molecules such as metabolites, giving rise to global measurements encompassing the metabolome. These technologies have also become essential to the drug-development process to identify and measure drug metabolites.

The ESI technology also facilitated the evolution of methods to characterize intact proteins. Because proteins are processed and modified post-translationally, their molecular weight rarely agrees with that predicted value from the gene sequence; consequently, methods to fragment intact proteins were needed[23]. Intact proteins are difficult to fragment in the gas phase because intramolecular hydrogen and electrostatic bonding produces very stable structures, and this has stimulated the development of new fragmentation methods such as electron capture dissociation. As the analysis of large proteins required very high mass resolution, the drive to develop 'top-down' analysis of proteins has led to the development of higher-magnetic-field Fourier transform mass spectrometers with ultrahigh mass resolution. Top-down analysis of proteins is advancing rapidly and will have an important role in understanding the functions of different protein species. Furthermore, the ability to perform top-down analysis of protein complexes provides an important avenue for structural characterization of these important cellular components[24].

Conclusions and outlook

Since its invention, mass spectrometry has had a prominent role in the science of the day. In the early part of the 1900s it was a critical technique in the study of atomic structure. During World War II mass spectrometry technology was used in more practical applications such as enrichment of uranium and process control to create better petroleum projects. By the 1960s mass spectrometry was developed to assist in the elucidation of natural product structures during the salad days of total chemical synthesis. Mass spectrometry has become an essential tool for the pharmaceutical industry to determine the metabolites of drugs and to measure important pharmacokinetic data on drug behavior. Large-scale studies of metabolites are emerging to drive the emerging field of metabolomics. The genome projects have made this the 'century of biology', and mass spectrometry is again an integral part of cutting edge science by driving biological discovery.

The success of proteomics has been driven by innovation in the development of mass spectrometers. Each new development increases capability. For example, mass spectrometry almost can be used to identify all proteins of a cellular proteome on a time scale compatible with experimentation. In time, deep sequencing experiments will be possible that identify all isoforms and modifications. Such goals may require a combination of bottom-up shotgun proteomics and top-down analysis of proteins. Large-scale identification of modifications such as phosphorylation are now possible with enrichment, and these types of experiments are allowing comparisons of phosphorylation states between tissues, which should in time contribute to a better understanding of diseases[25]. Mass spectrometry has already made enormous contributions to cell biology through the study of protein-protein interactions, organelles, cells and post-translational modifications. In addition, mass spectrometry has driven the search for markers of disease and has contributed to the discovery of the causes of diseases. Mass spectrometry has migrated from being an important technology for physics to chemistry and, if history is any guide, innovation in mass spectrometry will con-

tinue to make it an essential tool for biology. Sources used for the history of mass spectrometry included reference 26.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

I thank C. Delahunty and X. Han for reading early drafts and acknowledge funding from the US National Institutes of Health (P41 RR011823, R01 MH067880 and R01 HL079442).

COMPETING FINANCIAL INTERESTS

The author declares no competing financial interests.

1. Thomson, J.J. *Philos. Mag.* **44**, 293–316 (1897).
2. Aston, F.W. *Philos. Mag.* **39**, 449–455 (1920).
3. Nier, A.O. *Rev. Sci. Instrum.* **11**, 212 (1940).
4. Biemann, K. *J. Am. Soc. Mass Spectrom.* **13**, 1254–1272 (2002).
5. Biemann, K., Gapp, G. & Seibl, J. *J. Am. Chem. Soc.* **81**, 2274–2275 (1959).
6. Munson, M.S.B. & Field, F.H. *J. Am. Chem. Soc.* **88**, 2621–2630 (1966).
7. Gohlke, R.S. *Anal. Chem.* **31**, 535–561 (1959).
8. Ryhage, R. *Anal. Chem.* **36**, 759–764 (1964).
9. Nau, H., Kelley, J.A. & Biemann, K. *J. Am. Chem. Soc.* **95**, 7162–7164 (1973).
10. Lucas, F. *et al. Biochem. J.* **114**, 695–702 (1969).
11. Hunt, D.F., Buko, A.M., Ballard, J.M., Shabanowitz, J. & Giordani, A.B. *Biomed. Mass Spectrom.* **8**, 397–408 (1981).
12. Barber, M., Bordoli, R.S., Sedgwick, R.D. & Tyler, A.N. *Nature* **293**, 270–275 (1981).
13. Hunt, D.F., Bone, W.M., Shabanowitz, J., Rhodes, J. & Ballard, J.M. *Anal. Chem.* **53**, 1704–1708 (1981).
14. Scoble, H.A., Martin, S.A. & Biemann, K. *Biochem. J.* **245**, 621–622 (1987).
15. Karas, M. & Hillenkamp, F. *Anal. Chem.* **60**, 2299–2301 (1988).
16. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. *Science* **246**, 64–71 (1989).
17. Smith, R.D. & Udseth, H.R. *Nature* **331**, 639–640 (1988).
18. Henzel, W.J. *et al. Proc. Natl. Acad. Sci. USA* **90**, 5011–5015 (1993).
19. Eng, J., McCormack, A. & Yates, J. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
20. McCormack, A.L. *et al. Anal. Chem.* **69**, 767–776 (1997).
21. Link, A.J. *et al. Nat. Biotechnol.* **17**, 676–682 (1999).
22. Ficarro, S.B. *et al. Nat. Biotechnol.* **20**, 301–305 (2002).
23. Kelleher, N.L. *et al. J. Am. Chem. Soc.* **121**, 806–812 (1999).
24. Ruotolo, B.T. *et al. Science* **310**, 1658–1661 (2005).
25. Huttlin, E.L. *et al. Cell* **143**, 1174–1189 (2010).
26. Grayson, M. (ed.) *Measuring Mass: From Positive Rays to Proteins* (Chemical Heritage Press, Philadelphia, 2002).