Mass Spectrometry in the Postgenomic Era

Brian T. Chait

Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10021; email: chait@rockefeller.edu

Keywords

cellular systems, proteomics, protein complexes, native mass spectrometry, lipidomics

Abstract

Mass spectrometry (MS) is rapidly becoming an essential tool for biologists and biochemists in their efforts to throw light on molecular mechanisms within cellular systems. Used in unison with genome sequence data, MS has developed into the method of choice for identifying proteins, elucidating their posttranslational modifications, and reading out their functional interactions. Variations of the method have even begun to enable accurate mass determination of intact protein complexes, allowing for direct determination of subunit stoichiometry and the interactions between the subunits. Advances in mass spectrometric technologies have also led to great improvements in our ability to probe and define many of the other key molecular players in cells, including the all important lipid components. We provide here some perspectives on the explosion of applications of MS to protein science, systems biology, proteomics, lipidomics, and cell biology in general.
Cellular systems are distinguished by their extraordinarily complex hierarchy of structures, organized in space and time through the interactions and milieus set up by a very large ensemble of distinct biomolecules. To gain a better understanding of such systems, there is an urgent demand for tools that are able to elucidate all of these molecular players as well as their detailed molecular organization at all hierarchical levels. Toward this goal, effective methods have been developed to rapidly sequence entire genomes as well as to define the mRNA content of cells. More challenging has been the development of equally effective tools for analyzing the other constituents of cellular systems (including proteins and lipids) and for elucidating the dynamic functional interactions among these biomolecules.

Conceptually, one of the simplest and most powerful ways of identifying and characterizing a biological molecule is to determine its accurate molecular mass together with the masses of its component building blocks after fragmentation (1). Historically, this simple mass spectrometry (MS) approach proved remarkably elusive until robust methods were found to transfer highly polar biopolymers like peptides and proteins from the solution or solid state into their intact, ionized counterparts in the gas phase (2). Currently there are two dominant means for doing this.

The first is electrospray ionization (ESI), in which the molecular ions of interest are formed directly from solution by applying a high electric field to the tip of a capillary through which the solution passes (3). Properties that make this a method of choice for biological applications include (a) the “softness” of the phase conversion process, allowing very fragile molecules to be ionized intact and in some cases even noncovalent interactions to be preserved for MS analysis; (b) ready coupling to liquid chromatography (LC) such that the eluting fractions can be directly sprayed into the mass spectrometer, allowing for the analysis of complex mixtures; and (c) the natural and efficient production of multiply charged ions, allowing for the measurement of high-mass biopolymers using mass spectrometers with modest mass/charge ranges (because mass spectrometers measure mass/charge rather than mass). In addition, multiple charges on a molecule often facilitate improved fragmentation for structure elucidation and identification.

The second is matrix-assisted laser desorption/ionization (MALDI) in which the molecular ions of interest are formed by pulses of laser light impacting on the sample isolated within an excess of matrix molecules (most often a solid) (4). Advantageous properties of MALDI include its extraordinary robustness; high speed; and relative immunity to contaminants, biochemical buffers, and common additives.

In concert with the introduction of these enabling ionization techniques, a plethora of increasingly powerful mass analyzers have been developed that allow for the measurement of the mass/charge of intact ionized biomolecules with high accuracy (e.g., parts per million for peptides) as well as of their fragmentation spectra with high speed (as many as 5–10 spectra/s) (5). The measurement of these fragmentation spectra is termed tandem MS (or MS/MS) to differentiate the analysis from the single-stage MS of their intact precursor ions. The wide range of instrumentation that has become available confronts even seasoned practitioners with bewildering choices. Often, it is the specific application that dictates the optimal choice. So, for example, the LC-ESI-linear ion trap coupled to an Orbitrap (6) has become a favored instrument for many proteomics experiments, and the nano-ESI-quadrupole-orthogonal injection time-of-flight mass spectrometer (7) has become a favored tool for investigating noncovalent interactions. However, that said, many other combinations can also yield superb results, and MS instruments are constantly being refined and new ones developed. In practice, this choice is usually dictated by what is readily available to the investigator. However, if researchers wish to acquire mass spectrometers for their own use, my advice is to take well-defined test samples that represent their most important biological applications to the various manufacturers and determine for
themselves the efficacy of the instruments for their specific applications. In this way, in addition to the more usually quoted instrumental parameters, such as mass accuracy, sensitivity, data acquisition speed, and resolution, they can readily determine other important instrumental properties, such as robustness and ease of use of both the instrument and related software.

These advances have made MS a method of choice for a host of biological applications, including, for example, rapidly identifying proteins, determining details of their primary structures, and elucidating their interactions. In principle, there are two complementary lines of attack for the MS analysis of proteins: the bottom-up and top-down approaches (5, 8). In the bottom-up approach, proteins of interest are digested with an enzyme, such as trypsin, and the resulting “tryptic peptides” are analyzed by MS and MS/MS. In the top-down approach, the intact proteins are directly analyzed in the mass spectrometer without prior solution digestion. The advantages of the bottom-up approach are that the relatively small tryptic peptide ions are more uniform and easy to handle biochemically than intact protein ions, their masses are easier to determine with high accuracy, and they can be more readily induced to fragment. These advantages have made the bottom-up approach the dominant method for the majority of proteomic studies [see Cox & Mann (9)], even though the coverage of the protein via analysis of tryptic peptides is usually incomplete. The potential advantages of the top-down approach are that it can provide complete coverage of the protein as well as the positional correlation of multiple dispersed modifications. However, because of (a) the significant challenge in handling whole proteins (versus small peptide pieces), (b) issues relating to the omnipresent heterogeneity of intact proteins, and (c) the often complex nature of the analysis, such top-down analyses are still largely confined to low-throughput single-protein studies (10–12). However, it is noteworthy that an intermediate “middle-down” approach for analyzing proteolytic peptides larger than typical tryptic peptides is beginning to prove useful, as, for example, in the elucidation of the complex array of modifications on histone tails (13).

In the articles that follow, the authors review the application of cutting-edge biological MS to the study of proteins and lipids, with the ultimate goal of elucidating their detailed roles in biological processes. Examples are provided to illustrate how MS is enabling rapid developments in proteomics, lipidomics, and systems biology. These applications range from those that can be routinely achieved by the skilled investigator armed with appropriate instrumentation to those that just approach the bare edge of feasibility.

On the more routine side, it has become straightforward to use MS to rapidly identify proteins separated as bands on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels. All that is needed is the genome sequence of the organism under study, so that proteolytic fragments generated from the protein band of interest can be subjected to both single-stage MS and MS/MS and the resulting spectra rapidly compared with those expected theoretically from the genome sequence. Here, MS serves as a “generalized Western blot” without the need to generate antibodies, and in addition, MS does not suffer from the cross-reactivity common in antibodies. But, unlike the Western blot, which by its nature is targeted, MS detects all the protein components present above the detection limit of the mass spectrometer. Thus, MS often detects a large number of protein components (some present in only trace amounts) from what may appear to be a single band; the challenge then is to figure out which of these are proteins of interest.

Identifying proteins in gel bands by MS also has been used to great effect for the definition of components of protein complexes, especially for stoichiometric complexes that have been relatively cleanly isolated (see, e.g., References 14 and 15). An alternative strategy for such cases involves shotgun MS sequencing of the peptide mixture, generated by digesting the entire complex without prior separation of the proteins, and mapping these peptides onto...
the various proteins that are present using computer algorithms (5). The most vexing problem related to both of these strategies for determining the components of protein complexes is the issue of discerning which of the identified proteins are specific to the complex versus those that arise from nonspecific interactions or background contamination. Because of the sensitivity and high dynamic range of modern MS, it is not unusual to identify several hundred proteins when affinity isolating a complex that may only have a few to a few tens of components. Biologists confronted with such long lists of proteins can easily be overwhelmed—or at least be tempted to subjectively focus on those that may seem to be most interesting—so it is highly preferable to have an objective means for deciding which are specific interactors. Fortunately, this is often possible by the strategic use of cells labeled with heavy versus light stable isotopes (16, 17). Such objective quantitative means for determining specific from nonspecific interactions or for following cellular changes as a function of time and circumstance have become essential mainstream methods in MS-based proteomics [see Cox & Mann (9)].

Similarly, MS is now a preferred method for elucidating protein posttranslational modifications. Mass changes, characteristic of modifications, can be rapidly pinpointed to specific amino acids in the sequence using relatively low amounts of protein. Such applications have been greatly facilitated by improved MS instrumentation and fragmentation technology [using, e.g., so-called higher-energy collisional dissociation (18) or electron transfer dissociation (19) in addition to the more standard lower-energy collisional dissociation] as well as by the development of methods for enriching specifically modified proteolytic peptides, via either affinity isolation or chemical derivatization (20). For example, immobilized metal ion affinity chromatography (IMAC) has been used to enrich and facilitate identification of literally thousands of phosphorylation sites within whole proteomes as a function of cell state (18, 21–23; also see 9). At the same time, it should be noted that it often remains entirely nontrivial to comprehensively characterize the phosphorylation state of even a single protein of interest. This difficulty arises because these studies are usually performed on proteolytic fragments, using the bottom-up approach, and for a number of reasons (IMAC isolation bias, peptides too small, too large, too hydrophilic), it can be very difficult to detect all of the peptide fragment ions (or even assign the peptides unambiguously to a particular splice variant). The resulting gaps in the coverage of the protein are further exacerbated when the amount of the available protein is limited or the stoichiometry of the phosphorylated forms of the protein is low. In such cases, it may be necessary to resort to the classical approach of subjecting the protein of interest to multiple digestions with proteases having different specificities, provided of course that sufficient protein is available.

An alternative line of attack that is just beginning to gain traction relies on the top-down or middle-down approaches (5, 8). Instead of analyzing small proteolytic fragments, one obtains MS and MS/MS of the whole protein or large peptide components of it. Under ideal conditions, the problem then reduces to assembling pairwise puzzle pieces (protein ion fragments) that each add up to the mass of the protein (or that of a relatively large component of the protein) rather than trying to assemble a great many small puzzle pieces, many of which may be missing. Although considerable progress has been made in this endeavor, the top and middle-down approaches are still limited by difficulties relating to the isolation and handling of small amounts of protein as well as to introducing them into the mass spectrometer and fragmenting them efficiently.

Improvements in the speed and dynamic range of modern mass spectrometers now allow for the acquisition of proteomic information at all levels of organization, including organelles and even whole cells and tissues (9). However, the task of defining an organelle, for example, is not simple; it requires a means for differentiating bona fide components from contaminants and for ensuring that low-copy-number
components are detected in the presence of high-abundance components. Identifying the authentic organelar components is often accomplished by the classical approach of following the levels of individual protein as a function of enrichment, now using quantitative MS as the primary tool for this assay (24). Identifying low-abundance components is accomplished by ensuring that (a) there is present in the sample large enough amounts of the proteins to be detected and that (b) there is sufficient separation of the low-abundance components so that their signal is not drowned out by those of higher abundance (25). Even given the high sensitivity (∼femtomole) of modern MS instrumentation, it can be taxing to ensure that sufficient material is available for analysis—as, for example, when isolating a particular synapse type from a particular cell type in mouse brain (26). The question of ensuring sufficient separation of components either prior to or during MS is a topic currently under intense investigation (9). Here, it appears that the key is to find a level of separation prior to MS that is sufficient (without unduly increasing inevitable material losses and the time for analysis) and an appropriate level of single- and multistage MS to finish the job. As the complexity of the system increases further (e.g., whole-cell, tissue, or plasma analyses), problems relating to dynamic range and interferences become increasingly challenging (9).

As outlined above, rapid and accurate MS identification of affinity isolated complexes has greatly facilitated the definition of in vivo protein assemblies and protein interactions in general. However, such affinity-based experiments do not guarantee isolation of a single unique complex—just composites of protein complexes that interact with the bait protein (15). Thus, a more direct means for accurately determining the total masses of isolated complexes is highly desirable. One emerging technology for this task is the so-called native MS, wherein ESI is used to produce intact multiply charged protein complexes; this allows for the accurate mass determination of these noncovalent complexes, enabling the direct determination of subunit stoichiometry within a protein complex. In addition, dissociation of these complexes either prior to native MS or in the mass spectrometer itself yields information regarding subunit connectivity. Finally, it appears that native MS can even be used to gain information on the overall shape of complexes through measurements of the time taken for ionized complexes to drift through gas-filled tubes maintained at appropriate pressures. In Barrera & Robinson (27), the authors show how native MS, which was originally demonstrated on water-soluble complexes, can be extended to complexes whose native environment is the lipid bilayer.

An important consideration in the native MS approach is the degree of fidelity with which the complexes present in vivo, either in solution or within a lipid bilayer, are maintained into the gas phase, where they are essentially stripped naked of solvating species. There are several steps in this transfer process that might perturb this fidelity (28). The first involves classical isolation of the complex from the cellular system and is not peculiar to native MS. The second involves transfer of the complex to a volatile buffer solution that is appropriate for ESI-based native MS (often ammonium acetate), and because many interactions are very sensitive to buffer conditions, perturbations can certainly occur at this point. The third involves transfer of the complexes from solution to desolvated complexes in the gas phase using ESI. Perhaps surprisingly, there is now a substantial body of evidence showing that with appropriate care subunit stoichiometry measured in the gas phase can faithfully reflect that observed in solution (7, 27, 29). There is also a growing body of evidence indicating that native interactions and even detailed shape are to some extent maintained into the gas phase, although considerably more work is needed to determine the generality of these findings. A related issue is the relationship between the dissociation rates of components of solvated versus desolvated complexes. This is still a subject of intense debate and investigation, with findings ranging from the expected stabilization of ionic
interactions in the absence of water to the unexpected stabilization of hydrophobic interactions in the absence of water (30).

Up to this point, we have mainly considered the utility of MS for analyzing the proteinaceous components of cellular systems. Fortunately, the described advances in MS technology have also led to great improvements in our ability to probe and define many of the other key molecular players in cells, including the all important lipid components [see Harkewicz & Dennis (31)]. The resulting revolution in lipidomics has been somewhat quieter than those for genomics and proteomics, perhaps because lipidic components are of lower molecular mass and because there has been a reasonably successful tradition of their analysis by classical MS methods, facilitated by their oftentimes lower polarity as well as by effective chemistries for neutralizing the polarity of the more polar species. The more recent advances in MS technology are helping to throw a light on a host of outstanding questions concerning lipidic components of organisms, including, for example, a more global appreciation for the subset of the enormous possible number of these compounds that are functionally used within a given organism and a deeper understanding of the role of lipidic molecules in reprogramming cellular networks by modulating protein function. The potential also exists for defining the detailed lipidic milieu wherein ion channels and other integral membrane protein systems reside, although this necessitates the development of biochemical techniques that effectively isolate the proteinaceous systems together with their local endogenous environment.

Even though biological MS technologies have undergone remarkable advances over the past two decades, further improvements and innovations are urgently needed. These include the following:

1. Improved sensitivity to allow for the analysis of a smaller number of cells or samples that are more focused on cellular components with specialized function. For example, if we wish to define all the proteins at a particular locus on a chromosome for which the occupancy of any one of these proteins may be ~1 per cell, then >10⁹ cells would currently be needed for a successful analysis. To decrease the number of cells required for such an analysis, improvements are needed in sample handling to minimize losses prior to sample introduction into the mass spectrometer as well as more efficient utilization of the sample within the mass spectrometer.
2. Improved methods for measuring low-abundance components within complex samples in the face of high-abundance components, where the required dynamic range may be >10⁶.
3. Higher analysis speed to enable deeper and more routine analyses of complex samples.
4. More robust MS tools that can be more easily used by biologists themselves. Here, one envisages a great expansion of MS as a rapid-feedback tool in much the same way as is possible for SDS-PAGE and Western blotting.
5. Improved techniques for preparing samples that “freeze” in the desired information from the cellular milieu for subsequent MS readout. Such techniques should greatly assist the acquisition of both positional and temporal information on the biomolecular structures of interest.

DISCLOSURE STATEMENT
The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.
LITERATURE CITED

# Contents

## Preface
Past, Present, and Future Triumphs of Biochemistry  
*JoAnne Stubbe* ......................................................... v

## Prefatory
From Serendipity to Therapy  
*Elizabeth F. Neufeld* ................................................. 1

Journey of a Molecular Biologist  
*Masayasu Nomura* .................................................. 16

My Life with Nature  
*Julius Adler* .......................................................... 42

## Membrane Vesicle Theme
Protein Folding and Modification in the Mammalian Endoplasmic Reticulum  
*Ineke Braakman and Neil J. Bulleid* ............................... 71

Mechanisms of Membrane Curvature Sensing  
*Bruno Antony* ........................................................ 101

Biogenesis and Cargo Selectivity of Autophagosomes  
*Hilla Weidberg, Elena Shvets, and Zvulun Elazar* ............... 125

## Membrane Protein Folding and Insertion Theme
Introduction to Theme “Membrane Protein Folding and Insertion”  
*Gunnar von Heijne* .................................................... 157

Assembly of Bacterial Inner Membrane Proteins  
*Ross E. Dalbey, Peng Wang, and Andreas Kuhn* .................. 161

β-Barrel Membrane Protein Assembly by the Bam Complex  
*Christine L. Hagan, Thomas J. Silhavy, and Daniel Kahne* ........ 189
Transmembrane Communication: General Principles and Lessons from the Structure and Function of the M2 Proton Channel, K+ Channels, and Integrin Receptors
Gevorg Grigoryan, David T. Moore, and William F. DeGrado .......................... 211

Biological Mass Spectrometry Theme
Mass Spectrometry in the Postgenomic Era
Brian T. Chait ........................................................................................................ 239

Advances in the Mass Spectrometry of Membrane Proteins:
From Individual Proteins to Intact Complexes
Nelson P. Barrera and Carol V. Robinson ............................................................... 247

Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology
Jürgen Cox and Matthias Mann .................................................................................. 273

Applications of Mass Spectrometry to Lipids and Membranes
Richard Harkewicz and Edward A. Dennis .......................................................... 301

Cellular Imaging Theme
Emerging In Vivo Analyses of Cell Function Using Fluorescence Imaging
Jennifer Lippincott-Schwartz .................................................................................. 327

Biochemistry of Mobile Zinc and Nitric Oxide Revealed by Fluorescent Sensors
Michael D. Pluth, Elisa Tomat, and Stephen J. Lippard ............................................. 333

Development of Probes for Cellular Functions Using Fluorescent Proteins and Fluorescence Resonance Energy Transfer
Atsushi Miyawaki ..................................................................................................... 357

Reporting from the Field: Genetically Encoded Fluorescent Reporters Uncover Signaling Dynamics in Living Biological Systems
Sohum Mehta and Jin Zhang .................................................................................... 375

Recent Advances in Biochemistry
DNA Replicases from a Bacterial Perspective
Charles S. McHenry .................................................................................................. 403

Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors
Peter C. Hollenhorst, Lawrence P. McIntosh, and Barbara J. Graves ....................... 437
Contents ix

Signals and Combinatorial Functions of Histone Modifications
Tamaki Suganuma and Jerry L. Workman ............................................. 473

Assembly of Bacterial Ribosomes
Zahra Shajani, Michael T. Sykes, and James R. Williamson ......................... 501

The Mechanism of Peptidyl Transfer Catalysis by the Ribosome
Edward Ki Yun Leung, Nikolai Suslov, Nicole Tuttle, Raghuvir Sengupta,
and Joseph Anthony Piccirilli ................................................................. 527

Amyloid Structure: Conformational Diversity and Consequences
Brandon H. Toyama and Jonathan S. Weissman ...................................... 557

AAA+ Proteases: ATP-Fueled Machines of Protein Destruction
Robert T. Sauer and Tania A. Baker ......................................................... 587

The Structure of the Nuclear Pore Complex
André Hoelz, Erik W. Debler, and Günter Blobel .................................... 613

Benchmark Reaction Rates, the Stability of Biological Molecules
in Water, and the Evolution of Catalytic Power in Enzymes
Richard Wolfenden ................................................................................. 645

Biological Phosphoryl-Transfer Reactions: Understanding Mechanism
and Catalysis
Jonathan K. Lassila, Jesse G. Zalatan, and Daniel Herschlag ...................... 669

Enzymatic Transition States, Transition-State Analogs, Dynamics,
Thermodynamics, and Lifetimes
Vern L. Schramm ................................................................................... 703

Class I Ribonucleotide Reductases: Metallocofactor Assembly
and Repair In Vitro and In Vivo
Joseph A. Cotruvo Jr. and JoAnne Stubbe .............................................. 733

The Evolution of Protein Kinase Inhibitors from Antagonists
to Agonists of Cellular Signaling
Arvin C. Dar and Kevan M. Shokat ......................................................... 769

Glycan Microarrays for Decoding the Glycome
Cory D. Rillahan and James C. Paulson .................................................. 797

Cross Talk Between O-GlcNAcylation and Phosphorylation:
Roles in Signaling, Transcription, and Chronic Disease
Gerald W. Hart, Chad Slawson, Genaro Ramirez-Correa, and Olof Lagerlof ...... 825

Regulation of Phospholipid Synthesis in the Yeast
Saccharomyces cerevisiae
George M. Carman and Gil-Soo Han ....................................................... 859
Sterol Regulation of Metabolism, Homeostasis, and Development
Joshua Wollam and Adam Antebi ......................................................... 885

Structural Biology of the Toll-Like Receptor Family
Jin Young Kang and Jie-Oh Lee ............................................................. 917

Structure-Function Relationships of the G Domain, a Canonical
Switch Motif
Alfred Wittinghofer and Ingrid R. Vetter ................................................. 943

STIM Proteins and the Endoplasmic Reticulum-Plasma
Membrane Junctions
Silvia Carrasco and Tobias Meyer ......................................................... 973

Amino Acid Signaling in TOR Activation
Joungmok Kim and Kun-Liang Guan ......................................................... 1001

Mitochondrial tRNA Import and Its Consequences
for Mitochondrial Translation
André Schneider .................................................................................. 1033

Caspase Substrates and Cellular Remodeling
Emily D. Crawford and James A. Wells .................................................. 1055

Regulation of HSF1 Function in the Heat Stress Response:
Implications in Aging and Disease
Julius Anckar and Lea Sistonen ................................................................. 1089

Indexes
Cumulative Index of Contributing Authors, Volumes 76–80 ....................... 1117
Cumulative Index of Chapter Titles, Volumes 76–80 ................................. 1121

Errata
An online log of corrections to Annual Review of Biochemistry articles may be found at http://biochem.annualreviews.org/errata.shtml