Separation & Purification Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597294

A Birds-Eye View of Modern Proteomics

Abraham Lo[a]; Kevin A. Schug[a]
[a] Department of Chemistry and Biochemistry, The University of Texas at Arlington, Arlington, TX 76019-0065

To cite this Article Lo, Abraham and Schug, Kevin A.(2009) 'A Birds-Eye View of Modern Proteomics', Separation & Purification Reviews, 38: 2, 148 — 172
To link to this Article: DOI: 10.1080/15422110802555081
URL: http://dx.doi.org/10.1080/15422110802555081

Separation & Purification Reviews, 38: 148–172, 2009

ISSN 1542-2119 print/1542-2127 online
DOI: 10.1080/15422110802555081

Taylor & Francis
Taylor & Francis Group

A Birds-Eye View of Modern Proteomics

Abraham Lo and Kevin A. Schug*

Department of Chemistry and Biochemistry, The University of Texas at Arlington, Arlington, TX 76019-0065

Abstract: A broad general overview is provided on modern aspects of separation and detection in mass spectrometry-based proteomics research. This review is most suitable for undergraduate students and/or novices in the field.

Keywords: Proteomics, mass spectrometry, protein separation, protein identification

INTRODUCTION

The following review provides a broad perspective on the realm of proteomics and the relation of mass spectrometry to proteomics. Taking into account that the fields of proteomics and mass spectrometry are vast and extensive, this review does not attempt to provide a comprehensive or in-depth view of each field. This review does however present a general background/understanding for the novice or aspiring scientist with an interest to venture into such areas of research.

The first section describes what is meant by the term ''Proteomics,'' giving a brief historical account and placing the field of research into context with respect to its importance for understanding life's processes. This includes a brief introduction to proteins, as well as to the development of the field of proteomics in terms of its initiation and development through to modern times. Next, a general description of mass spectrometry is provided, which is expanded in the final section in

Received 30 May 2008, Accepted 10 October 2008
Address correspondence to Kevin A. Schug, Department of Chemistry and Biochemistry, The University of Texas at Arlington, Arlington, TX 76019-0065. E-mail: kschug@uta.edu

terms of its application to proteomics research. Focus is placed mainly on the use of high performance liquid chromatography – electrospray ionization – mass spectrometry in this regard.

PROTEOMICS

Proteomics is a relatively new discipline spawned within the field of biochemistry. In contrast to the early origins of biochemistry beginning in the 1800s, the genesis of proteomics can be traced back to the recent 1970s–1980s. The word proteomics was derived by Marc Wilkins, a Macquarie University Ph.D. candidate, in 1994 as ''the protein complement of the genome'' at a meeting in Siena, Italy (1).

The definition of proteomics is the study of the proteome where the standard definition of the proteome is ''the complete set of proteins within a certain organism''. Proteomics not only involves the study and the identification of proteins but also the interactions between proteins that compose various systems, so-called ''functional proteomics''. These systems may range from proteins involved in simple biological pathways to more complex systems such as organ systems or the whole organism itself. Along with this, the relative quantification of proteins in various biological systems underlies the blossoming field of biomarker discovery (2, 3). Monitoring up-regulation and down-regulation of various proteins and enzymes over periods of time or during experimental treatment provides the potential for personal clinical care and the (early) detection of various diseases associated with these biomarkers (4–6).

Within the short span of several decades, the field of proteomics has become a very hot topic with funding from both the private and public sector ballooning to billions of dollars a year. A significant portion of the NIH's yearly budget ($28.4 billion in 2007 (7)) is dedicated towards proteomic and genomic research. For example, 1 of 10 proteomics centers in the U.S. established by the National Heart, Lung, and Blood Institute (NHLBI) to enhance and develop innovative proteomics technologies was given a total budget of $157 million over seven years by the NIH (8). Furthermore, The National Institute of Allergy and Infectious Diseases (NIAID), part of the NIH, awarded a $8.7 million contract for the formation of the NIAID Resource Center for Biodefense Proteomic Research (9). These examples of judicious funding are just a snippet of the NIH's commitment for advancing proteomic research. The overwhelming consensus from the scientific community is that we have only begun to scratch the surface within the field of proteomics and that there are historic breakthroughs waiting to be discovered. With proteins being involved in almost all biological processes, the study of proteomics is directly related to questions and problems that plague the 21st century:

Can we cure diseases like cancer, Alzheimer's disease, Parkinson's disease, etc.? What effects do environmental toxins have on the human condition? Or even, what type of vaccines do we need to design to defend against the pressing threat of drug-resistant biological pathogens?

In general, proteomics is important for a number of reasons. It allows scientists to identify proteins in normal and disease conditions, to identify pathogenic mechanisms, to improve medicine by matching a person's protein fingerprint to the effective treatment, and to contribute to the understanding of gene function. The current goals of proteomics research are geared towards the determination of the properties of proteins which include sequence, quantity, state of modification, interactions with other proteins, activity, subcellular distribution, and structure (9).

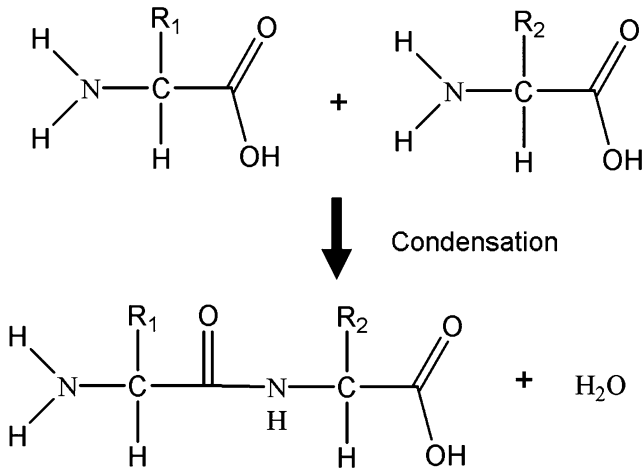
What are Proteins?

In proteomics, the protein is the common thread from which everything begins. Without proteins, life would not exist. All functions of living organisms are related to proteins. Proteins are the end result of the "central dogma" of life (10) where DNA is transcribed to RNA and RNA is translated to form proteins.

Proteins are amino acid based macromolecules that are vital in numerous biological processes. These biological processes are often complicated and convoluted, and scientists have only recently begun to unravel the intricate functions of proteins in biological systems. These functions range from enzymatic activity (i.e., pepsin breaking down food products in the stomach (11)) to providing mechanical function (actin and myosin contracting in the muscles (12)).

From an assortment of 20 amino acids available, proteins are constructed by the cell's machinery in a linear fashion by connecting the C-terminus of one amino acid to the N-terminus of another amino acid. Through a condensation reaction, as shown in Figure 1, an amide peptide bond forms between the two amino acids. A peptide is created after several amino acids join together to form a linear polymer. The general definition of a peptide is a molecule composed of two or more amino acids. Larger peptides are often referred to as polypeptides or proteins; however, a vital distinguishing factor between a peptide and protein is that a protein has a definitive 3-dimensional structure through the folding of the chain of amino acids.

The structure of a protein can be delineated as shown in Figure 2 into four distinct categories: primary, secondary, tertiary, and quaternary structure. The primary structure is the amino acid sequence of the protein. The secondary structures are localized formations stabilized by

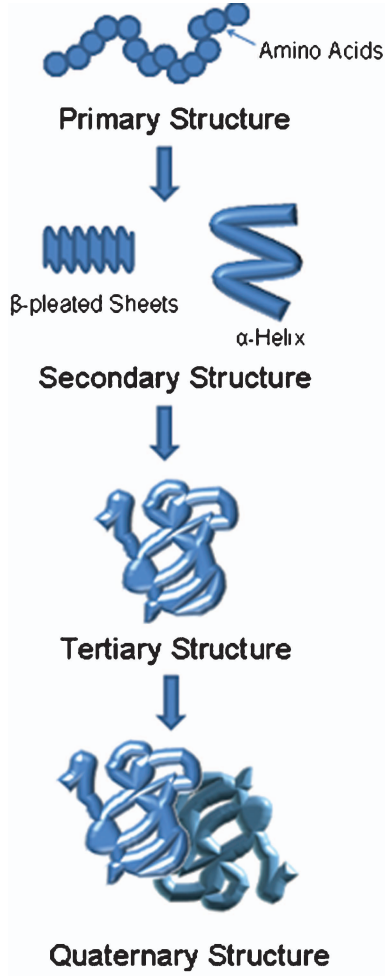


hydrogen bonds composing ordered arrangements such as alpha helixes and beta sheets. The tertiary structures are the gathering of secondary structures together through hydrogen bonding, disulfide bonds, and salt bridges to form the overall shape of the protein molecule. Finally, the quaternary structures are the aggregation of two or more protein molecules to form a larger multi-protein complex (i.e., hemoglobin).

Development of Proteomics

Four major scientific discoveries pertaining to proteins set the backdrop for the genesis of proteomics. The history of protein chemistry begins first when the word protein was coined by Jons Jakob Berelius in 1838 (13). Derived from Greek and meaning "of primary importance," the word protein was used for the large organic compounds with equivalent empirical formulas which he studied. In 1926, the next major step in protein research was made by James B. Sumner when he isolated and crystallized the enzyme urease (14). In 1955, Sir Frederick Sanger determined the entire amino acid sequence of the protein insulin (15, 16). In 1958, the 3-dimensional structures of the proteins, hemoglobin (17) and myoglobin (18) were elucidated through X-ray diffraction analysis by Max Perutz and Sir John Cowdery Kendrew and coworkers, respectively.

Building upon the momentous discoveries in protein chemistry, the subsequent scientific study of proteins has journeyed through several stages. Patterson and Aebersold breaks down the emergence and maturation of proteomic concepts and technology into three phases (10): (1) Transition from protein chemistry to proteomics as a platform for scientific advancement; (2) present diversification of proteomic



technologies and tools used to ascertain properties of proteins; and (3) comprehensive understanding of the working of biological systems through proteomic data and new science technologies.

In the infant stages of proteomics (phase 1), the main goals of protein chemistry conducted in the 1980s and early 1990s were to provide the link between the observed activity of a biochemically isolated protein and the gene that encoded it (10). These main goals were facilitated by the very important advent of two-dimensional gel electrophoresis (2DE), which allowed for the separation of complex protein matrixes based primarily on size and isoelectric point. With the separation of proteins, technological developments for analytical protein chemistry were commenced to improve the sensitivity of detection. Numerous achievements were

accomplished; however, the reliance on 2DE as the main means for protein research was a bottleneck for advancement. Several ideas were proposed in the 1970s and 1980s to use 2DE to create a database of proteins, but this endeavor failed initially. In the early 1990s, the sequencing of proteins through Edman degradation and PCR provided the link between amino acid sequence and the corresponding protein activity. Much effort was put into the mass DNA sequencing of cDNA derived from mRNA (19) with the well publicized Human Genome Project as its pinnacle. As a result, these endeavors created massive libraries of protein sequence databases.

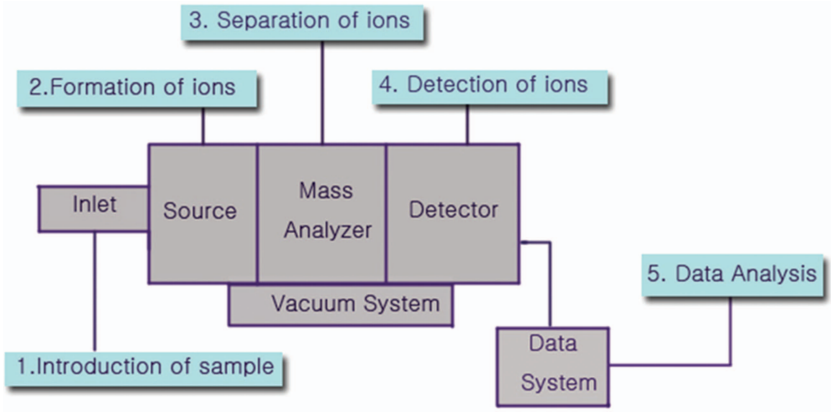
From the amalgamation of these protein sequence databases, creation of algorithms, and the availability of existing mass spectrometry techniques in concert with a wide variety of separation techniques, the study of proteins was viewed in a new light. While determining the mass of an unknown protein or peptide with a mass spectrometer is not a unique identifying feature, a protein cleaved with an enzyme of known specificity to create a mixture of peptides could distinctively identify a protein. By extracting peptide masses from the mass spectra after running a digested protein through reversed phase-high performance liquid chromatography-mass spectrometry (RP-HPLC-MS), one could match the mass list to a predetermined proteolytic sequence database of known proteins (20). Scientists now had an extremely powerful tool where they could identify and study most, if not all, proteins (proteome) on a much larger scale than before. This is when proteomics was born.

MASS SPECTROMETRY

Mass spectrometry (MS) is the analysis of the mass to charge ratio (m/z) of ions. It is a technique which facilitates the capability of 'weighing' charged molecules. In general, mass spectrometry is a very powerful technique because it allows for the analysis of gas, liquid, and solid samples. Through highly sensitive instrumentation and different analytical techniques, mass spectrometry is capable of elemental and molecular analysis. Mass spectrometry is now considered as an indispensable part of proteomics research.

MS Setup

The general MS instrumental setup consists of 5 major components: the inlet, the source, the mass analyzer, the detector, and the data acquisition system as shown in Figure 3. The inlet is where the introduction of sample occurs. Once the sample has been introduced, the formation of ions



(ionization) occurs in the source. The two common methods of ionization of proteins/peptides are through electrospray ionization (ESI) (21, 22) and matrix-assisted laser desorption/ionization (MALDI) (23). These ionization techniques are "soft" in nature, largely preserving ionized species as intact molecular ions. After ionization, the ions are transported to the analyzer through ion optics *via* the application of highly controlled electric and/or magnetic fields. The detection of ions and analysis are then accomplished by the detector and the data acquisition systems, respectively.

Inlet

The inlet is where the introduction of sample occurs. Since the creation of gas phase molecules is a requirement at the source, the inlet must somehow accommodate solution to gas phase transfer at the inlet/source interface. Although there are numerous types of inlets (direct vapor inlet, gas chromatography, liquid chromatography, direct insertion probe, etc.) which interface to a mass spectrometer, the inlet choice depends solely upon the sample. In order to choose which inlet is suitable, one must consider the characteristics of the sample. Samples with high volatility and thermal stability are easily accommodated by a variety of inlets involving thermal vaporization such as GC, direct vapor inlet, and direct insertion probes. Since proteins/peptides have low volatility and are thermally labile, they require direct ionization from the condensed phase (21). Liquid chromatography provides such an introduction of thermally labile and nonvolatile liquid compounds and mixtures not amenable to gas chromatography.

Liquid chromatography (LC) is a technique where a sample is forced through a column packed with solid particles (stationary phase) by a

liquid (mobile phase). The origins of liquid chromatography began in 1900 when Mikhail Tswett coined the word 'chromatography' in his seminal work on plant pigment separations (24, 25). Within his research, Tswett was able to separate a mixture of chlorophylls and xanthophylls by passing a solution in petroleum ether through a glass column packed with calcium carbonate (26). For several decades, similar methods employed packed columns as such for the separation of mixtures. Moving forward, the 1960s used packed columns with small particles resulting in high performance liquid chromatography or high pressure liquid chromatography (HPLC) where higher pressures were used to force liquid through the column at accelerated and highly controlled flow rates. HPLC is highly regarded as an analytical separation technique because of its sensitivity, its high efficiency, its ready adaptability to accurate quantitative determinations, its suitability for separating nonvolatile or thermally fragile species, and its widespread applicability to a variety of substances (27).

With the manipulation of the polarity of the stationary and mobile phases, a sample containing a mixture of analytes can be separated through the partitioning between the two phases. Multiple modes of separation (e.g. reversed phase, normal phase, ion exchange, size exclusion, chiral, etc.) are possible with HPLC. For proteins and peptides, the reversed phase mode of separation is the most widely employed. Reversed phase-HPLC (RP-HPLC) uses a non-polar stationary phase (i.e., C_{18} bonded silica) and an aqueous/polar organic mobile phase (i.e., acetonitrile or methanol). Through hydrophobic interactions, analytes are retained by the hydrophobic stationary phase in high aqueous mobile phase conditions. Then, they are typically eluted by gradually increasing the percentage of polar organic solvent in the mobile phase.

Gradient elution is commonly used to separate complex mixtures of peptides in RP-HPLC on a C_{18} column prior to mass spectrometric detection. A typical elution buffer for RP-HPLC consists of two reversed phase solvents labeled A and B containing 100% water with a small percentage of acidic modifier (formic acid (FA), trifluoroacetic acid, or acetic acid) and 100% Acetonitrile (ACN), respectively. The acidic modifier is used to facilitate good chromatographic peak shape and retention by suppressing ionization of acidic functional units on the peptides; but also to provide a source of protons for the ionization process. The ratio of solvents is typically varied in a programmed way with a low concentration of organic at the beginning of each run (i.e. 98:2 H_2O (0.5% FA):ACN) and increased to a high organic content toward the end of the run (i.e., 98:2 ACN: H_2O (0.5%FA)). Volitile buffers, such as ammonium acetate and ammonium formate (5–20 millimolar) are also commonly incorporated to facilitate electrospray ionization. The elution

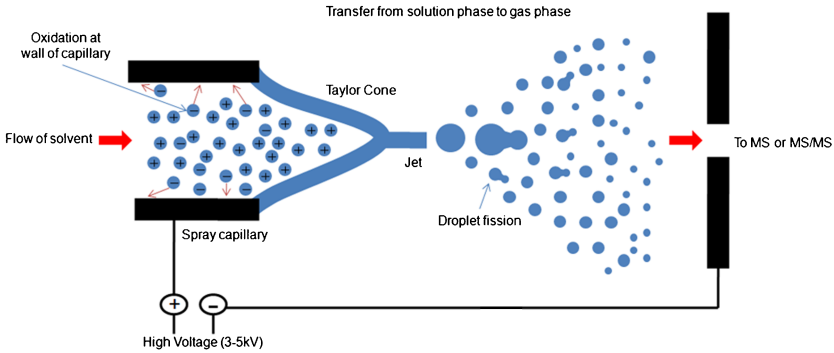
of many proteolytically derived peptides (e.g., via enzymatic digestion; see below) from a RP-HPLC C_{18} column commonly occurs around the 30–40% organic solvent mark in a typical linear gradient analysis. Peptides that are more hydrophobic are retained longer while those that are more hydrophilic are eluted earlier.

ESI Source

The source is where the formation of ions occurs. While interfacing LC to MS, two major hurdles of volatilizing biochemical analytes and removing solvent present in large excess, have to be overcome. In 1989, Fenn's group (22) pioneered a new atmospheric pressure ionization (API) source called electrospray ionization (ESI), which addressed the aforementioned issues, allowing for a seamless interface between LC and MS. It is now one of the most commonly used amongst other established interfaces (atmospheric pressure photoionization (APPI), matrix-assisted laser desorption/ionization (MALDI), Thermospray, fast atom bombardment (FAB), atmospheric pressure chemical ionization (APCI), etc.) for soft ionization of biomolecules. Most modern LC-MS instruments can readily utilize at least two atmospheric pressure ionization modes: electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI).

ESI has the advantages of ionizing low and high molecular weight non-volatile or volatile polar and ionic compounds. It can accommodate conventional HPLC flow rates ($\leqslant$ 1000 ul/min with the aid of N_2 pneumatic nebulization) and exhibits high sensitivity due to its relatively high ionization efficiency. As the most commonly employed ion source for LC-MS, ESI is geared towards polar, ionic molecules (i.e., proteins and peptides) and has the advantage of detecting very large biomolecules through multiple charging (28). For all its virtues, ESI still has several disadvantages, including little to no fragmentation with possible adduct formation, no ionization for non-polar compounds, and no universal library.

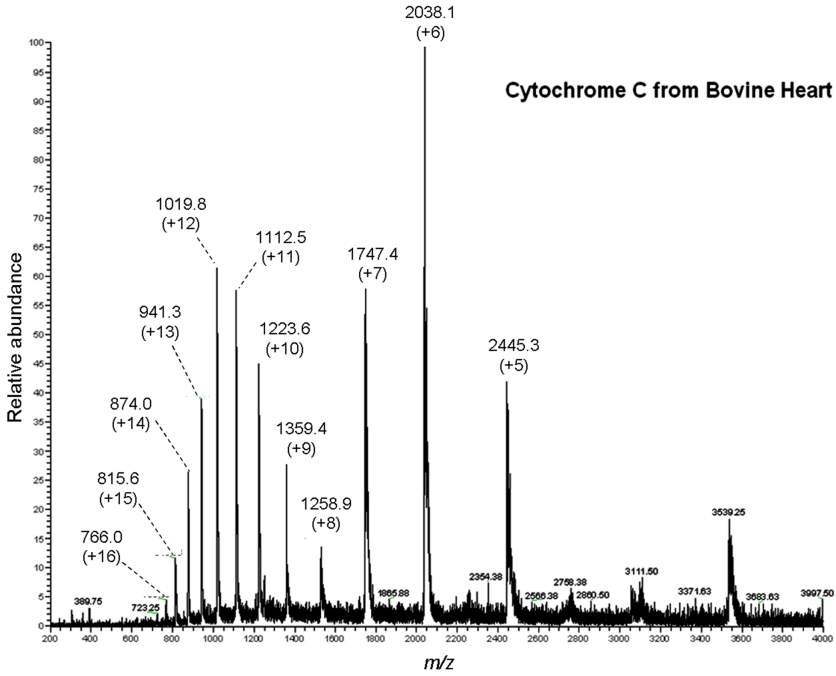
The exact science behind the mechanism of ESI is not fully understood at this point, but two prevailing models have been proposed: the ion evaporation model (29) and the charged residue model (30, 31). Droplet size, surface charge, liquid surface tension, solvent volatility, and analyte ion solvation energies are several factors which must be considered to understand the ESI process. The general mechanism for ESI ion generation, as shown in Figure 4, can be broken down into four steps. First, the capillary through which the sample solution is passed is held at a high potential ($\pm$2–5 kV). Second, in response to the potential gradient placed across the atmospheric pressure spray chamber, the liquid solvent disperses into a mist of highly charged droplets. Third, the



droplets reduce in size by evaporation of the solvent or by "Coulombic explosion." Coulombic explosion refers to the phenomena where the surface tension holding the droplet together is overcome by charge repulsion at the surface of the shrinking droplet. Last, the fully desolvated ions resulting from complete evaporation of the solvent are transferred to the mass spectrometer for mass analysis. ESI can be operated in either the positive or the negative ionization mode by varying the potential applied to the capillary. For proteins and peptides, positive mode is ideal. Analysis in negative mode may be hampered by arcing, corona discharge, and unstable spray current; for these reasons, sensitivity in the negative ionization mode is often less than that found in the positive ionization mode.

The high sensitivity of ESI, without excessive fragmentation, proves worthy in its application towards protein/peptide research. In contrast to other ionization techniques, where upon ionization significant fragments (offspring ions) may occur, ESI utilizes a 'soft ionization' technique which preserves the molecular ion (precursor ion) due to the low internal energy imparted onto the molecule. The softness of the ESI technique has even been used extensively for preserving and studying non-covalent interactions between protein/ligand and peptide/ligand complexes (32–45).

As mentioned above, ESI has the potential to produce multiply charged ions allowing higher molecular weight compounds to be observed at lower mass to charge ratio (m/z) values using standard mass analyzers. For example, one might predict that a protein such as Cytochrome c (MW ~11.3 kDa) would be difficult to observe on a typical ion trap mass spectrometer with an effective scan range of 100–4000 m/z. However under optimum solvent conditions, Cytochrome c appears as an envelope of low m/z multiply charged ions as shown in Figure 5. By applying an averaging algorithm and examining the relative position of isotopic ion abundances, one is able to assign number of



charges to the peaks in the ion envelope and estimate the average molecular mass of the protein.

Nanoelectrospray Ionization (nESI) Source

The reduction in source dimensions to accommodate lower flow rates, higher sensitivity, and increased linearity has been achieved and is commonly employed through the use of nanoelectrospray ionization sources (46). While the mechanism of ionization is conceptually identical to that in conventional ESI, the use of microbore spray capillaries (low μm internal diameter) allows for higher ionization efficiency by virtue of the production of smaller charged droplets. At reduced flow rates (nL/min), the spray capillary can be placed directly in front of the mass spectrometer inlet to allow more ions to enter the mass analyzer, improving sensitivity. While such sources have been traditionally prone to clogging, some manufacturers (e.g., Advion and Agilent) provide options in a microchip format that alleviate this problem through the robotic manipulation and placement of a new nanospray nozzle for each analysis.

MALDI Source

Matrix-assisted laser desorption/ionization (MALDI) was pioneered in the 1980s as a new tool for the analysis of large biomolecules (23, 47).

Analyte samples are co-crystallized with an appropriate matrix onto a target plate and irradiated by a laser to desorb ions into the gas phase for mass analysis. The matrix has two functions. These small molecule organic molecules possess a chromophore and absorb the bulk of the laser radiation, protecting the analyte, which is typically present in an order of magnitude lower amount in the sample spot, allowing the mixture to be desorbed into the gas phase. The matrix is also believed to be responsible for transferring charge to the analyte molecule through ion–molecule collisions in the gas phase, just following desorption. This soft ionization technique is amenable to a wide range of analytes and suitable matrices can be chosen which maximize the ionization of samples ranging from peptide mixtures (proteolytic digests) and small molecules to polymers and intact proteins. The pulsed laser radiation makes this source most amenable to combinations with time-of-flight mass analyzers, which in turn, place little limit on the size of potential analytes to be investigated, but other mass analyzer configurations are also available to combine with MALDI. Recently, a variant of MALDI, termed surface enhanced laser desorption/ionization (SELDI), has also found wide-spread use in proteomics research. In SELDI, proteins are selectively adsorbed to the surface of a chemically-modified target plate, allowing the user to remove unwanted impurities by washing prior to analysis (6).

Mass Analyzer and Detector

Many types of mass analyzers are available on the market today including time-of-flight (TOF), ion trap (IT) (linear, quadrupole, orbitrap), triple quadrupole, magnetic sector, and Fourier transform ion cyclotron resonance (FT-ICR). The choice of mass analyzer is dependent on the analytical goals requiring a certain sensitivity, resolution, mass accuracy, mass range, rate of scanning, ease of use, and cost. Hybrid mass analyzers such as the hybrid quadrupole time of flight (Q-TOF), linear ion trap-Fourier transform ion cyclotron resonance (LIT-FTICR), ion trap time-of-flight (IT-TOF), triple quadrupole (QQQ), and TOF-TOF are commonly employed for proteomics but cost can be a limiting factor. These hybrid analyzers exploit the advantages of various analyzers while minimizing the disadvantages by combining several analyzers into one instrument.

Currently, hybrid mass analyzers such as Q-TOF, IT-TOF, and TOF-TOF, along with FT-ICR, are the principal mass analyzers used in proteomic research. These mass analyzers are exploited for their high resolution, mass accuracy, and sensitivity. Quadrupole ion traps are also used in proteomic research due to their advantages of being

inexpensive, easily interfaced to many ionization methods, and having MS/MS capabilities. Although the resolution(<4000), mass accuracy($\sim \pm 0.5$ amu), mass range(<4000 m/z), and scanning rate might be inferior in comparison to more expensive mass analyzers, the quadrupole ion trap still remains a reliable workhorse in proteomic research.

The MS^n (tandem MS to the nth degree) capabilities are especially useful in the hybrid and ion trap mass analyzers. Hybrid mass analyzers such as the Q-TOF utilize the first quadrupole to select the precursor ion and then fragment the selected ion with the second quadrupole through Collision Induced Dissociation (CID) for MS^2 analysis. The time-of-flight portion of the mass analyzer then analyzes the selected product ions. The ion trap operates under slightly different conditions. By isolating a particular ion, ejecting everything else, and then employing CID, the relevant fragmentation data of the precursor ion is obtained. The process can be repeated by isolating one of the fragment ions and so on (MS^n), as long as sufficient ion signal remains. For detection, the ions are focused onto the detector to produce the mass spectrum. Most detectors are comprised of a conversion dynode and a channel electron multiplier.

Current mass spectrometers often have the capability to provide full scan followed along with data-dependent MS/MS or MS^n information to obtain data for interacting with proteomics databases. Through specified parameters, data dependent scanning allows the selection of one or more ions of significance for analysis. The software program picks the most intense ion in the full-scan spectrum and extracts MS/MS data from that particular ion. The program then stores the mass into an exclusion list after acquisition and seeks out less intense components that have not yet been examined. This automated process ensures that both high and low abundance proteolytic peptide ions are fragmented to maximize chances of identifying the protein(s) of interest.

Data Acquisition System

The data acquisition aspect of mass spectrometry has gone through radical changes in the last 30 years. As strip chart recorders sit around gathering dust as relics of the past, the digital age has brought the necessary advancements for progressive proteomic research. Computer hardware and software play a vital role in the management, categorization, and analysis of data relayed from the mass spectrometer. The internet also plays an important role in the dissemination of information and rapid exchange of ideas, including interfacing collected data with numerous databanks for proteomic-based applications.

Relation between Proteomics and Mass Spectrometry

As mentioned previously, Patterson and Aebersold have broken down the emergence and maturation of proteomic concepts and technology into three phases (10). Phase 1 highlights the transition from protein chemistry to proteomics as a platform for scientific advancement. Phase 2 ascertains the properties of proteins through the diversification of proteomic technologies and platforms. Last, phase 3 builds towards a comprehensive understanding of the working of biological systems through proteomic data and new science technologies. The following discussion centers largely on methods employed relative to Phase 1. Proteomics is carried out by two predominant experimental setups, the top-down and the bottom-up approach. These classifications are defined first, followed by a description of the methods needed to carry them out. Experimental methods are highly dictated by the type of information desired (e.g. sequencing, identification, molecular weight determination, quantification, etc.) and often rely heavily on an ever-increasing set of bioinformatics tools available to the researcher.

Top-Down Proteomics

The top-down proteomic approach focuses on the protein level analysis where intact protein ions or large protein fragments are subjected to gas phase fragmentation. The fragmentation pattern from an intact protein is often difficult to interpret since, in ESI, different charge states have different fragmentation patterns (48–51). As a result, most top-down approaches for determining sequential information rely on high resolution mass analyzers (FT-ICR) to help deconvolute data. Top-down proteomics is often considered not as information rich as the bottom-up approach and is still a limited technique. Some of its uses involve the direct determination of molecular mass which can be useful in determining post-translational modifications where sufficient mass resolution and sensitivity are available. Either ESI-MS or MALDI-TOF-MS of intact proteins may be performed if molecular weight information is desired. A protein's primary structure is also a target of the top-down proteomic approach. *De novo* sequencing is a technique where the sequence of a previously unknown protein is derived through the use of the tandem mass spectrometry data. The recent interest in ion mobility mass spectrometry (52) also provides an interesting route to studying protein 3° and 4° structure (e.g., protein folding dynamics).

Bottom-up Proteomics

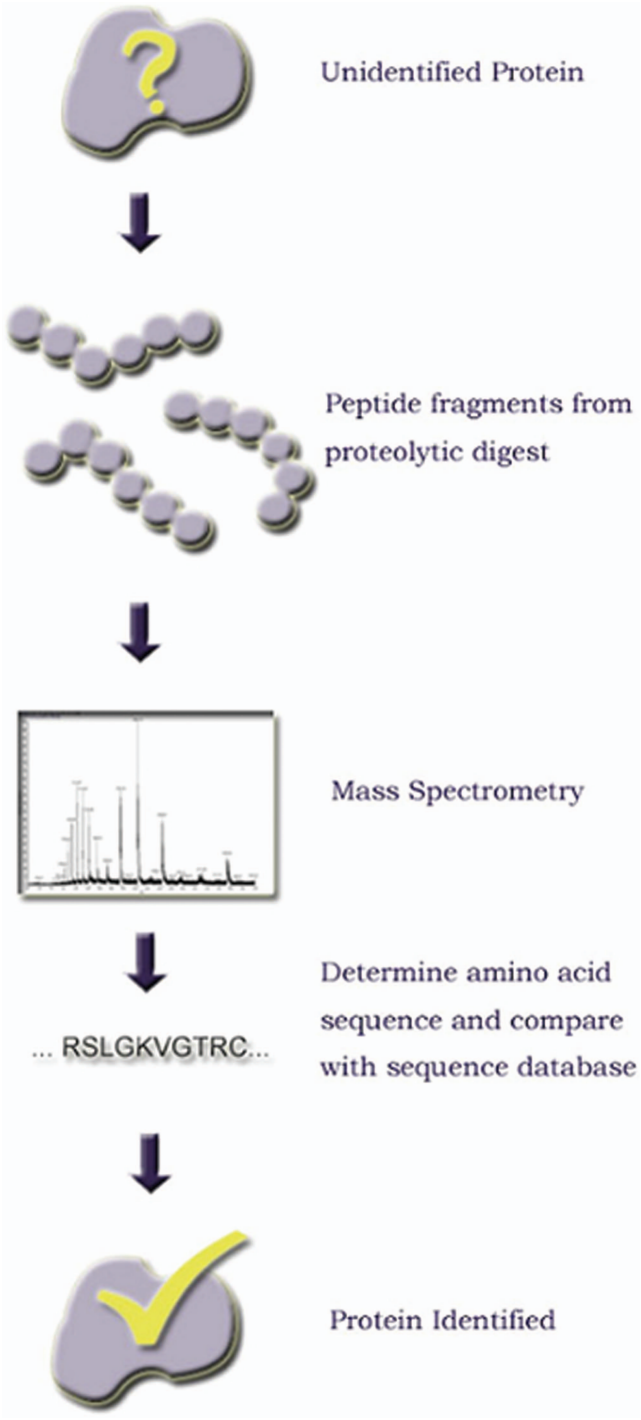
In bottom-up proteomics, an isolated and proteolytically digested protein of interest yields peptide fragments. Specific cleavage is often accomplished by the protease enzyme trypsin, but other proteases such as chymotrypsin, Glu-C, Asp-N, and others are also utilized. More on specific proteases is described below. MALDI is commonly utilized for peptide mass fingerprinting of the complex mixture (53). The masses of peptides are extracted from the mass spectra and matched against a database. Figure 6 provides a general schematic of the bottom-up proteomic approach. Just like a human fingerprint, each protein has a distinct pattern of peptide signals for a given protein. Another technique called peptide fragmentation incorporates ESI to take the sequential information based on collisional dissociation of polypeptides from a protein and matches them against a database (54–57). Rather than providing the exact information on protein sequence, bottom-up approaches are mainly used for protein identification. Various types of collisional dissociation processes can be employed for experiments ranging from protein identification to the localization of post-translational modifications (58, 59).

Mapping of Primary Structure and Bioinformatics

Protease mapping is a way to determine the primary structure sequence of a protein. It has gone through an evolutionary change from its early roots in chromatography/gel electrophoresis combined with Edman degradation (60) sequencing to a more streamlined technique where the combination of mass spectrometry, proteolytic techniques, and computer-driven data analysis software (Mascot, PEAKS, OMSSA, SEQUEST, Protein-Prospector, etc.) are used to determine the primary structure of a protein. By conducting mass analysis on a peptide mixture resulting from a protein proteolyzed by a sequence specific enzyme, the fragmentation pattern can be statistically compared to databases (NCBInr, Genpept, Swiss Prot, Owl, Ludwignr, Unknome, etc.) populated by already sequenced proteins. For example, NCBI's RefSeq database contains over 5 million proteins. With an appropriately selected protease that generates a significant number of peptide fragments, there is a reasonable probability of identifying the target protein if it is contained in the database.

Denaturation

Proteins fold into compact structures stabilized through their secondary and tertiary structures. In order to effectively map the primary structure



of a protein, both the hydrophilic surface and hydrophobic core of the protein must be made available to the enzymes for proteolysis. Denaturation of the protein causes the secondary and tertiary structures to destabilize thereby allowing the enzymes to access previously inaccessible sites. Heat, acids, bases, heavy metal salts, detergents, and organic solvents are sometimes employed to achieve denaturing conditions (61). Urea and guanidinium chloride are commonly used chemicals for the denaturation process (62). These chaotropic reagents disrupt intramolecular hydrogen bonds which would otherwise hold the protein in a folded conformation. Tertiary elements such as disulfide bonds connecting cysteines to one another are main points of emphasis when attempting to denature a protein. Dithiothreitol can then bez a commonly used reagent to reduce disulfide bonds. Iodoacetamide can then be used to cap the cysteines through an alkylation reaction following treatment with dithiothreitol (63).

Proteases

A protease is an enzyme that causes the degradation of a protein. Proteases are exploited in proteomics for their specificity at cleaving certain amino acids within the primary sequence of a protein. Trypsin, Lys C, Lys N, CNBr, Arg C, Asp N, Chymotrypsin, Pepsin, Proteinase K are several site-specific proteases utilized in biochemistry. Although proteases are commonly employed in a standalone fashion, they are occasionally used in tandem for increased proteolysis. The most common and robust of the enzymes is trypsin. Trypsin primarily cleaves at the C-terminal side of the amino acids lysine and arginine, except when either is followed by proline. Another common enzyme, chymotrypsin, cleaves peptides at the C-terminal side of tyrosine, tryptophan, and phenylalanine. Also commonly employed for digestion at an acidic pH, pepsin cleaves at the C-terminal side of phenylalanine and leucine. Table 1 provides a summary of different specificities for enzyme-catalyzed proteolysis.

Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) is a commonly employed method for the mapping of primary structure. MS/MS techniques enable fragmentation of a peptide or protein, providing an in-depth analysis of the amino acid sequence and possible post-translational modifications, for either top-down or bottom-up applications. The fragmentation can be accomplished by two methods: in-source fragmentation and post-source fragmentation. The

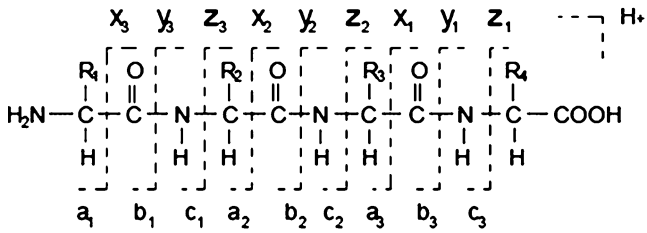
Table 1. Protease enzyme specificities

Enzyme or Reagent	Cleaves where?	Exceptions
Trypsin	C-terminal side of K or R	if P is C-term to K or R
Lys C	C-terminal side of K	
Lys N	N-terminal side of K	
CNBr	C-terminal side of M	
Arg C	C-terminal side of R	if P is C-term to R
Asp N	N-terminal side of D	
Asp N / Lys C	N-terminal side of D, C-terminal side of K	
Asp N + N-terminal Glu	N-terminal side of D or E	
Glu C (bicarbonate)	C-terminal side of E	if P is C-term to E, or if E is C-term to E
Glu C (phosphate)	C-terminal side of D or E	if P is C-term to D or E, or if E is C-term to D or E
Chymotrypsin	C-terminal side of F, L, M, W, Y	if P is C-term to F, L, M, W, Y, if P is N-term to Y
Pepsin (pH 1.3)	C-terminal side of F, L	
Pepsin (pH > 2)	C-terminal side of F, L, W, Y, A, E, Q	
Proteinase K	C-terminal side of A, F, Y, W, L, I, V	

most common post-source fragmentation uses collision-induced dissociation (CID) in a tandem mass analyzer. The idea behind MS/MS is to select a precursor ion, focus the ion in a collision cell, collide the ion with an inert gas, and then mass analyze the resulting product ions. The dissociation of a peptide usually requires an energy of 10–100 eV. This low energy process tends to form product ions formed by small neutral losses and cleavage of peptide bonds (61). Higher energy processes tend to dissociate backbone and side-chain bonds of peptides providing more information but also convoluting the data (64, 65).

Peptide Fragmentation

The resulting fragmentation ions from MS/MS of a peptide by collisional dissociation are called b-ions and y-ions. Electron capture dissociation results primarily in c- and z-ions. a, b, or c-ions are the ions with the charge retained on the N-terminus while x, y, z-ions have the charge retained on the C-terminus. Figure 7 depicts peptide fragmentation nomenclature. Roepstorff and Fohlman (66) initially formed the nomenclature for product ion formation and Johnson (65)



and Biemann (67, 68) subsequently added modifications to the nomenclature. The general mechanism of peptide fragmentation has been studied extensively (69–74) and the process is well understood, in general. However, researchers constantly seek innovative ways to generate information-rich fragmentation data through new dissociation techniques.

Mapping of Tertiary Structure

The determinations of the interface region of the protein in contact with the ligand (associated protein, enzyme, drug compound, etc.) and the functionally significant amino acid residues are of interest in understanding biochemical function. Previously, x-ray crystallography and NMR have been the common methods for probing protein structure at the molecular level. Recently, mass spectrometry integrated with a combination of hydrogen-deuterium exchange, chemical modification, and limited proteolysis has emerged as an effective tool for identifying interaction regions.

H/D Exchange

Hydrogen-deuterium (H/D) exchange combined with mass spectrometry can be used to determine interaction regions of a protein (75). Hydrogen atoms are exchanged with deuterium atoms when the compound is dissolved in deuterium oxide. Those hydrogen atoms which are accessible to the solvent are rapidly exchanged with the deuterium while those located within the hydrophobic core of the protein are left untouched. For complexes, the ligand provides protection from H/D exchange at the interaction region. Protein in the presence and absence of ligand are deuterated, quenched by acid, enzymatically proteolyzed by pepsin, and then mass analyzed. By comparing the masses of peptic fragments, the interaction site can be determined.

Chemical Modification

Chemical modification is also a method for studying tertiary structure, but is also useful for locating specific residues during primary sequence mapping. Two chemical modification techniques include reactions towards specific functional groups on the protein and mutations during protein synthesis. The most common form of chemical modification is the addition of a crosslinker to covalently link subunits of a protein complex. After cross-linking, factors such as the tertiary structure of the protein, stoichiometry of the noncovalent protein complex (76), and the protein complex interface region (77–79) can be determined.

Limited Proteolysis

In limited proteolysis, the protein is subjected to a protease in non-denaturing conditions. Since certain polypeptide chain regions of the proteins are buried within the core, they are protected from proteolysis. Only the exposed surface of the protein and the flexible sites are accessible to the protease. Like the H/D exchange example above, the unassociated protein is compared to the associated protein by mass analysis and the interaction site is determined. One of the most cited examples of limited proteolysis in relation to non-covalent protein complex is the transcription factor Max and Max-specific DNA complex (80). Cohen et al. were able to effectively probe the solution structure of MAX even though NMR or X-ray crystallography was unable to provide insight at the time.

CONCLUSION

Although still maturing, proteomics and mass spectrometry have reached new heights as well-established platforms for scientific discovery. The importance of proteomics and mass spectrometry is highlighted by the vast amount of published and ongoing research in each area. Currently, proteomics along with mass spectrometry are crucial for advancing biochemical work centered on elucidating protein function. Indications of positive outlook for the future include the development of new techniques such as ion mobility mass spectrometry (52), the advancement of quantitative mass spectrometric approaches, innovation in bioinformatics, and achieving greater understanding of systems biology.

REFERENCES

1. Wilkins, M.R., Sanchez, J.C., and Gooley, A.A. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.*, 13: 19–50.
2. Bachi, A. and Bonaldi, T. (2008) Quantitative proteomics as a new piece of the systems biology puzzle. *J. Proteom.*, 71: 357–367.
3. Veenstra, T.D., Conrads, T.P., Hood, B.L., Avellino, A.M., Ellenbogen, R.G., and Morrison, R.S. (2005) Biomarkers: Mining the biofluid proteome. *Molec. Cell. Proteom.*, 4: 409–418.
4. Jurisicova, A., Jurisica, I., and Kislinger, T. (2008) Advances in ovarian cancer proteomics: the quest for biomarkers and improved therapeutic interventions. *Expert Rev. Proteom.*, 5: 551–560.
5. Dihazi, H. and Müller, G.A. (2007) Urinary proteomics: a tool to discover biomarkers of kidney diseases. *Expert. Rev. Proteom.*, 4: 39–50.
6. Li, J., Zhang, Z., Rosenzweig, J., Wang, Y.Y., and Chan, D.W. (2002) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin. Chem.*, 48: 1296–1304.
7. Zerhouni, E.A. (2008) National Institutes of Health. FY 2007 Director's Budget Request Statement. http://www.nih.gov/about/director/budgetrequest/fy2007directorsbudgetrequest.htm (accessed April 3, 2008).
8. Sali, A. (2003) NIH workshop on structural proteomics of biological complexes. *Structure*, 11: 1043–1047.
9. Resource Center for Biodefense Proteomics Research. (2008) http://www.proteomicsresource.org/AboutProject.aspx (Accessed April 3, 2008).
10. Patterson, S.D. and Aebersold, R.H. (2003) Proteomics: the first decade and beyond. *Nat. Genet.*, 33: 311–323.
11. Crick, F. (1970) Central dogma of molecular biology. *Nature*, 227: 561–563.
12. Florkin, M. (1957) Discovery of pepsin by Theodor Schwann. *Rev. Med. Liege*, 12: 139–144.
13. Adams, R.J. and Pollard, T.D. (1986) Propulsion of organelles isolated from Acanthamoeba along actin filaments by myosin-1. *Nature*, 322: 754–756.
14. Wisniak, J. (2000) Jöns Jacob Berzelius: A guide to the perplexed chemist. *Chem. Educ.*, 5: 343–350.
15. Dounce, A.L. (1955) Prof. James B. Sumner. *Redox Rept.*, 176: 859.
16. Sanger, F. (1960) Chemistry of insulin. *Br. Med. Bull.*, 16: 183–188.
17. Muirhead, H. and Perutz, M. (1963) Structure of hemoglobin. A three-dimensional fourier synthesis of reduced human hemoglobin at 5.5 A resolution. *Nature*, 199: 633–638.
18. Kendrew, J., Bodo, G., Dintzis, H., Parrish, R., Wyckoff, H., and Phillips, D. (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 181: 662–666.
19. Adams, M., Kelley, J., and Gocayne, J. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252: 1651–1656.
20. Chait, B.T. (2006) Chemistry. Mass spectrometry: bottom-up or top-down. *Science*, 314: 65–66.

21. Van Bramer, S.E. (1997) *An Introduction to Mass Spectrometry*. Widener University, Chester, PA.
22. Fenn, J.B., Mann, M., Meng, C.K., and Wong, S.F. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246: 64–71.
23. Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons. *Anal. Chem.*, 60: 2299–2301.
24. Leicester, H.M. (1968) *Source Book in Chemistry, 1900–1950*; Harvard University Press: Cambridge, MA, 316–223.
25. Teich, M. (1992) *A Documentary History of Biochemistry, 1770–1940*; Fairleigh Dickinson University Press: Rutherford, NJ.
26. Zechmeister, L. (1946) Mikhail Tswett—The inventor of chromatography. *Isis*, 36: 108–109.
27. Skoog, D.A., Holler, F.J., and Nieman, T.A. (1998) *Principles of Instrumental Analysis*. 5th; Brooks/Cole Thomson Learning: California, 726.
28. Yamashita, M. and Fenn, J.B. (1984) Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.*, 88: 4451–4459.
29. Tal'roze, G.V. (1968) Molecular beam of macroions. *J. Phys. Chem.*, 42: 2240–2249.
30. Iribarne, J.V. and Thomson, B.A. (1976) On evaporation of small ions from charged droplets. *J. Chem. Phys.*, 64: 2287–2294.
31. Thomson, B.A. and Iribarne, J. V. (1979) Field induced ion evaporation from liquid surfaces at atmospheric pressure. *J. Chem. Phys.*, 71: 4451–4463.
32. Wendt, S., Mccombie, G., Daniel, J., Kienjofer, A., Hilvert, D., and Zenobi, R. (2003) Quantitative evaluation of noncovalent chorismate mutase-inhibitor binding by ESI-MS. *J. Am. Soc. Mass Spectrom.*, 14: 1470–1476.
33. Loo, J.A., Holler, T.P., Foltin, S.K., McConnell, P., Banotai, C.A., Horne, N.M., Mueller, W.T., Stevenson, T.I., and Mack, D.P. (1998) Application of electrospray ionization mass spectrometry for studying human immunodeficiency virus protein complexes. *Proteins*, 2: 28–37.
34. Kraunsoe, J.A., Aplin, R.T., Green, B., and Lowe, G. (1996) An investigation of the binding of protein proteinase inhibitors to trypsin by electrospray ionization mass spectrometry. *FEBS Lett.*, 396: 108–112.
35. Daniel, J.M., McCombie, G., Wendt, S., and Zenobi, R. (2003) Mass spectrometric determination of association constants of adenylate kinase with two noncovalent inhibitors. *J. Am. Soc. Mass Spectrom.*, 14: 442–448.
36. White, H.D. and Ashcroft, A.E. (2007) Real-time measurement of myosin-nucleotide noncovalent complexes by electrospray ionization mass spectrometry. (2007) *Biophys. J.*, 3: 914–919.
37. Conway, M.C., Whittal, R.M., Baldwin, M.A., Burlingame, A.L., and Balhorn, R. (2006) Electrospray mass spectrometry of NeuAc oligomers associated with the C fragment of the tetanus toxin. *J. Am. Soc. Mass Spectrom.*, 17: 967–976.
38. Fermas, S., Gonnet, F., Varenne, A., Gareil, P., and Daniel, R. (2007) Frontal analysis capillary electrophoresis hyphenated to electrospray ionization mass spectrometry for the characterization of the antithrombin/heparin pentasaccharide complex. *Anal. Chem.*, 79: 4987–4993.

39. McClements, D.J. (2006) Non-covalent interactions between proteins and polysaccharides. *Biotechnol. Adv.*, 24: 621–625.
40. Wang, W., Kitova, E.N., and Klassen J.S. (2003) Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanelectrospray FTICR-MS. *Anal. Chem.*, 75: 4945–4955.
41. Krishnaswamy, S.R., Williams, E.R., and Kirsch, J.F. (2006) Free energies of protein–protein association determined by electrospray ionization mass spectrometry correlate accurately with values obtained by solution methods. *Protein Sci.*, 15: 1465–1475.
42. Dennhart, N. and Letzel, T. (2006) Mass spectrometric real-time monitoring of enzymatic glycosidic hydrolysis, enzymatic inhibition and enzyme complexes. *Anal. Bioanal. Chem.*, 386: 689–698.
43. Jørgensen, T.J.D., Roepstorff, P., and Heck, A. (1998) Direct determination of solution binding constants for noncovalent complexes between bacterial cell wall peptide analogues and vancomycin group antibiotics by electrospray ionization mass spectrometry. *Anal. Chem.*, 70: 4427–4432.
44. Kempen, E.C. and Brodbelt, J.S. (2002) A method for the determination of binding constants by electrospray ionization mass spectrometry. *Anal. Chem.*, 72: 5411–5416.
45. Powell, K.D., Ghaemmaghami, S., Wang, M.Z., Ma, L., Oas, T.G., and Fitzgerald, M.C. (2002) A general mass spectrometry-based assay for the quantitation of protein-ligand binding interactions in solution. *J. Am. Chem. Soc.*, 124: 10256–10257.
46. Wilm, M. and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal. Chem.*, 68: 1–8.
47. Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988) Protein and polymer analysis up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*, 2: 151–153.
48. Hogan, J.M. and McLuckey, S.A. (2003) Charge state dependent collision-induced dissociation of native and reduced porcine elastase. *J. Mass Spectrom.*, 38: 245–256.
49. Reid, G.E., Wu, J., Chrisman, P.A., Wells, J.M., and McLuckey, S.A. (2001) Charge-state-dependent sequence analysis of protonated ubiquitin ions via ion trap tandem mass spectrometry. *Anal. Chem.*, 73: 3274–3281.
50. Newton, K.A., Chrisman, P.A., Reid, G.E., Wells, J.M., and McLuckey, S.A. (2001) Gaseous apomyoglobin ion dissociation in a quadrupole ion trap: $(M + 2H)^{2+}$ - $(M + 21H)^{21+}$. *Int. J. Mass Spectrom.*, 212: 359–376.
51. Wells, J.M., Stephenson, J.L., and McLuckey, S.A. (2000) Charge dependence of protonated insulin decompositions. *Int. J. Mass Spectrom.*, 203: A1–A9.
52. Kanu, A.B., Dwivedi, P., Tam, M., Matz, L., and Hill, H.H. (2008) Ion mobility-mass spectrometry. *J. Mass Spectrom.*, 43: 1–22.
53. Bienvenut, W.V., Deon, C., and Pasquarello, C. (2002) Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics*, 2: 868–876.

54. Wood, T.D., Chen, L.H., White, C.B., Babbitt, P.C., Kenyon, G.L., and McLafferty, F.W. (1995) Sequence verification of human creatine kinase (43 kDa) isoenzymes by high-resolution tandem mass spectrometry. *Proc. Natl. Acad. Sci. USA*, 92: 11451–11455.
55. Amunugama, R., Hogan J.M., Newton K.A., and McLuckey S.A. (2004) Whole protein dissociation in a quadrupole ion trap: identification of an a priori unknown modified protein. *Anal. Chem.*, 76: 720–727.
56. Kelleher, N.L., Lin H.Y., Valaskovic G.A., Aaserud D.J., Fridriksson E.K., and McLafferty F.W. (1999) Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.*, 121: 806–812.
57. Reid, G.E. and McLuckey S.A. (2002) 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.*, 37: 663–675.
58. Zubarev, R. (2006) Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry. *Expert Rev. Proteom.*, 3: 251–261.
59. Kjeldsen, F., Haselmann K.F., Budnik B.A., Sorensen E.S., and Zubarev R.A. (2003) Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. *Anal. Chem.*, 75: 2355–2361.
60. Edman, P. (1950) Method for the determination of the amino acid sequence in peptides. *Acta Chem. Scand.*, 4: 283–293.
61. Mangino, M. and Harper, J. (2008) Food Science 822. Protein Denaturation. http://class.fst.ohio-state.edu/FST822/lectures/Denat.htm (accessed April 4, 2008).
62. Monera, O.D., Kay, C.M., and Hodges, R.S. (1994) Protein denaturation with guandinium hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Prot. Sci.*, 3: 1984–1991.
63. Yen, T.Y., Yan H., and Macher B.A. (2002) Characterizing closely spaced, complex disulfide bond patterns in peptides and proteins by liquid chromatography/electrospray ionization tandem mass spectrometry. *J. Mass Spectrom.*, 37: 15–30.
64. Johnson, R.S., Martin, S.A., and Biemann, K. (1988) Collision-induced fragmentation of $(M+H)^+$ ions of peptides: side chain specific sequence ions. *Int. J. Mass Spectrom. Ion Proc.*, 86: 137–154.
65. Johnson, R.S., Martin, S.A., Biemann, K., Stultz, J.T., and Watson, J.T. (1987) Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.*, 59: 2621–2625.
66. Roepstorff, P. and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.*, 11: 601.
67. Biemann, K. (1988) Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.*, 16: 99–111.
68. Biemann, K. (1990) Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.*, 193: 886–887.

69. Yalcin, T., Csizmadia, I.G., Peterson, M.R., and Harrison, A.G. (1996) The structure and fragmentation of B_n (n>3) ions in peptide spectra. *J. Am. Soc. Mass Spectrom.*, 7: 233–242.
70. Yalcin, T., Khouw, C., Csizmadia, I., Peterson, M., and Harrison, A. (1995) Why are B ions stable species in peptide spectra? *J. Am. Soc. Mass Spectrom.*, 6: 1165–1174.
71. Nold, M.J., Wesdemiotis, C., Yalcin, T., and Harrison, A.G. (1997) Amide bond dissociation in protonated peptides. Structures of the N-terminal ionic and neutral fragments. *Int. J. Mass Spectrom. Ion Proc.*, 164: 137–153.
72. Nair, H., Somogyi, A., and Wysocki, V.H. (1996) Effect of alkyl substitution at the amide nitrogen on amide bond cleavage: electrospray ionization/surface-induced dissociation fragmentation of substance P and two alkylated analogs. *J. Mass Spectrom.*, 31: 1141–1148.
73. Summerfield, S.G., Whiting, A., and Gaskell, S.J. (1997) Intraionic interactions in electrosprayed peptide ions. *Int. J. Mass Spectrom. Ion Proc.*, 162: 149–161.
74. Tsaprailis, G., Nair, H., Somogyi, Á, Wysocki, V.H., Zhong, W., and Futrell, J.H. (1999) Influence of secondary structure on the fragmentation of protonated peptides. *J. Am. Chem. Soc.*, 121: 5142–5154.
75. Zhang, Z. and Smith D.L. (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.*, 2: 522–531.
76. Farmer, T.B. and Caprioli R.M. (1998) Determination of protein–protein interactions by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.*, 33: 697–704.
77. Sinz, A. (2003) Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.*, 38: 1225–1237.
78. Back, J.W., de Jong, L., Muijsers, A.O., and de Koster, C.G. (2003) Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.*, 331: 303–313.
79. Nadeau, O. and Carlson, G. (2002) Chemical cross-linking in studying protein-protein interactions. *Protein-Protein Interactions: A Molecular Cloning Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY; 75–92.
80. Cohen, S.L., Ferre-D'amare, A.R., Burley, S.K., and Chait, B.T. (1995) Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Prot. Sci.*, 4: 1088–1099.