1 Posttranslational Modification (PTM) of Proteins

The study of posttranslational modification (PTM) of proteins using mass spectrometry (MS) approaches has now become a well-matured area of study. There are numerous approaches toward applying chromatography coupled with MS for PTM studies. The liquid chromatography (LC) front-end separation approach of choice is now nanoflow/nano-electrospray, which allows increased sensitivity over previous LC methodology. This book looks at recent developments in PTM studies using MS and proteomic techniques with a focus upon a number of actual studies designed to instruct and highlight modern methodological approaches. A brief overview of nano-electrospray/nanoflow LC-MS is presented in Section 1.3.

COPYRIGHTED MATERIAL

1.1 OVER 200 FORMS OF PTM OF PROTEINS

In the genomic sequencing field, the use of robotic gene sequencers allowed large-scale sequencing that was essentially automated. The robotic automation of determining gene sequences is possible because the sequences involved with genes involve only four bases (see "Overview of Nucleic Acids" in Section 1.4), and there are no variations induced in the form of postmodification. This has resulted in the well-publicized entire sequencing of the human genome (Human Genome Project, *Nature*, February 2001). This is not the case with proteins where there is not only the observance of spliced variants from alternative splicing from the messenger ribonucleic acid (mRNA), there are also PTMs that can take place with the amino acids contained within the protein. There are over 200 PTMs that can take place with proteins as has been described by Wold.[1] As examples, here are 22 different types

Proteomics of Biological Systems: Protein Phosphorylation Using Mass Spectrometry Techniques, First Edition. Bryan M. Ham.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

of PTMs that can take place with proteins: acetylation, amidation, biotinylation, C-mannosylation, deamidation, farnesylation, formylation, flavinylation, gamma-carboxyglutamic acids, geranyl-geranylation, hydroxylation, lipoxylation, myristoylation, methylation, *N*-acyl diglyceride (tripalmitate), *O*-GlcNAc, palmitoylation, phosphorylation, phosphopantetheine, pyrrolidone carboxylic acid, pyridoxyl phosphate, and sulfation.[2] There are also artifactual modifications such as oxidation of methionine (Met). A brief overview of proteins and an introduction to proteomics is presented in Section 1.5.

1.2 THREE MAIN TYPES OF PTM STUDIED BY MS

Of these, the three types of PTM that are primarily observed and studied using mass spectrometric techniques are glycosylation, sulfation, and phosphorylation. The observance of PTM is increasingly being used in expression studies where a normal state proteome is being compared with a diseased state proteome. However, the PTM of a protein during a biological or physiological change within an organism may take place without any change in the abundance of the protein involved and often, is one piece of a complex puzzle. Methods that measure PTM using mass spectrometric methodologies often focus on the degree (increase or decrease, or alternatively, upregulation or downregulation) of PTM for any given protein or proteins. We shall briefly look at glycosylation and sulfation, which are less involved in cellular processes than phosphorylation, a major signaling cascade pathway for the response to a change in cellular condition(s).

1.3 OVERVIEW OF NANO-ELECTROSPRAY/ NANOFLOW LC-MS

1.3.1 Definition and Description of MS

During the past decade, MS has experienced a tremendously large growth in its uses for extensive applications involved with complex biological sample analysis. MS is basically the science of the measurement of the mass-to-charge ratio (m/z) of ions in the gas phase (GP). Mass spectrometers are generally composed of three components: (1) an ionization source that ionizes the analyte of interest and effectively transfers it into the GP, (2) a mass analyzer that separates positively or negatively charged ionic species according to their mass-to-charge ratio

(m/z), and (3) a detector used to measure the subsequently separated GP ions. Mass spectrometers are computer controlled, which allows the collection of large amounts of data and the ability to perform various and complex experiments with the mass spectral instruments. Applications of MS include unknown compound identification, known compound quantitation, structural determination of molecules, GP thermochemistry studies, ion–ion and ion molecule studies, and molecule chemical property studies. MS is routinely used to determine elements such as $Li^+, Na^+, Cl^-, Mg,^{2+}$ inorganic compounds such as $Li^+(H_2O)_x$ or $(TiO_2)^+_x$, and organic compounds including lipids, proteins, peptides, carbohydrates, polymers, and oligonucleotides (deoxyribonucleic acid [DNA]/RNA).

1.3.2 Basic Design of Mass Analyzer Instrumentation

Typical mass spectrometric instrumentation that is used in laboratories and research institutions is composed of six components: (1) an inlet, (2) an ionization source, (3) a mass analyzer, (4) a detector, (5) a data processing system, and (6) a vacuum system. Figure 1.1 illustrates the interrelationship of the six components that make up the fundamental construction of a mass spectrometer. The *inlet* is used to introduce a sample into the mass spectrometer and can be a solid probe, a manual syringe or syringe pump system, a gas chromatograph, or a liquid chromatograph. The inlet system can be either at atmospheric pressure as is shown in Figure 1.1 or at a reduced pressure under vacuum. The *ionization source* functions to convert neutral molecules into charged analyte ions, thus enabling their mass analysis. The ionization source can also be part of the inlet system. A typical inlet system and ionization source that is used with high-performance liquid chromatography (HPLC) is electrospray ionization (ESI). In an HPLC/ESI inlet system and ionization source, the effluent coming from the HPLC column is transferred into the ESI capillary that has a high voltage applied to it inducing the ESI process. In this configuration, the inlet system and ionization source are located at atmospheric pressure outside of the mass spectrometric instrumentation that is under vacuum. The spray that is produced passes through a tiny orifice that separates the internal portion of the mass spectrometer that is under vacuum from its ambient surroundings that are at atmospheric pressure. This orifice is also often called the *inlet* and/or the *source*. In the case of the coupling of a gas chromatograph to the mass spectrometer, the capillary column of the gas chromatograph is inserted through a heated transfer capillary directly into the internal portion of the mass spectrometer that is under

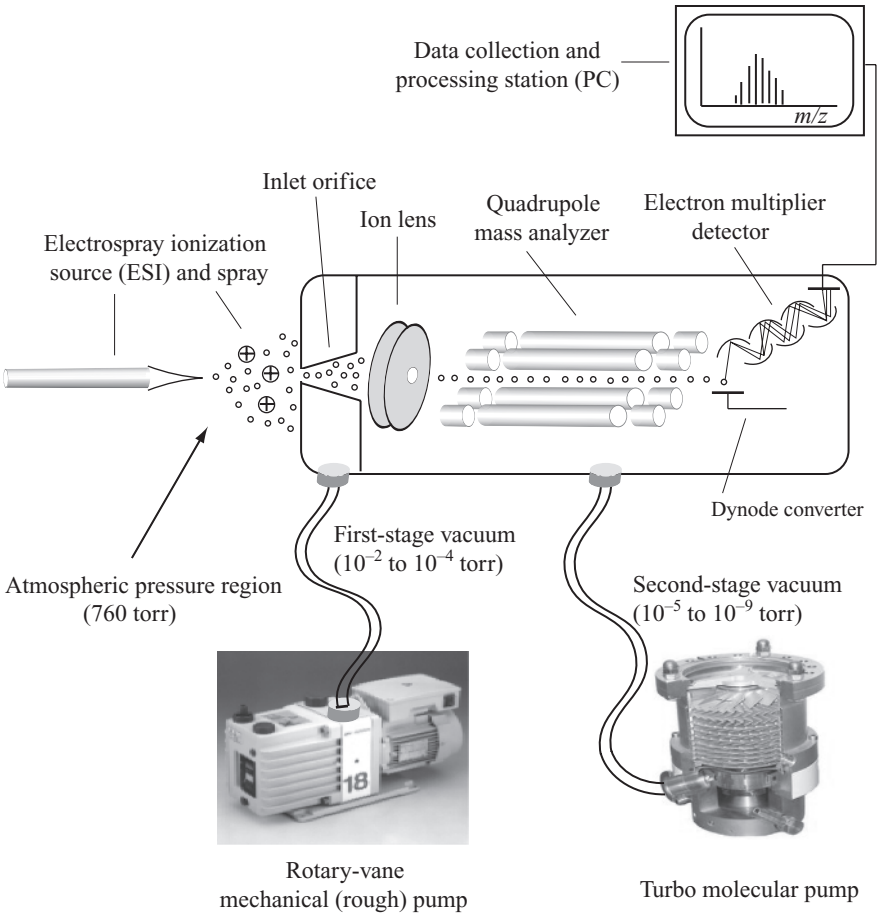

Figure 1.1. The six components that make up the fundamental configuration of mass spectrometric instrumentation composed of (1) inlet and ionization system, (2) inlet orifice (source), (3) mass analyzer, (4) detector, (5) vacuum system, and (6) data collection and processing station (PC). [See Wikipedia, *turbomolecular pump*, http://en.wikipedia.org/w/index.php?title=Turbomolecularpump&oldid=71160479 (as of August 24, 2006, 17:45 GMT)].

vacuum. This is possible due to the fact that the species eluting from the capillary column are already in the GP, making their introduction into the mass spectrometer more straightforward as compared with the liquid eluant from an HPLC where analytes must be transferred from the solution phase to the GP. An example of an ionization process that takes place under vacuum in the front end of the mass spectrometer is a process called matrix-assisted laser desorption ionization or MALDI. In this ionization technique, a laser pulse is directed toward a MALDI

target that contains a mixture of the neutral analytes and a strongly UV-absorbing molecule, often times a low-molecular-weight organic acid such as dihydroxybenzoic acid (DHB). The analytes are lifted off of the MALDI target plate directly into the GP in an ionized state. This is due to transference of the laser energy to the matrix and then to the analyte. The MALDI technique takes place within a compartment that is at the beginning of the mass spectrometer instrument and is under vacuum. The compartment that this takes place is often called the ionization source, thus combining the *inlet system* and the *ionization source* together into one compartment. As illustrated in Figure 1.1, the analyte molecules (small circles), in an ionized state, pass from atmospheric conditions to the first stage of vacuum in the mass spectrometer through an inlet orifice that separates the mass spectrometer that is under vacuum from ambient conditions. The analytes are guided through a series of ion lenses into the mass analyzer. The *mass analyzer* is the heart of the system, which is a separation device that separates positively or negatively charged ionic species in the GP according to their respective mass-to-charge ratios. The mass analyzer GP ionic species separation can be performed by an external field such as an electric field or a magnetic field or by a field-free region such as within a drift tube. For the detection of the GP-separated ionic species, electron multipliers are often used as the *detector*. Electron multipliers are mass impact detectors that convert the impact of the GP-separated ionic species into a cascade of electrons, thereby multiplying the signal of the impacted ion many times fold.

The *vacuum system* ties into the inlet, the source, the mass analyzer, and the detector of the mass spectrometer at different stages of increasing vacuum as movement goes from the inlet to the detector (left to right in Fig. 1.1). It is very important for the mass analyzer and detector to be under high vacuum as this removes ambient gas, thereby reducing the amount of unwanted collisions between the mass-separated ionic species and gas molecules present. As illustrated in Figure 1.1, ambient, atmospheric conditions are generally at a pressure of 760 torr. The first-stage vacuum is typically at or near 10^{-3} torr immediately following the inlet orifice and around the first ion transfer lenses. This stage of vacuum is obtained using two-stage rotary vane mechanical pumps that are able to handle high pressures such as atmospheric and large variation in pressures but are not able to obtain the lower pressures that are required further into the mass spectrometer instrument. The two-stage rotary vane mechanical pump has an internal configuration that utilizes a rotating cylinder that is off-axis within the pump's hollow body. The off-axis-positioned rotor contains two vanes that are opposed and

directed radially and are spring controlled to make pump body contact. As the cylinder rotates, the volume between the pump's body and the vanes changes; the volume increases behind each vane that passes a specially placed gas inlet port. This will cause the gas to expand behind the passing vane, while the trapped volume between the exhaust port and the forward portion of the vane will decrease. The exhaust gas is forced into a second stage and is then released by passing through the oil that is contained within the pump's rear oil reservoir. This configuration is conducive for starting up at atmospheric pressure and working toward pressures usually in the range of 10^{-3} to 10^{-4} torr.

The lower stages of vacuum are obtained most often using turbo molecular pumps as illustrated in Figure 1.1. Turbo molecular pumps are not as rugged as the mechanical pumps described previously and need to be started in a reduced pressure environment. Typically, a mechanical pump will perform the initial evacuation of an area. When a certain level of vacuum is obtained, the turbo molecular pumps will then turn on and bring the pressure to higher vacuum. Using a mechanical vane pump to provide a suitable forepump pressure for the turbo molecular pump is known as roughing or "rough out" the chamber. Therefore, two-stage rotary vane mechanical pumps are often referred to as rough pumps. As illustrated in Figure 1.1, the turbo molecular pump contains a series of rotor/stator pairs that are mounted in multiple stages. The principle of turbomolecular pumps is to transfer energy from the fast rotating rotor (turbo molecular pumps operate at very high speeds) to the molecules that make up the gas. After colliding with the blades of the rotor, the gas molecules gain momentum and move to the next lower stage of the pump and repeat the process with the next rotor. Eventually, the gas molecules enter the bottom of the pump and exit through an exhaust port. As gas molecules are removed from the head or beginning of the pump, the pressure before the pump is continually reduced as the gas is removed through the pump, thus achieving higher and higher levels of vacuum. Turbo molecular pumps can obtain much higher levels of vacuum (up to 10^{-9} torr) as compared with the rotary vane mechanical pumps (up to 10^{-4} torr).

The final component of the mass spectrometer is a data processing system. This is typically a personal computer (PC) allowing the mass spectrometric instrumentation to be software controlled, enabling precise measurements of carefully designed experiments and the collection of large amounts of data. Commercially bought mass spectrometers will come with its own software that is used to set the operating parameters of the mass spectrometer and to collect and interpret the data, which is in the form of mass spectra.

1.3.3 ESI

ESI is a process that enables the transfer of compounds in solution phase to the GP in an ionized state, thus allowing their measurement by MS. The use of ESI coupled with MS was pioneered by Whitehouse et al.[3] and Fenn[4] in 1985 and 1993 by extending the work of Dole et al.[5] in 1968, who demonstrated the production of GP ions by spraying macromolecules through a steel capillary that was electrically charged and subsequently monitoring the ions with an ion-drift spectrometer. The process by which ESI works has received much theorization, study, and debate,[6–12] in the scientific community, especially the formation of the ions from the Taylor[13] cone droplets and offspring droplets. Figure 1.2

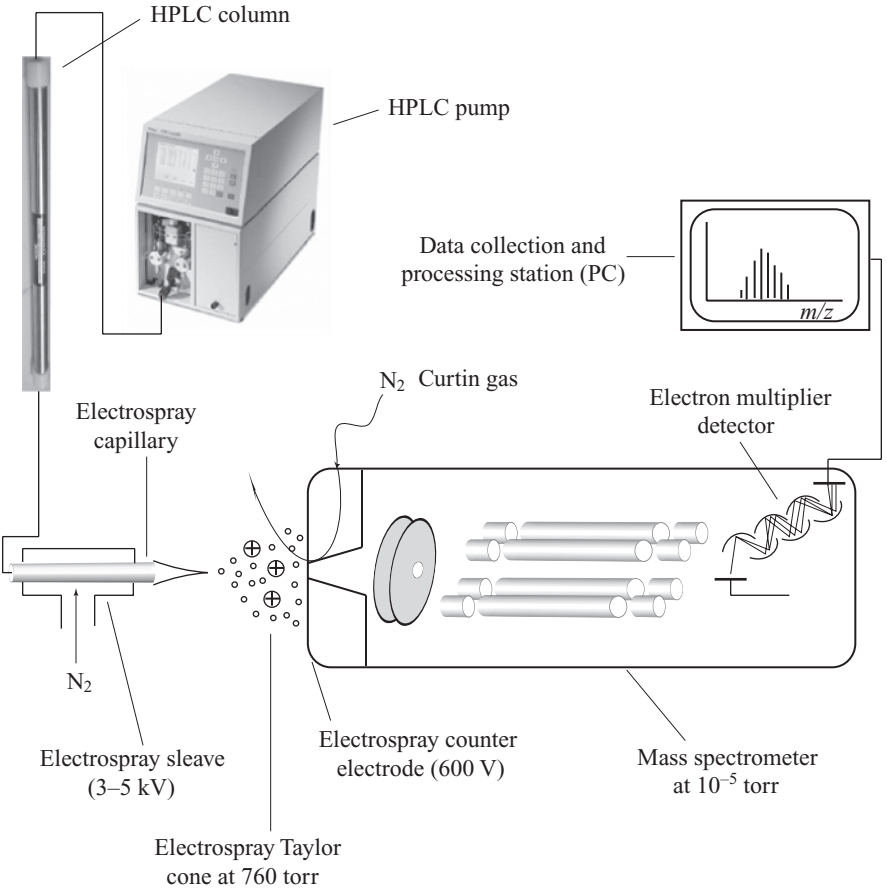

Figure 1.2. General setup for ESI when measuring biomolecules by electrospray mass spectrometry.

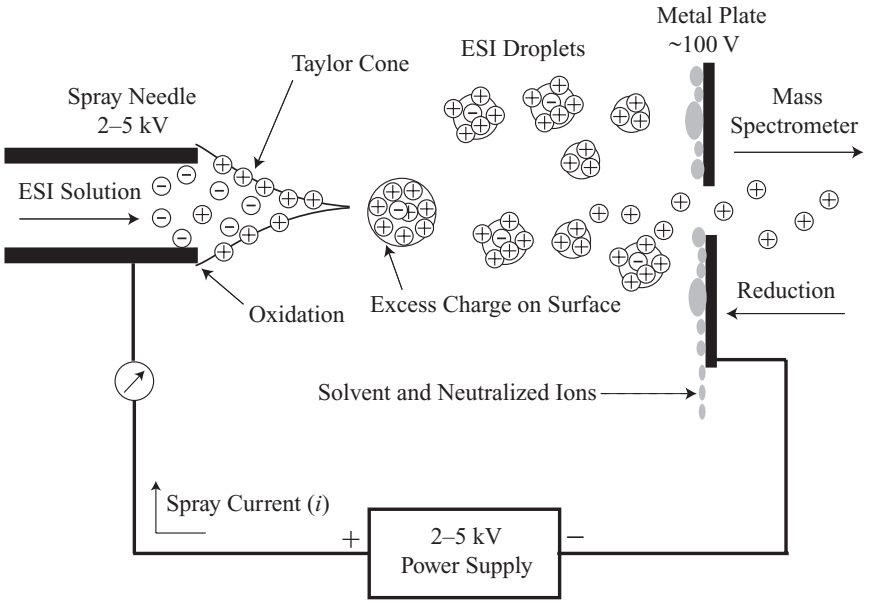

Figure 1.3. Electrospray ionization process illustrated in positive ion mode. (Reprinted with permission of John Wiley & Sons, Inc. Cech, N.B., and Enke, C.G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrometry Reviews*, 2001, *20*, 362–387. Copyright 2001.)

shows the general setup for ESI when measuring biomolecules by electrospray MS. The electrospray process is achieved by placing a potential difference between the capillary and a flat counter electrode. This is illustrated in Figure 1.3 where the "spray needle" is the capillary and the "metal plate" is the flat counter electrode. The generated electric field will penetrate into the liquid meniscus and create an excess abundance of charge at the surface. The meniscus becomes unstable and protrudes out, forming a Taylor cone. At the end of the Taylor cone, a jet of emitting droplets (number of drops estimated at 51,250 with radius of 1.5 μm) will form that contains an excess of charge. Pictures of jets of offspring droplets are illustrated in Figure 1.4. As the droplets move toward the counter electrode, a few processes take place. The drop shrinks due to evaporation, thus increasing the surface charge until columbic repulsion is great enough that offspring droplets are produced. This is known as the Rayleigh limit, producing a columbic explosion. The produced offspring droplets have 2% of the parent droplets' mass and 15% of the parent droplets' charge. This process will continue until the drop contains one molecule of analyte and charges that are associated with basic sites (positive ion mode). This is referred to as the

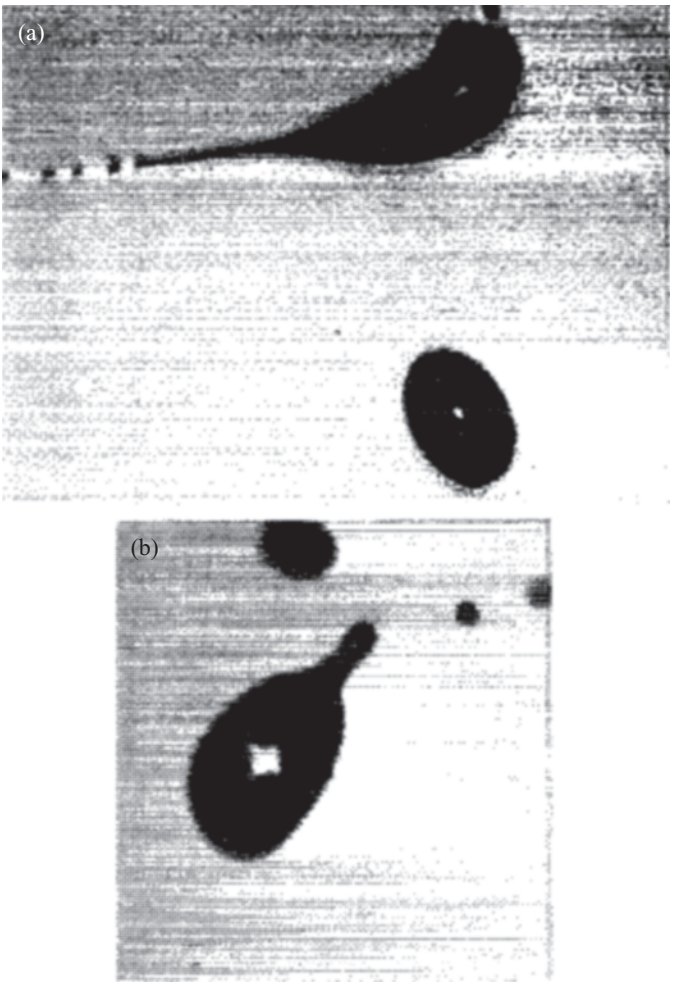

Figure 1.4. Pictures illustrating the jet production of offspring droplets. (Reprinted with permission from Alessandro Gomez, *Physics of Fluids*, 6, 404 (1994). Copyright 1994, American Institute of Physics.)

"charged residue model" that is most important for large molecules such as proteins. This process is illustrated in Figure 1.5. As the droplets move toward the counter electrode, a second process also takes place known as the "ion evaporation model." In this process, the offspring droplet will allow evaporation of an analyte molecule from its surface along with charge when the charge repulsion of the analyte with the solution is great enough to allow it to leave the surface of the drop. This usually takes place for droplets with a radius that is less than 10 nm. This type of ion formation is most important for small molecules.

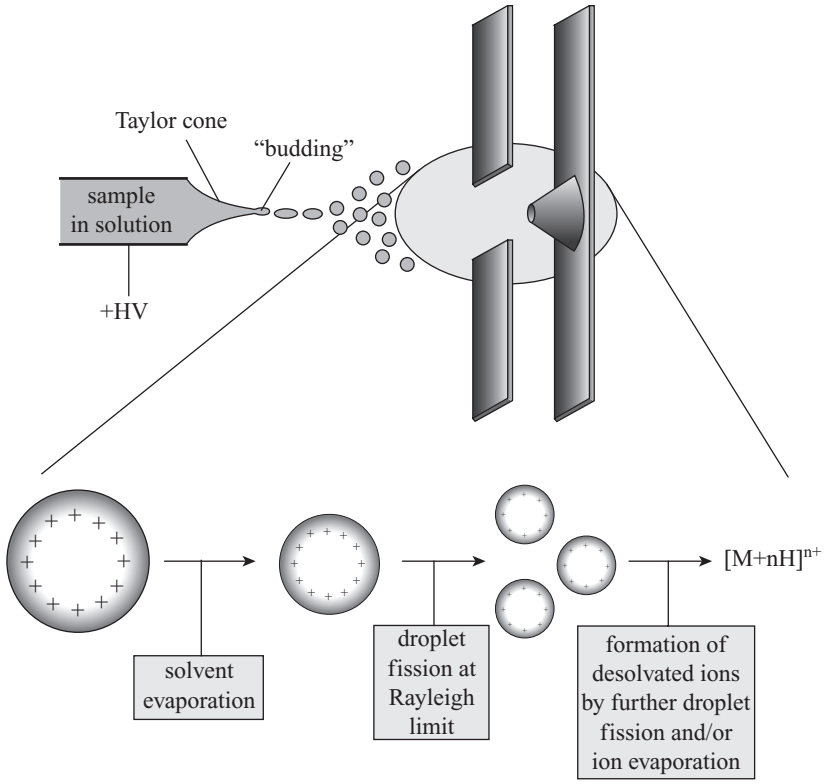

Figure 1.5. Gas-phase ion formation process from electrospray droplets. (Reprinted from Gaskell, S.J. Electrospray: principles and practice. *J. Mass Spectrom.* 1997, 32, 677–688. Copyright John Wiley & Sons Limited 1997. Reproduced with permission.)

In the ensuring years since its introduction, electrospray MS has been used for structural elucidation and fragment information,[14–16] and noncovalent complex studies,[17,18] just to name a few recent examples of its overwhelmingly wide range of applications.

Electrospray[3,4,7,8] is an ionization method that is now well known to produce intact GP ions with very minimal, if any, fragmentation being produced during the ionization process. In the transfer process of the ions from the condensed phase to the GP, several types of "cooling" processes of the ions are taking place in the source: (1) cooling during the desolvation process through vibrational energy transfer from the ion to the departing solvent molecules, (2) adiabatic expansion of the electrospray as it enters the first vacuum stage, (3) evaporative cooling, and (4) cooling due to low-energy dampening collisions with ambient gas molecules. The combination of these effects, and the fact that

electrospray can effectively transfer a solution phase complex to the GP with minimal interruption of the complex, makes the study of non-covalent complexes from solution by ESI-MS attractive.

1.3.4 Nano-ESI

A major application of biomolecule analysis using MS has been the ability to allow liquid flows to be introduced into the source of the mass spectrometer. This has enabled the coupling of HPLC to MS where HPLC is used for a wide variety of biomolecule analysis. Normal ESI, introduced in the preceding section, typically has flow rates in the order of microliters per minute (~1–500 μL/min). Traditional analytical HPLC systems designed with UV/Vis detectors generally employ flow rates in the range of milliliters per minute (~0.1–1 mL/min). A recent advancement in the ESI technique has been the development of nano-electrospray where the flows employed are typically in the range of nanoliters per minute (~1–500 nL/min). Following the progression of the development of electrospray from Dole's original reporting in 1968 through Fenn's work reported in 1984 and 1988, a more efficient electrospray process was reported by Wilm et al.[19] employing flows in the range of 25 nL/min. This early reporting of low flow rate electrospray was initially termed as microelectrospray by Wilm et al. but was later changed to nano-electrospray.[20] At the same time that Wilm et al.[19] reported the microelectrospray, Caprioli et al.[21] also reported a miniaturized ion source that they had named microelectrospray. The name "nanoelectrospray" for Wilm's source is actually more descriptive due to flow rates used in the nanoliter per minute range and the droplet sizes that are produced in the nanometer range. Conventional electrospray sources before the introduction of nano-electrospray produced droplets on the order of 1–2 μm. The nano-electrospray source produces droplets in the size range of 100–200 nm, which is 100–1000 times smaller in volume. When spraying standard solutions at concentrations of 1 pmol/μL, it is estimated that droplets of the nanometer size contain only one analyte molecule per droplet.

The original nano-electrospray sources that were used were composed of pulled fused-silica capillary tips 3–5 cm long with orifices of 1–2 μm in diameter. The tips also have thin gold plating that allows current flow. The tips are loaded with 1–5 μL of sample directly using a pipette[22] and coupled to the electrospray source, completing the closed circuit required for the production of the applied voltage electrospray Taylor cone generation. This is illustrated in Figure 1.6 where in the top portion of the figure a sample is being loaded into the

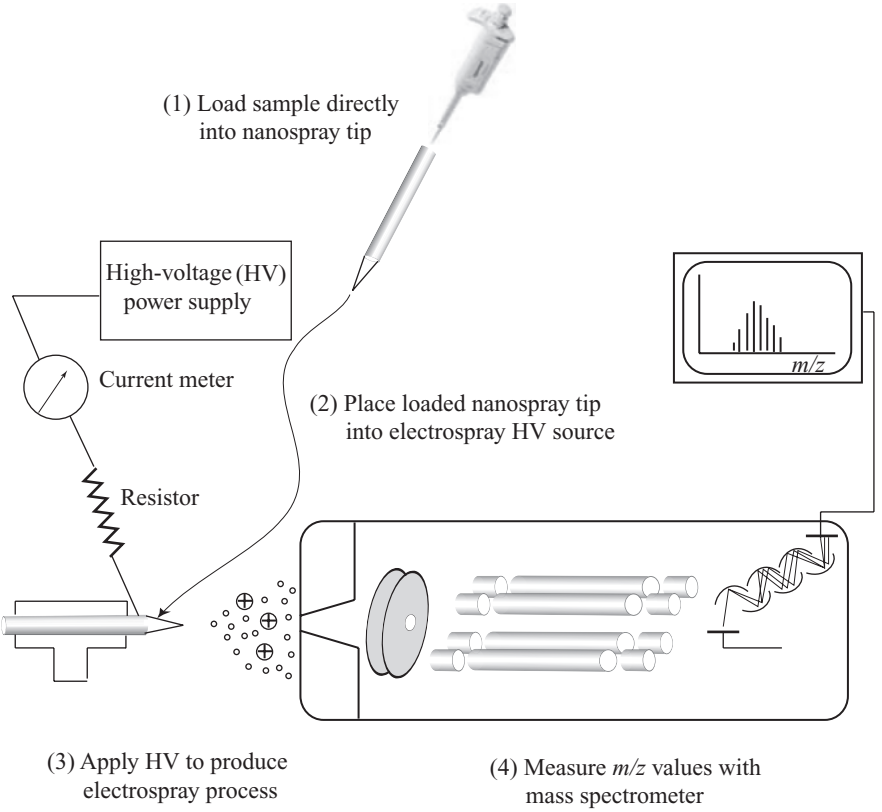

Figure 1.6. Top of figure illustrates the loading of a nano-electrospray tip. Bottom of figure illustrates the coupling of the nano-electrospray tip to the closed-circuit system.

nanospray tip using a pipette. The tip is then placed into the closed-circuit system for the electrospray to take place. The sample flow rate is very low using the nanospray tips allowing the measurement of a very small sample size over an extended period of time. It has also been observed that nanospray requires a lower applied voltage for the production of the electrospray that helps to reduce problems with corona electrical discharges that will interrupt the electrospray. In nano-electrospray, the flow rate is lower than in conventional electrospray and is felt to have a direct impact on the production of the droplets within the spray and the efficiency of ion production. The lower flow rate produces charged droplets that are reduced in size as compared with conventional electrospray. This has been described in detail by Wilm et al.,[19] by Fernandez de la Mora et al.,[23] and by Pfeifer and Hendricks.[24] There are fewer droplet fission events required with smaller

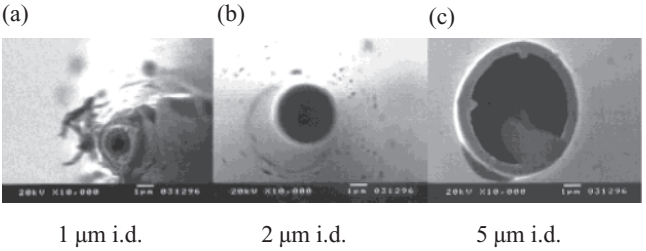

Figure 1.7. Illustration of different nano-electrospray tip orifice diameters. Scanning electron microscopy images of employed nanospray emitters: (a) 1-, (b) 2-, and (c) 5-μm tip. Images were obtained after 2 hours of use. (Reprinted with permission from Li, Y.; Cole, R.B. Shifts in Peptide and Protein Charge State Distributions with Varying Spray Tip Orifice Diameter in Nano-Electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* 2003, 75, 5739–5746. Copyright 2003 American Chemical Society.)

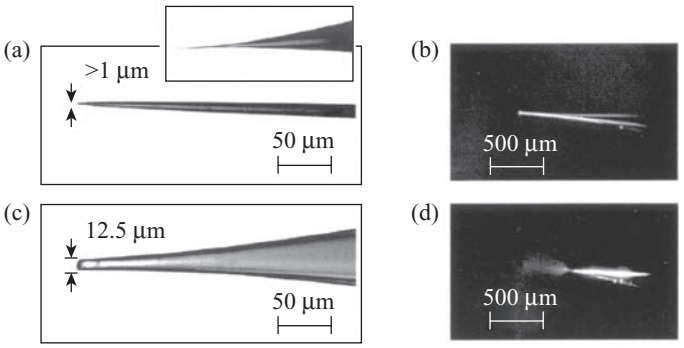

Figure 1.8. Examples of nanospray tip sizes and the influence upon the ESI Taylor cone. The cone is not observed in (b) at a diameter of >1 μm. The cone is observed in (d) for a diameter of 12.5 μm. (Reprinted with permission. This article was published in *J Am Soc Mass Spectrom*, Schmidt, A., Karas, M. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI?, 2003, 14, 492–500. Copyright Elsevier 2003.)

initial droplets in conjunction with less solvent evaporation taking place before ion release into the GP.[25,26] A result of this is that a larger amount of the analyte molecule is transferred into the mass spectrometer for analysis. Though the efficiency of ionization is increased with nano-electrospray, the process is also influenced by the size and shape of the orifice tip.[27,28] Pictures of nano-electrospray orifice tips are illustrated in Figure 1.7. Figure 1.8 shows an example of the production and observance of and ESI Taylor core.

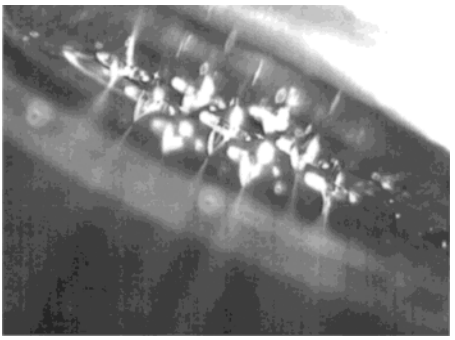

Figure 1.9. Photograph of nine stable electrosprays generated from the nine-spray emitter array. (Reprinted with permission from Tang, K.; Lin, Y.; Matson, D.W.; Kim, T.; Smith, R.D. *Anal. Chem.* 2001, 73, 1658–1663. Copyright 2001 American Chemical Society.)

While Figure 1.8d does show a Taylor cone formed, Figure 1.9 gives a good picture of an array of Taylor cones formed from a microelectrospray emitter. In the figure, multiple cones can be seen along with their associated spray produced from the electrospray process.

As mentioned previously, nano-HPLC is increasingly being coupled to nano-electrospray for biomolecule analysis. A nano-HPLC-ESI system is illustrated in Figure 1.10. The flow involved in nano-HPLC-ESI often ranges between 10 and 100 nL/min. The fused-silica capillary columns that are used in nano-HPLC have very small diameters, often around 50 μm. These small-diameter columns can often create high back pressures in the HPLC system. One way to achieve the very low flow rate through the fused-silica nano-HPLC column is to use a flow splitter that is located in-stream between the column and the HPLC pump as illustrated in Figure 1.10. The tubing from the splitter to waste is called a restrictor and is used to regulate the flow through the nanocolumn. A smaller-diameter restrictor used will increase the back pressure, forcing more mobile phase through the nanocolumn. If a larger diameter restrictor is used, the back pressure will be lower, resulting is less flow being directed through the column. The nanocolumns have a nano-ESI tip coupled to them (diameters can range from 1 μm up to 100 μm) to produce the electrospray. Another difference observed here as compared with the atmospheric pressure source is the absence of a nebulizing gas or a drying gas. These are not needed or used in nano-ESI.

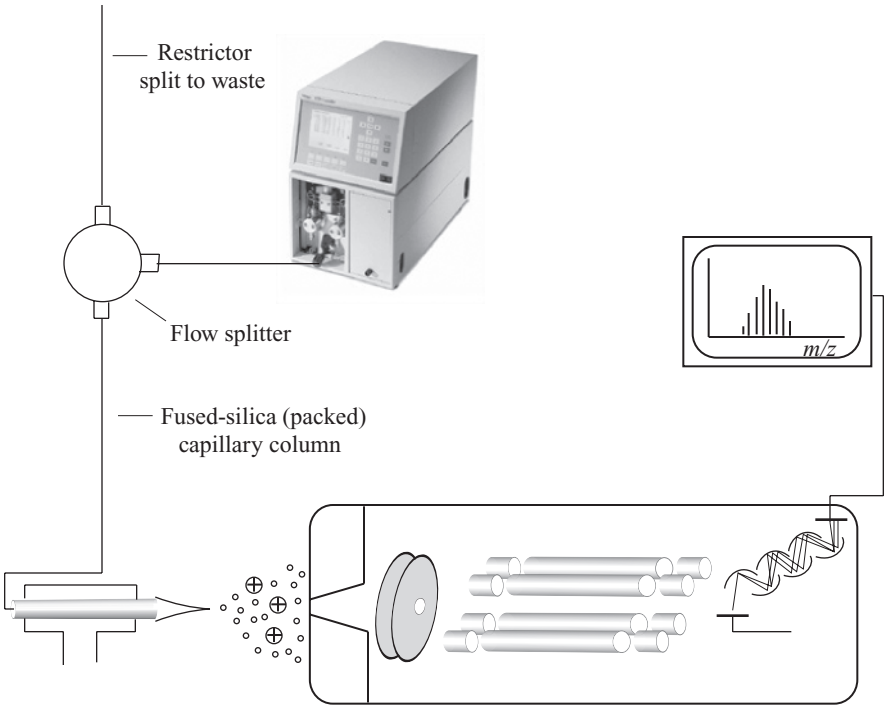

Figure 1.10. Design of a nano-HPLC nano-ESI system for mass spectrometric analysis of biomolecules.

1.4 OVERVIEW OF NUCLEIC ACIDS

Nucleic acids are an important consideration in PTM study; they are present in cellular protein extracts and must be separated. In traditional studies, the Trizol precipitation method was used to isolate a nucleic acid fraction and a protein fraction. This afforded the opportunity to study both from an extraction. Due to their close relationship and importance, we will look at a brief overview of nucleic acids and their measurement by MS.

Nucleic acids are also analyzed using mass spectrometric techniques, and we will start with a background look at the makeup of nucleic acids before looking at the MS. In contrast to polysaccharides, and similar to proteins, nucleic acids are specifically directional in their makeup and contain nonidentical monomers that have a distinct sequence that produce informational macromolecules. The nucleic acids reside in the nucleus of the cell and are the storage, expression, and transmission of genetic information of living species. The two types of nucleic acids are

DNA and RNA. There are two distinct parts of their chemical structure and makeup that differentiate the two. First, DNA contains the five-carbon sugar deoxyribose, while RNA contains ribose; and second, DNA contains the base thymine (T), while RNA contains uracil (U). The molecules that make up the DNA and RNA structures are illustrated in Figure 1.11. This constitutes the purine bases adenine (A) and guanine (G), the pyrimidine bases cytosine (C), uracil (U), and thymine. Also illustrated in Figure 1.11 are the two sugars D-deoxyribose and D-ribose and finally, the phosphate group that acts as the backbone of the nucleic acids linking the nucleotides together. Nucleotides are the monomeric units that make up the nucleic acids. There are actually only four nucleotides that make up DNA and RNA, a much smaller number than the 20 amino acids found in proteins. Examples of nucleotides found in DNA and RNA are illustrated in Figure 1.12. Figure 1.12a is a DNA nucleotide where the number 2′ carbon in the sugar ring contains a hydrogen atom for D-deoxyribose. One of the bases will be attached to the 1′ carbon of the sugar through an aromatic nitrogen, and the phosphate will be attached to the number 5′ sugar carbon with a phosphoester bond. The RNA nucleotide illustrated in Figure 1.12b has the same types of bonding as illustrated for the DNA nucleotide but to a D-ribose sugar. In the case that the phosphate group is removed from the nucleotide, the remaining base sugar structure is called a nucleoside.

The nucleotides are linked to each other through the phosphate group, forming a linear polymer. The nucleotides undergo a condensation reaction through the linking of the phosphate group on the 5′ carbon to the 3′ carbon of the next nucleotide known as a 3′,5′ phosphodiester bond. The resulting polynucleotide therefore has a 5′ hydroxyl group at the start (by convention) and a 3′ hydroxyl group at the end (by convention). Representative linear nucleotide structures are illustrated for RNA and DNA in Figure 1.13.

A similar naming scheme that is used for the fragmentation ions generated by collision-induced dissociation (CID) of peptides was proposed by Glish et al.[29] for nucleic acids and is illustrated in Figure 1.14. There are four cleavage sites producing fragmentation along the phosphate backbone from CID. When the product ion contains the 3′-OH portion of the nucleic acid, the naming includes the letters w, x, y, and z, where the numeral subscript is the number of bases from the associated terminal group. When the product ion contains the 5′-OH portion of the nucleic acid, the naming includes the letters a, b, c, and d. Losses are also more complicated than that shown in Figure 1.14 due to the neutral loss of base moieties. Figure 1.15 illustrates an actual structural

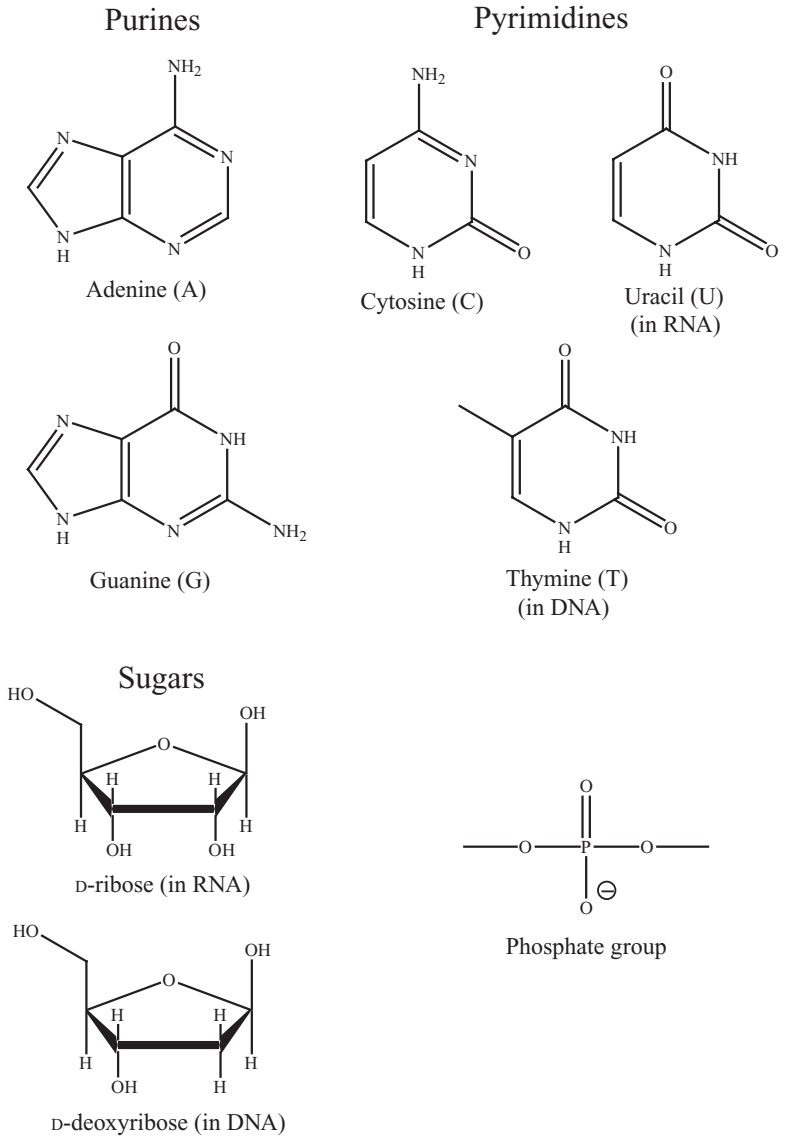

Figure 1.11. Structures of the molecules that make up the nucleic acids DNA and RNA. The purine bases adenine (A) and guanine (G), the pyrimidine bases cytosine (C), uracil (U), and thymine, the sugars D-deoxyribose and D-ribose, and the phosphate group.

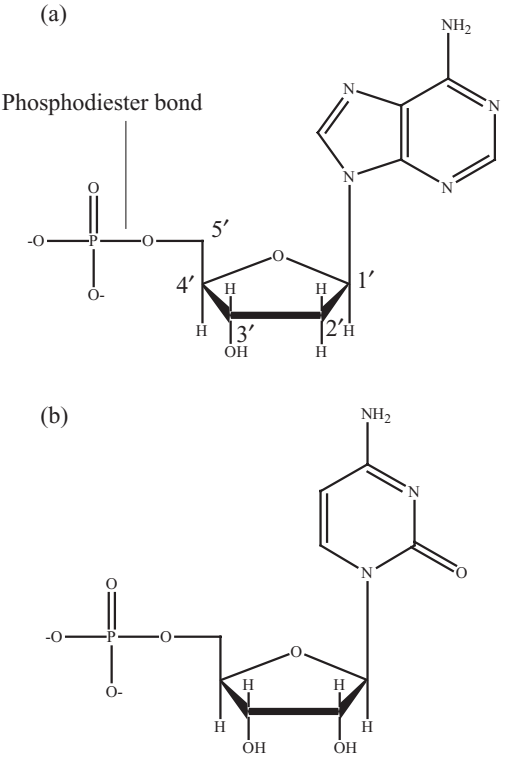

Figure 1.12. (a) DNA nucleotide and (b) RNA nucleotide.

cleavage at the w_2/a_2 site of a 4-mer nucleic acid's phosphate backbone according to the naming scheme of Figure 1.14. Numerous mechanisms have also been reported for the fragmentation pathways leading to charged base loss and also neutral base loss. These are losses that are observed in product ion spectra other than the cleavage along the phosphate backbone that is illustrated in Figures 1.14 and 1.15. Neutral and charged base losses add to the complexity of the product ion spectra but also add information concerning the makeup of the oligonucleotide. Figure 1.16 illustrates a couple examples of proposed fragmentation pathway mechanisms for neutral and charged base losses. In Figure 1.16a, a simple nucleophilic attack on the C-1′ carbon atom by the phosphodiester group results in the elimination of a charged base.[30] Figure 1.16b illustrates a two-step reaction where in the first step there is neutral base loss followed by breakage of the 3′-phosphoester bond.[31] There are other proposed fragmentation pathways for a number of other possible mechanisms for the production of the product ions observed in tandem mass spectra of the nucleic acids.

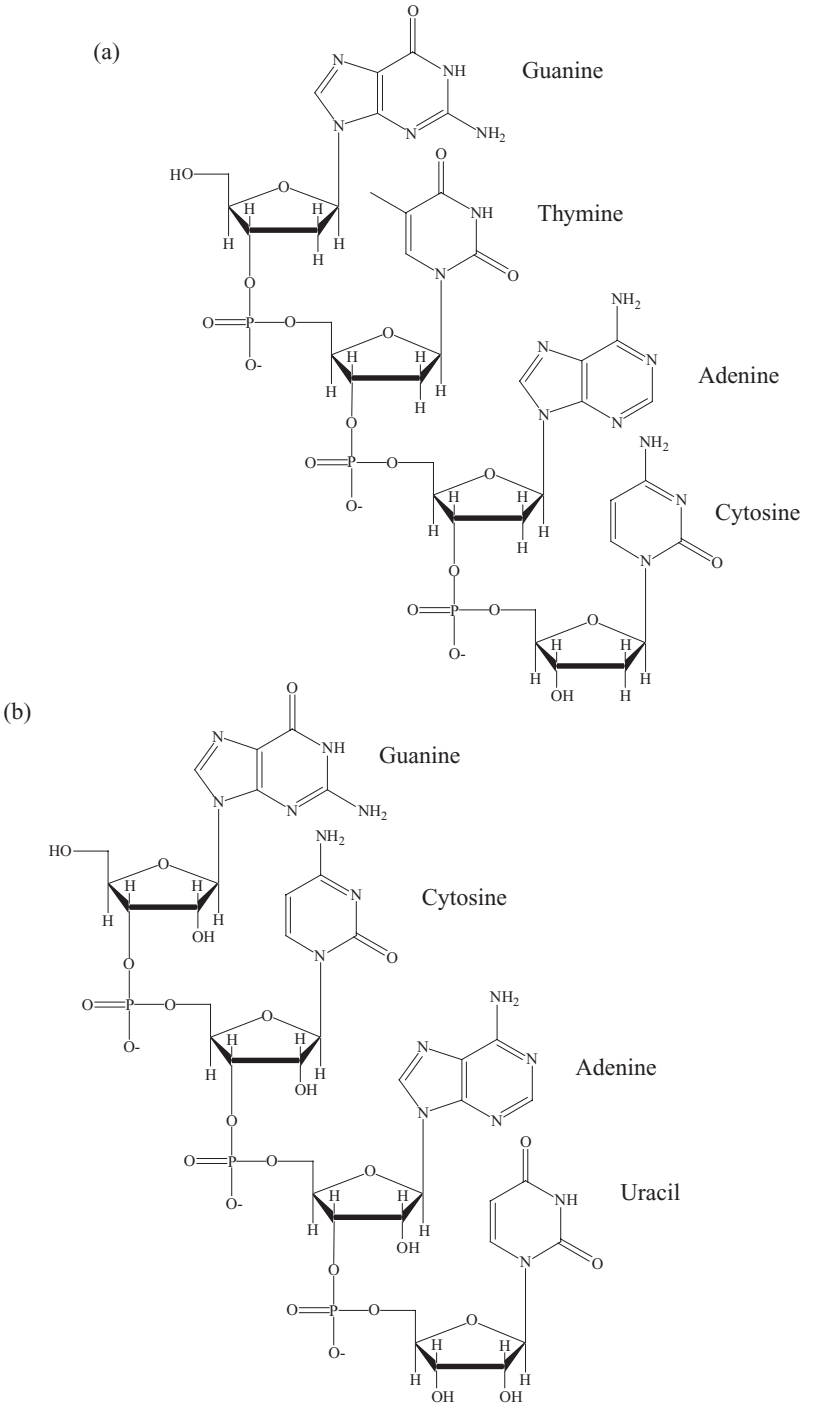

Figure 1.13. Linear nucleic acid structures for (a) DNA and (b) RNA.

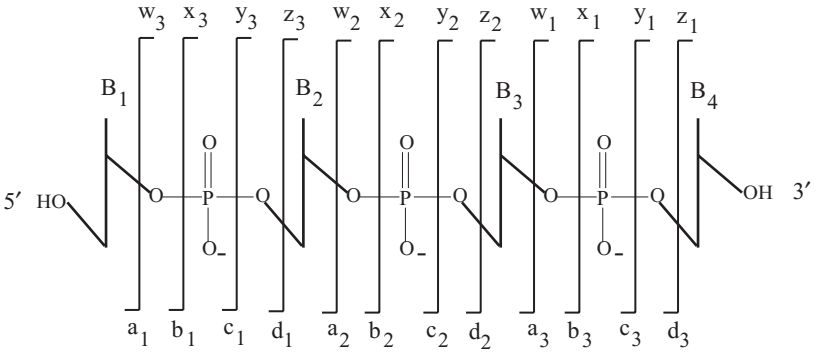

Figure 1.14. Naming scheme for nucleic acid product ions. When the product ion contains the 3′-OH portion of the nucleic acid, the naming includes the letters w, x, y, and z. When the product ion contains the 5′-OH portion of the nucleic acid, the naming includes the letters a, b, c, and d. In both 3′-OH and 5′-OH containing product ions, the numeral subscript is the number of bases from the associated terminal group.

1.5 PROTEINS AND PROTEOMICS

1.5.1 Introduction to Proteomics

Proteomics, the study of a biological system's compliment of proteins (e.g., from cell, tissue, or a whole organism) at any given state in time, has become a major area of focus for research and study in many different fields and applications. In proteomic studies, MS can be employed to analyze either the intact, whole protein or the resultant peptides obtained from enzyme-digested proteins. The mass spectrometric analysis of whole, intact proteins is often called top-down proteomics where the measurement study starts with the analysis of the intact protein in the GP and subsequently investigating its identification and any possible modifications through CID measurements. The mass spectrometric analysis of enzyme-digested proteins that have been converted to peptides is known as bottom-up proteomics. Finally, MS is also used to study PTMs that have taken place with the proteins such as glycosylation, sulfation, and phosphorylation. We shall begin with a look at bottom-up proteomics, the most common approach, followed by top-down proteomics, which is seeing more applications and study lately, and finally, the PTMs of glycosylation, sulfation, and phosphorylation. Bioinformatics has become an important tool used in the interpretation of results obtained from MS studies. In the last part of this chapter, we will briefly look at what bioinformatics is and what it can be used for in relation to MS and proteomic studies. Due to the enormous impact

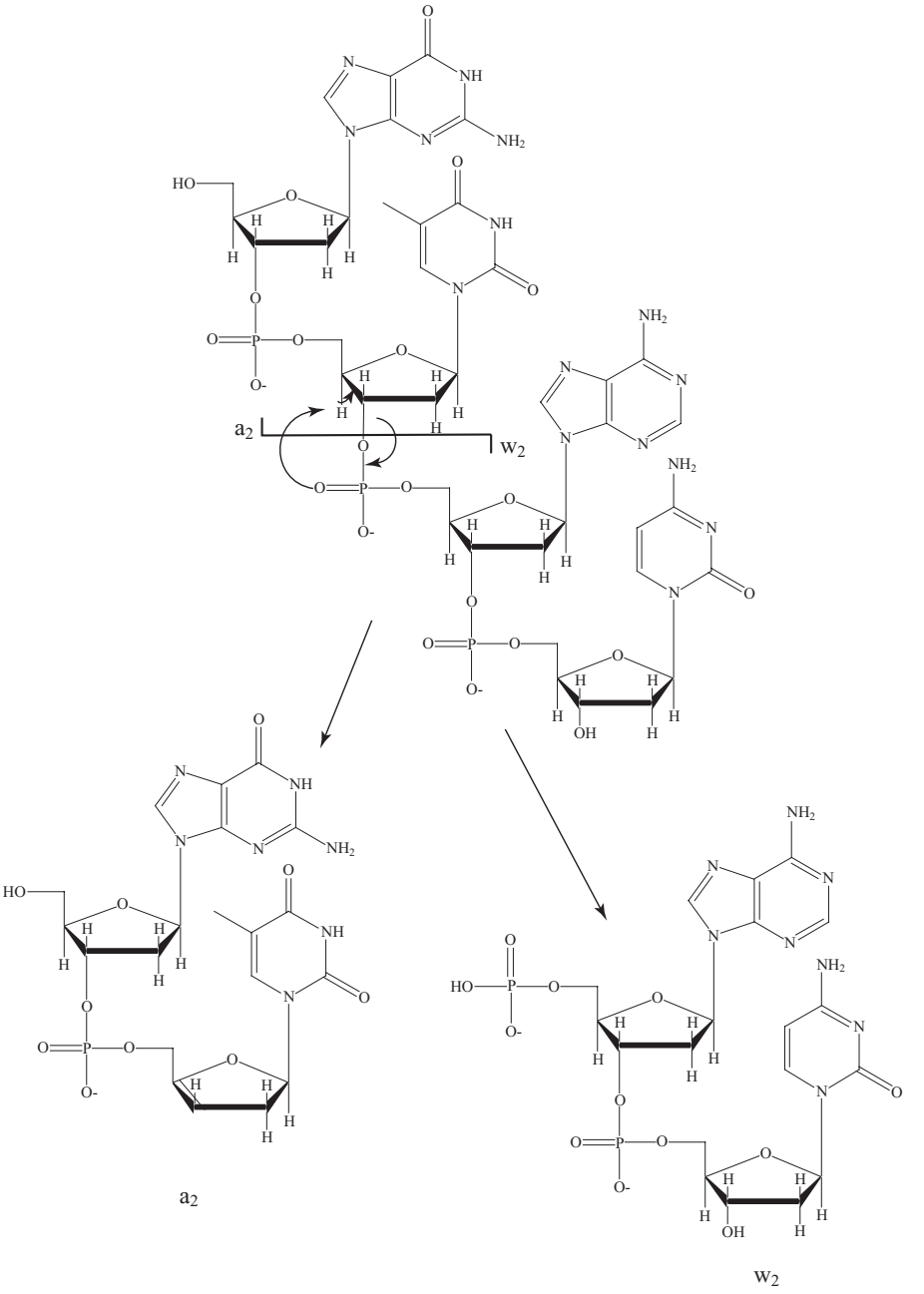

Figure 1.15. Example of structural cleavage at the w_2/a_2 site of a 4-mer nucleic acid's phosphate backbone according to the naming scheme of Figure 1.14.

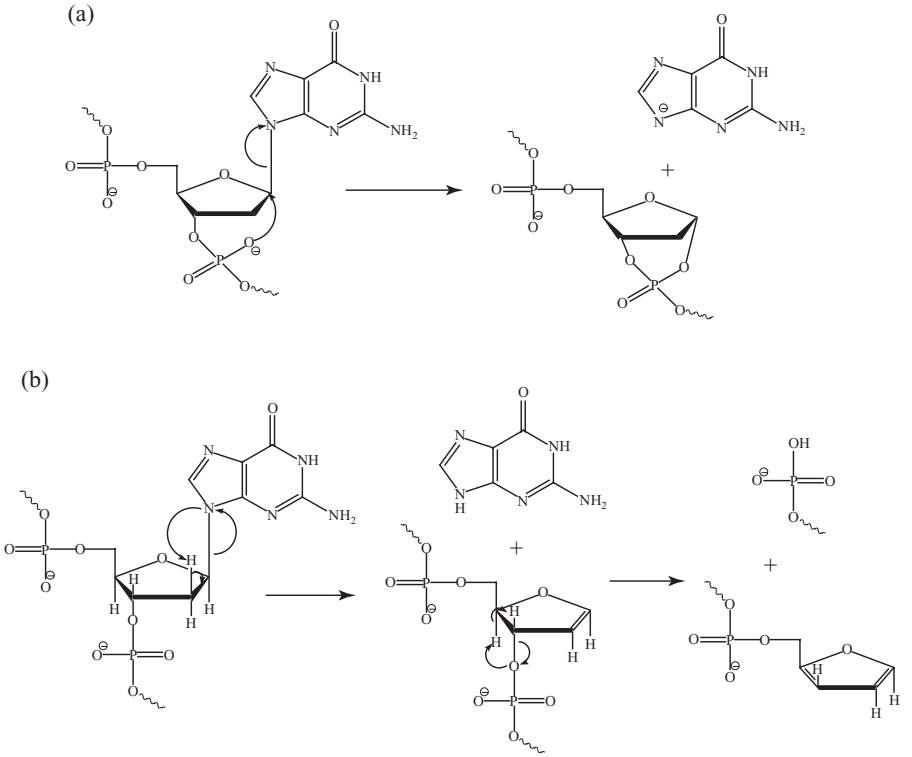

Figure 1.16. Proposed fragmentation pathways associated with the base substituent groups. (a) Nucleophilic attack on the C-1′ carbon atom by the phosphodiester group results in the elimination of a charged base. (b) Two-step reaction mechanism where in the first step there is neutral base loss followed by breakage of the 3′-phosphoester bond.

of proteomics on research into biological processes, organisms, diseased states, tissues, and so on, we will begin this section starting with a brief overview of proteins including their structure and makeup.

1.5.2 Protein Structure and Chemistry

Of all biological molecules, proteins are one of the most important, next only to the nucleic acids. All living cells contain proteins, and their name is derived from the Greek word *proteios*, which has the meaning of "first."[32] There are two broad classifications for proteins related to their structure and functionality: water-insoluble fibrous proteins and water-soluble globular proteins. The three-dimensional configuration of a protein is described by its primary, secondary,

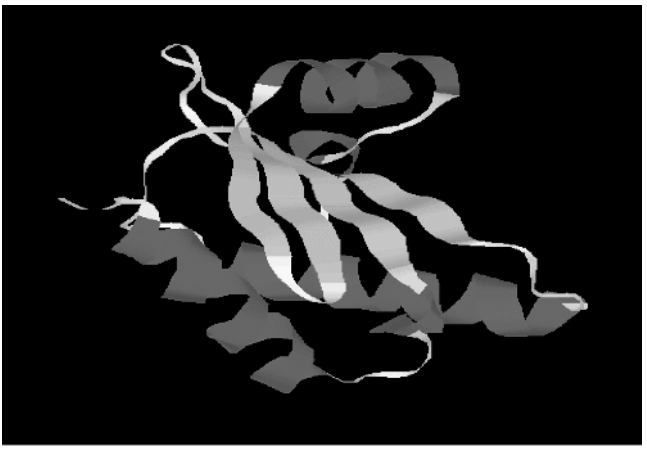

Figure 1.17. Ribbon structure representation of the RNase protein illustrating substructures of alpha helices and beta sheets.

tertiary, and quaternary structures. Figure 1.17 is a three-dimensional ribbon representation of the protein RNase. The primary structures of proteins are made up of a sequence of amino acids forming a polypeptide chain. Typically, if the chain is less than 10,000 Da, the compound is called a polypeptide; if greater than 10,000 Da, the compound is called a protein. There are 20 amino acids that make up the protein chains through carbon to nitrogen peptide bonds. Figure 1.18 illustrates the 20 amino acid structures that make up the polypeptide backbone chain of proteins. Amino acids possess an amino group (NH_2) and a carboxyl group (COOH) that are bonded to the same carbon atom that is alpha to both groups; therefore, amino acids are called alpha amino acids (α-amino acid). At physiological pH (~7.36), the amino acids can be subdivided into four classes according to their structure, polarity, and charge state: (1) negatively charged composed of aspartic acid (Asp) and glutamic acid (Glu); (2) positively charged composed of lysine (Lys), arginine (Arg), and histidine (His); (3) polar composed of serine (Ser), threonine (Thr), tyrosine (Tyr), cysteine (Cys), glutamine (Gln), and asparagine (Asn); and (4) nonpolar composed of glycine (Gly), leucine (Leu), isoleucine (Ile), alanine (Ala), valine (Val), proline (Pro), Met, tryptophan (Trp), and phenylalanine (Phe). The carbon to nitrogen peptide bonds are formed through condensation reactions between the carboxyl and amino groups. An example condensation reaction between the amino acids Leu and Tyr is illustrated in Figure 1.19. The peptide C-N bonds are found to be

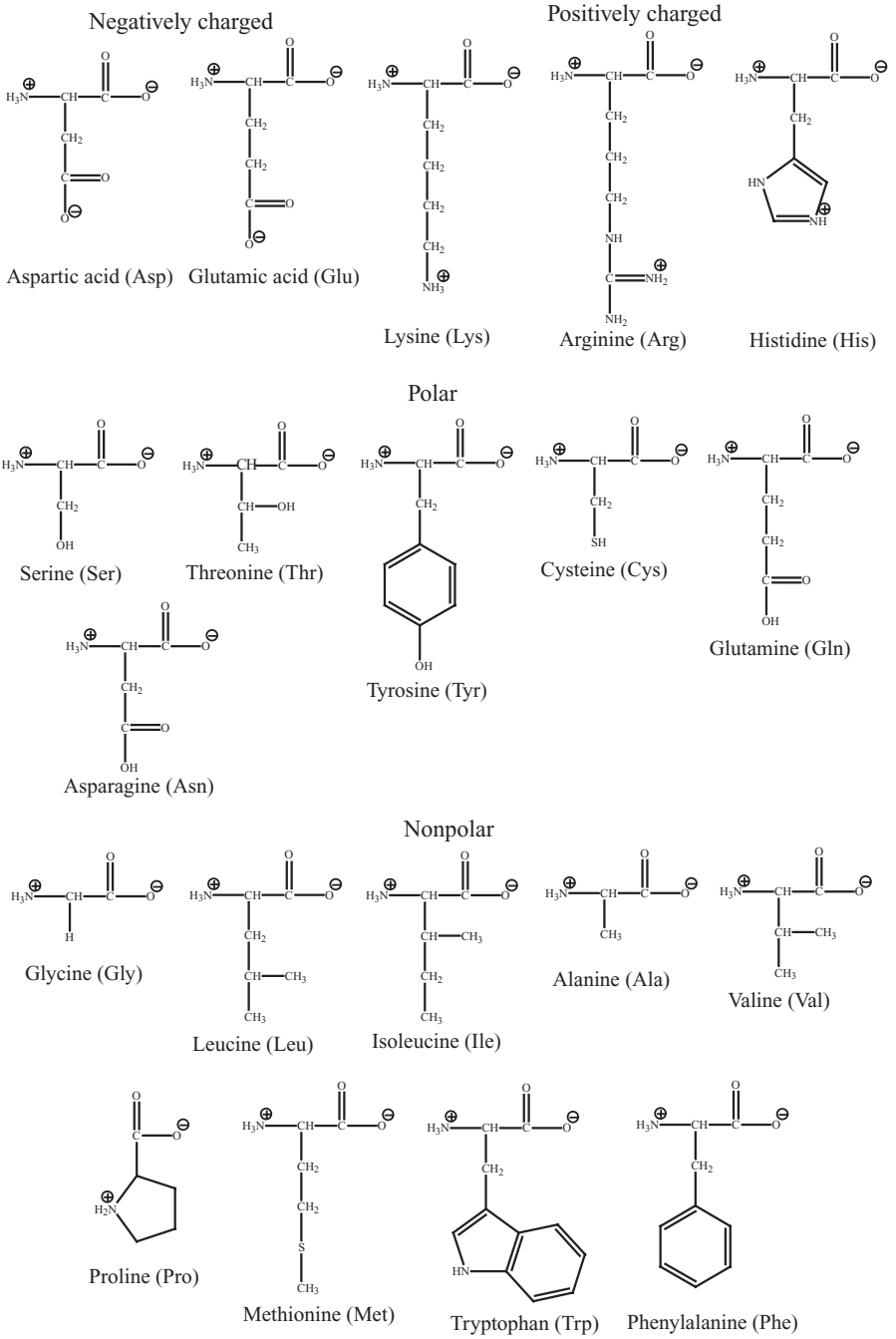

Figure 1.18. Structures of the 20 amino acids that make up the polypeptide backbone of proteins. Divisions include negatively charged, positively charged, polar, and nonpolar.

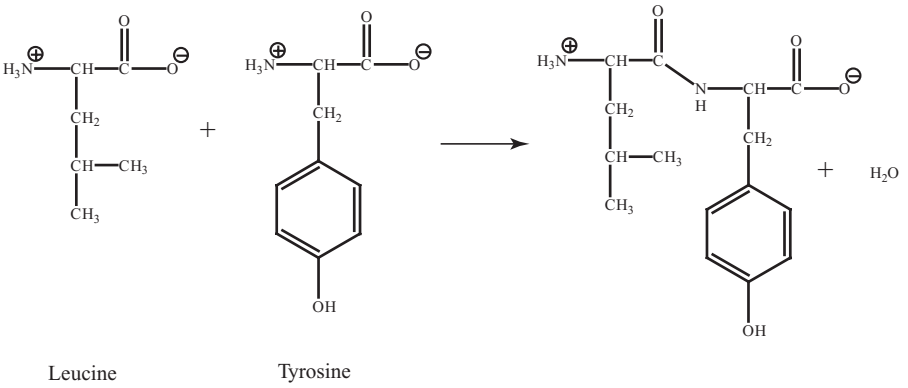

Figure 1.19. Condensation reaction between the amino acids leucine and tyrosine forming a peptide bond.

shorter than most amine C–N bonds due to a double-bond nature that contributes to 40% of the peptide bond.[33] This double-bond character lessons the free rotation of the bond, thus affecting the overall structure of the protein.[34] The secondary structure of the protein is described by two different configurations and turns. The two configurations are α-helices (first proposed by Linus Pauling and Robert B. Corey in 1951) and β-sheets (parallel and antiparallel) and are illustrated in Figure 1.20. The α-helix is described as a right-hand-turned spiral that has hydrogen bonding between oxygen and the hydrogen of the nitrogen atoms of the chain backbone. This hydrogen bonding stabilizes the helical structure. The R-group side chains that make up the amino acid residues extrude out from the helix. The β-sheet is a flat structure that also has hydrogen bonding between oxygen and the hydrogen of the nitrogen atoms but from different β-sheets (parallel and antiparallel) that run along side each other. These hydrogen bonds also work to stabilize the structure. The R-group side chains alternatively extrude out flat with the sheet from the sides of the sheet. The third secondary structure, the turn, basically changes the direction of the polypeptide strand. The tertiary structure, which includes the disulfide bonds, is composed of the ordering of the secondary structure, which is stabilized through side chain interactions. The quaternary structure is the arrangement of the polypeptide chains into the final working protein. All four structures describe what is actually a folded protein, where the apolar regions of the protein are tucked away inside the structure, away from the aqueous medium they are found in naturally, and more polar regions are on the surface.

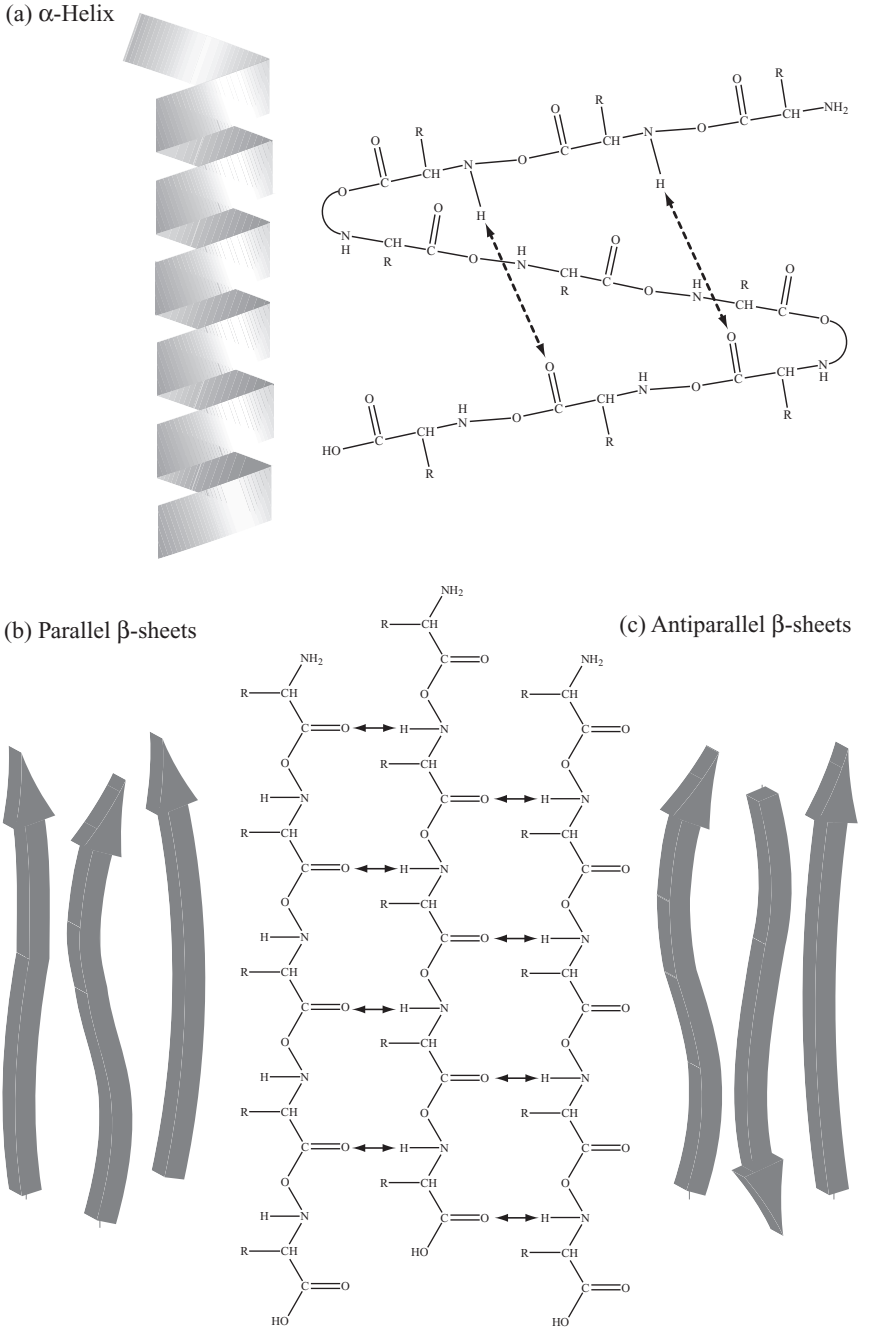

Figure 1.20. (a) α-Helices. β-Sheets: (b) parallel and (c) antiparallel.

1.5.3 Bottom-Up Proteomics: MS of Peptides

1.5.3.1 History and Strategy. The proteomic approach was composed of measuring the enzymatic products of the protein digestion (after protein extraction from the biological sample), namely the peptides, using MS is known as bottom-up proteomics. In the bottom-up approach using nano-ESI-HPLC/MS, the peptides are chromatographically separated and subjected to collision-induced dissociation in the GP. The product ion spectra thus obtained of the separated peptides are then used to identify the proteins present in the biological system being studied. Prior to the use of nano-ESI-HPLC/MS for peptide measurement, Edman degradation was used to sequence unknown proteins. The method of Edman sequencing involves the removal of each amino acid residue one by one from the polypeptide chain starting from the N-terminus of the peptide or protein.[35] The method worked well for highly purified protein samples that contained a free amino N-terminus, but the analysis was slow, usually taking a day to analyze the sequence of one protein. MS was first coupled with Edman sequencing in 1980 by Shimonishi et al.,[36] where the products of the Edman degradation were measured using field desorption (FD) MS. FD, introduced in 1969 by Beckey, is an ionization technique not commonly in use today. FD consists of depositing the sample, either solid or dissolved in solvent, onto a needle and applying a high voltage. The process of desorption and ionization are obtained simultaneously. The analyte ions produced from FD are then introduced into the mass spectrometer for mass analysis. Fast atom bombardment was also used as an ionization technique to measure peptides obtained from the Edman sequencing approach.[37]

Another early approach to proteomics using MS was the application of MALDI time-of-flight (TOF) MS (MALDI-TOF/MS) to the measurement of peptides obtained from in-gel digestions of proteins separated by gel electrophoresis. This technique was reported by several groups and was called peptide mass fingerprinting (PMF).[38–40] In the PMF approach, proteins are first separated using two-dimensional gel electrophoresis (2-DE), a protein separation technique first introduced in the 1970s.[41] The gel used in electrophoresis is a rectangular gel composed of polyacrylamide. The protein sample is loaded onto the gel and the proteins are separated according to their isoelectric point (pH where the protein has a zero charge). This is the first dimension of the separation. The second dimension is a linear separation of the proteins according to their molecular weights. In preparation for sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE), the proteins

are first denatured (usually with 8 M urea and boiling) and sulfide bonds are cleaved effectively unraveling the tertiary and secondary structure of the protein. SDS, which is negatively charged, is then used to coat the protein in a fashion that is proportional to the proteins' molecular weight. The proteins are then separated within a polyacrylamide gel by placing a potential difference across the gel. Due to the potential difference across the gel, the proteins will experience an electrophoretic movement through the gel, thus separating them according to their molecular weight with the lower-molecular-weight proteins having a greater mobility through the gel and the higher molecular weight proteins having a lower mobility through the gel. The resultant 2-DE separation is a collection of spots on the gel that can be up to a few thousand in number. In 2-D SDS-PAGE, the proteins have been essentially separated into single protein spots. This allows the digestion of the protein within the spot (excised from the gel) using a protease with known cleavage specificity into subsequent peptides that are unique to that particular protein. The peptides extracted from the in-gel digested proteins separated by 2-D SDS-PAGE are then measured by MALDI-TOF/MS, creating a spectrum of peaks that represent the molecular weight of the protein's enzymatic generated peptides. This list of measured peptides can be compared with a theoretical list according to the specificity of the enzyme used for digestion. There is an extensive list of references and searching software that has been introduced for the PMF approach to proteomics that has been reviewed.[42] The 2-D SDS-PAGE and peptide mass fingerprint approach to proteomics is illustrated in Figure 1.21.

In bottom-up proteomics, the proteins are generally extracted from the sample of interest, which can include a sample of cultured cells, bacterium, tissue, or a whole organism. A general scheme for the extraction and peptide mass fingerprint mass spectrometric analysis typically followed in early proteomic studies is illustrated in Figure 1.21. The initial sample is lysed and the proteins are extracted and solubilized. The proteins can then be separated using one-dimensional (1-D) or 2-D SDS-PAGE. Proteins can be digested in the gels, or the proteins in solution are digested using a protease such as trypsin. Trypsin is an endopeptidase that cleaves within the polypeptide chain of the protein at the carboxyl side of the basic amino acids Arg and Lys (the trypsin enzyme has optimal activity at a pH range of 7–10 and requires the presence of Ca^{+2}). It has been observed though that trypsin does not efficiently cleave between the residues Lys–Pro and Arg–Pro. Tryptic peptides are predominantly observed as doubly or triply charged when using electrospray as the ionization source. This is due to the amino

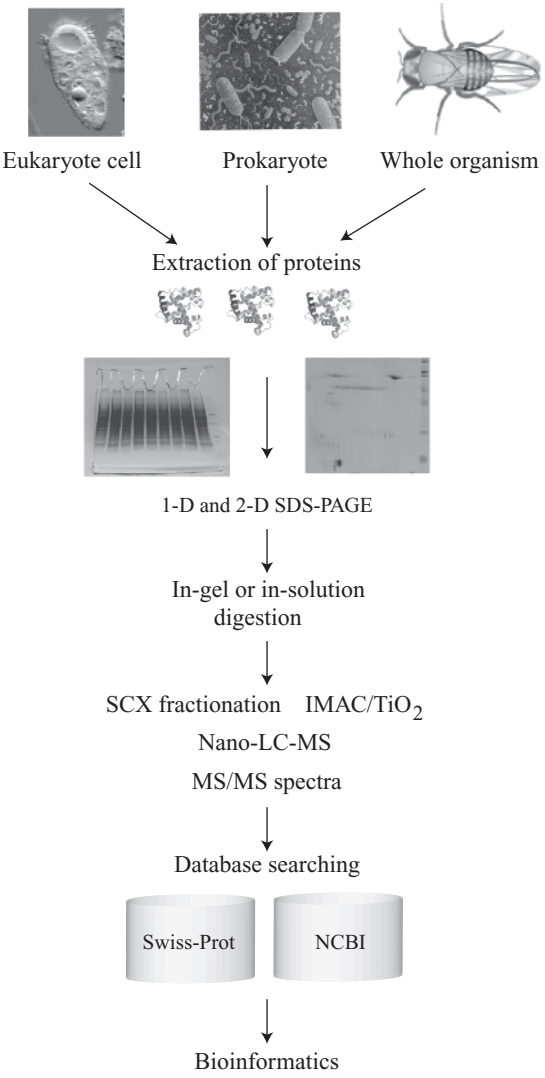

Figure 1.21. General strategy and sample flow involved in proteomics. IMAC, immobilized metal affinity chromatography.

terminal residue being basic in each peptide, except for the C-terminal peptide. There exist a number of proteases that are available to the mass spectrometrist when designing a digestion of proteins into peptides. These can be used to target cleavage at specific amino acid residues within the polypeptide chain. Examples of available proteases and their cleavage specificity are listed in Table 1.1. The enzymes will cleave the proteins into smaller chains of amino acids (typically from five

TABLE 1.1. Examples of Proteases Available for Polypeptide Chain Cleavage

Protease	Polypeptide Cleavage Specificity
Trypsin	At carboxyl side of arginine and lysine residues
Chymotrypsin	At carboxyl side of tryptophan, tyrosine, phenylalanine, leucine, and methionine residues
Proteinase K	At carboxyl side of aromatic, aliphatic, and hydrophobic residues
Factor Xa	At carboxyl side of Glu-Gly-Arg sequence
Carboxypeptidase Y	Sequentially cleaves residues from the carboxy (C) terminus
Submaxillary Arg-C protease	At carboxy side of arginine residues
Staphylococcus aureus V-8 protease	At carboxy side of glutamate and aspartate residues
Aminopeptidase M	Sequentially cleaves residues from the amino (N) terminus
Pepsin	Nonspecifically cleaves at exposed residues favoring the aromatic residues
Ficin	Nonspecifically cleaves at exposed residues favoring the aromatic residues
Papain	Nonspecifically cleaves at exposed residues

amino acid residues up to 100 or so). These short-chain amino acids are mostly water soluble and can be directly analyzed by MS. However, often, a lysis and extract from a biological system will constitute a very complex mixture of proteins that requires some form of separation to decrease the complexity prior to mass spectral measurement.

1.5.3.2 Protein Identification through Product Ion Spectra. More recently, nano-ESI-HPLC-MS/MS has been employed using reversed-phase (RP) C18 columns to initially separate the peptides prior to introduction into the mass spectrometer. If a highly complex compliment of digested proteins are being analyzed such as those obtained from eukaryotic cells or tissue, a greater degree of complexity reduction is employed such as strong cation exchange (SCX) fractionation, which can separate the complex peptide mixture up to 25 fractions or more. The coupling of online SCX with nano-ESI C18 RP HPLC-MS/MS has also been employed and is called 2-D HPLC and multidimensional protein identification technology (MudPIT).[43] This is a gel-free approach that utilizes multiple HPLC-MS analysis of in-solution digestions of protein fractions. The separated peptides are introduced into the mass spectrometer, and product ion spectra are obtained. The

product ions within the spectra are assigned to amino acid sequences. A complete coverage of the amino acid sequence within a peptide from the product ion spectrum is known as *de novo* sequencing. This can unambiguously identify a protein (except for a few anomalies that will be covered shortly) according to standard spectra stored in protein databases. Two examples of protein databases are NCBInr, a protein database composed of a combination of most public databases compiled by the National Center for Biotechnology Information (NCBI), and Swiss-Prot, a database that includes an extensive description of proteins including their functions, PTMs, and domain structures. The correlation of peptide product ion spectra with theoretical peptides was introduced by Yates et al. in 1994.[44] At the same time, Mann et al.[45] proposed a partial sequence error-tolerant database searching for protein identifications from peptide product ion spectra. There exists now a rather large choice of searching algorithms that are available for protein identifications from peptide product ion spectra. A list of identification algorithms and their associated uniform resource locators (URLs) is illustrated in Table 1.2. The final step in the proteomic analysis of a biological system is the interpretation of the identified proteins, which has been called bioinformatics. Bioinformatics attempts to map and decipher interrelationships between observed proteins and the genetic description. Valuable information can be obtained in this way concerning biomarkers for diseased states, the descriptive workings of a biological system, biological interactions, and so on.

In the identification of proteins from peptide collision-induced dissociation using MS is performed to fragment the peptide and identify its amino acid residue sequence. In most mass spectrometers used in proteomic studies such as the ion trap, the quadrupole TOF, the triple quadrupole, and the Fourier transform ion cyclotron resonance (FTICR), the collision energy is considered low (5–50 eV) and the product ions are generally formed through cleavages of the peptide bonds. According to the widely accepted nomenclature of Roepstorff and Fohlman,[46] when the charge is retained on the N-terminal portion of the fragmented peptide, the ions are depicted as *a*, *b*, and *c*. When the charge is retained on the C-terminal portion, the ions are denoted as *x*, *y*, and *z*. The description of the dissociation associated with the peptide chain backbone and the nomenclature of the produced ions is illustrated in Figure 1.22. The ion subscript, for example, the "2" in y_2, indicates the number of residues contained within the ion, two amino acid residues in this case. The weakest bond is between the carboxyl carbon and the nitrogen located directly to the left in the peptide chain. At low-energy collision-induced dissociation of the peptide in MS, the

TABLE 1.2. List of Identification Algorithms

MS identification algorithms and URLs PMF
- Aldente http://www.expasy.org/tools/aldente/
- Mascot http://www.matrixscience.com/search_form_select.html
- MOWSE http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse
- MS-Fit http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm
- PeptIdent http://www.expasy.org/tools/peptident.html
- ProFound http://65.219.84.5/service/prowl/profound.html

MS/MS identification algorithms and URLs PFF
- Phenyx http://www.phenyx-ms.com/
- Sequest http://fields.scripps.edu/sequest/index.html
- Mascot http://www.matrixscience.com/search_form_select.html
- PepFrag http://prowl.rockefeller.edu/prowl/pepfragch.html
- MS-Tag http://prospector.ucsf.edu/ucsfhtml4.0/mstagfd.htm
- ProbID http://projects.systemsbiology.net/probid/
- Sonar http://65.219.84.5/service/prowl/sonar.html
- TANDEM http://www.proteome.ca/opensource.html
- SCOPE N/A
- PEP_PROBE N/A
- VEMS http://www.bio.aau.dk/en/biotechnology/vems.htm
- PEDANTA N/A

De novo sequencing
- SeqMS http://www.protein.osaka-u.ac.jp/rcsfp/profiling/SeqMS.html
- Lutefisk http://www.hairyfatguy.com/Lutefisk
- Sherenga N/A
- PEAKS http://www.bioinformaticssolutions.com/products/peaksoverview.php

Sequence similarity search
- PeptideSearch http://www.narrador.embl-heidelberg.de/GroupPages/Homepage.html
- PepSea http://www.unb.br/cbsp/paginiciais/pepseaseqtag.htm
- MS-Seq http://prospector.ucsf.edu/ucsfhtml4.0/msseq.htm
- MS-Pattern http://prospector.ucsf.edu/ucsfhtml4.0/mspattern.htm
- Mascot http://www.matrixscience.com/search_form_select.html
- FASTS http://www.hgmp.mrc.ac.uk/Registered/Webapp/fasts/
- MS-Blast http://dove.embl-heidelberg.de/Blast2/msblast.html
- OpenSea N/A
- CIDentify http://ftp.virginia.edu/pub/fasta/CIDentify/

Congruence analysis
- MS-Shotgun N/A
- MultiTag N/A

Tag approach
- Popitam http://www.expasy.org/tools/popitam/
- GutenTag http://fields.scripps.edu/GutenTag/index.html

Reprinted with permission of John Wiley & Sons, Hernandez, P., Muller, M., Appel, R.D. Automated protein identification by tandem mass spectrometry: Issues and strategies. *Mass Spectrom. Rev.* 2006, 25, 235–254.

primary breakage will take place at the weakest bond, generally along the peptide backbone chain, and produce *a*, *b*, and *y* fragments. Notice that the *c* ions and the *y* ions contain an extra proton that they have abstracted from the precursor peptide ion. There has also been a proposed third structure for the *b* ion that is formed as a protonated oxazolone, which is suggested to be more stable through cyclization[47] (see b_2 ion in Fig. 1.23). The stability of the *y* ion can be attributed to the

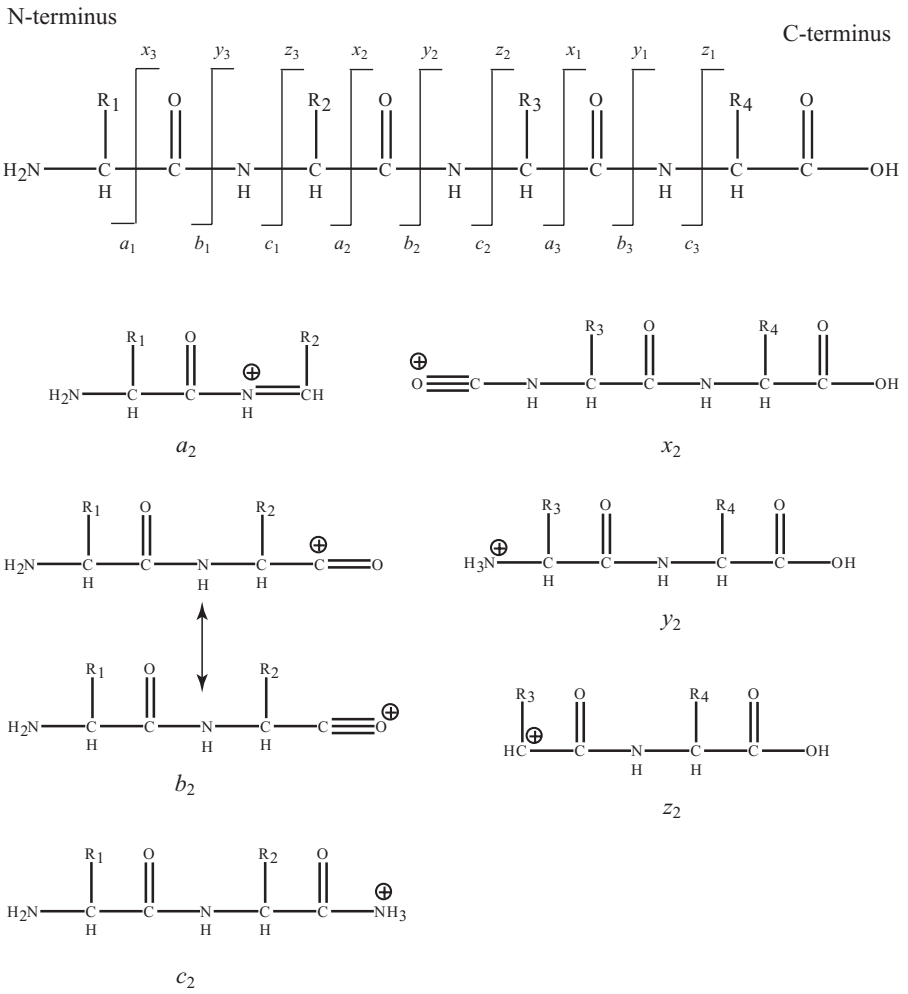

Figure 1.22. Dissociation associated with the peptide chain backbone and the nomenclature of the produced ions. Charge retained on the N-terminal portion of the fragmented peptide the ions are depicted as *a*, *b*, and *c*. Charge retained on the C-terminal portion the ions are denoted as *x*, *y*, and *z*. Ion subscript, for example, "2" in y_2, indicates the number of residues (two) contained within the ion.

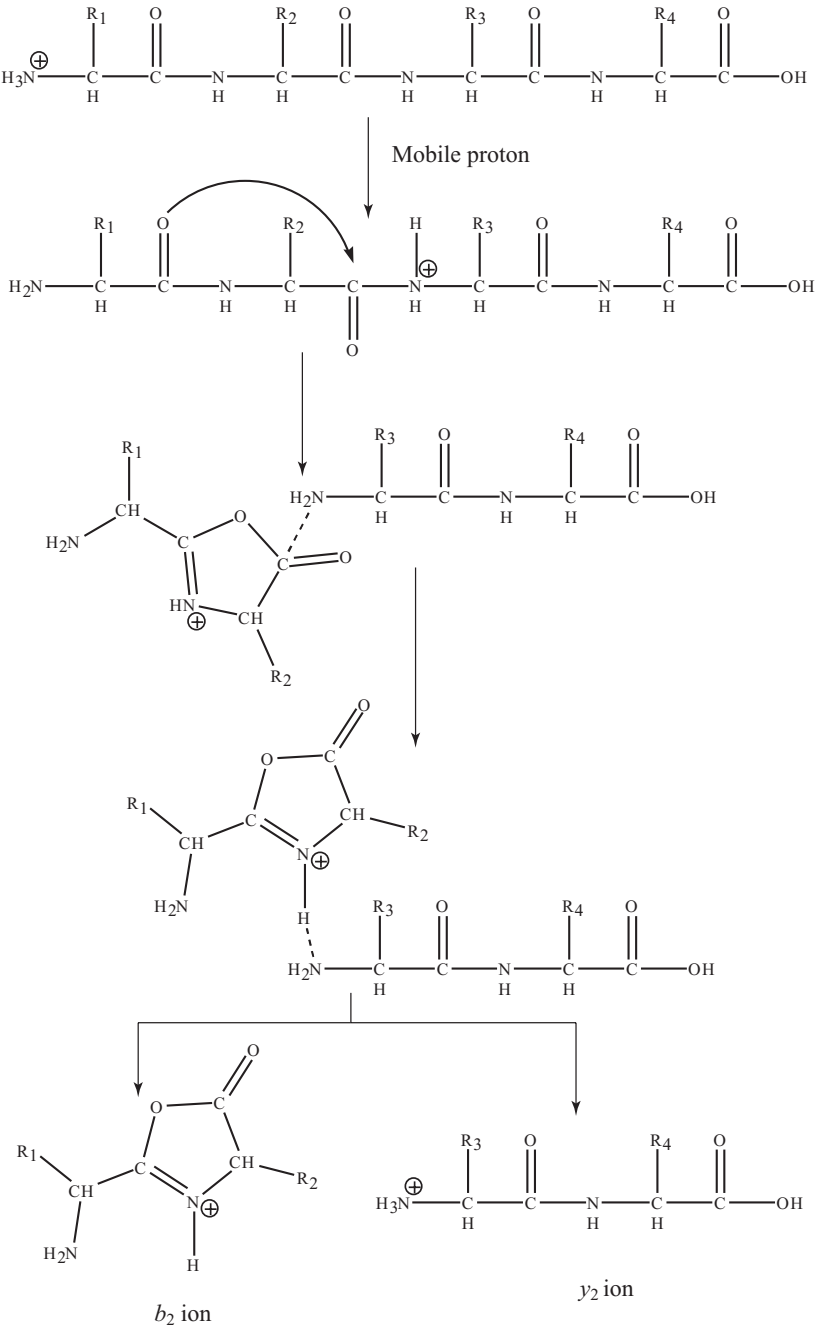

Figure 1.23. Fragmentation pathway leading to the production of the *b* and *y* ions from collision-induced dissociation from the polypeptide backbone chain.

transfer of the proton that is producing the charge state to the terminal nitrogen, thus inducing new bond formation and a lower energy state. The model that describes the dissociation of protonated peptides during low-energy collision-induced excitation is called the "mobile proton" model.[48] Peptides fragment primarily from charge-directed reactions where protonation of the peptide can take place at side chain groups, amide oxygen and nitrogen, and at the terminal amino acid group. On the peptide chain backbone, protonation of the amide nitrogen will lead to a weakening of the amide bond inducing fragmentation at that point. However, it is more thermodynamically favored, as determined by molecular orbital calculations,[48,49] for protonation to take place on the amide oxygen, which also has the effect of strengthening the amide bond. Inspection of peptide product ion fragmentation spectra has demonstrated though that the protonating of the amide nitrogen is taking place over the protonating of the amide oxygen. This is in contrast to the expected site of protonation from a thermodynamic point of view that indicates the amide oxygen protonation and not the amide nitrogen. This discrepancy has been explained by the "mobile proton model," introduced by Wysocki et al.,[48,50] which describes that the proton(s) added to a peptide, upon excitation from CID, will migrate to various protonation sites provided they are not sequestered by a basic amino acid side chain prior to fragmentation. The fragmentation pathway leading to the production of the *b* and *y* ions is illustrated in Figure 1.23. The protonation takes place first on the N-terminus of the peptide. The next step is the mobilization of the proton to the amide nitrogen of the peptide chain backbone where cleavage is to take place. The protonated oxazolone derivative is formed from nucleophilic attack by the oxygen of the adjacent amide bond on the carbon center of the protonated amide bond. Depending on the location of the retention of the charge, either a *b* ion or a *y* ion will be produced.

Besides the amide bond cleavage producing the *b* and *y* ions that are observed in low-energy collision product ion spectra, there are also a number of other product ions that are quite useful in peptide sequence determination. Ions that have lost ammonia (–17 Da) in low-energy collision product ion spectra are denoted as *a**, *b**, and *y**. Ions that have lost water (–18 Da) are denoted as a^{o}, b^{o}, and y^{o}. The *a* ion illustrated in Figure 1.22 is produced through loss of CO from a *b* ion (–28 Da). Upon careful inspection of the structures in Figure 1.22 for the product ions, it can be seen that the *a* ion is missing CO as compared with the structure of the *b* ion. When a difference of 28 is observed in product ion spectra between two *m*/*z* values, an *a*–*b* ion pair is suggested and can be useful in ion series identification. Internal cleavage

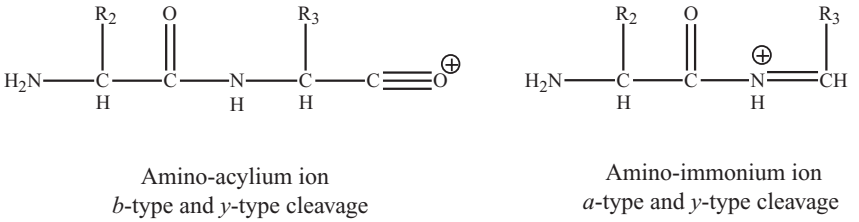

Figure 1.24. Structure of (left) an amino-acylium ion produced through a combination of *b*- and *y*-type cleavage and (right) an amino-immonium ion through a combination of *a*- and *y*-type internal cleavage.

ions are produced by double backbone cleavage, usually by a combination of *b*- and *y*-type cleavage. When a combination of *b*- and *y*-type cleavage takes place, an amino-acylium ion is produced. When a combination of *a*- and *y*-type internal cleavage takes place, an amino-immonium ion is produced. The structures of an amino-acylium ion and an amino-immonium ion are illustrated in Figure 1.24. These types of product ions that are produced from internal fragmentation are denoted with their one-letter amino acid code. Though not often observed, *x*-type ions can be produced using photodissociation.

1.5.3.3 High-Energy Product Ions. Thus far, the product ions that have been discussed, the *a*-, *b*-, and *y*-type ions, are produced through low-energy collisions such as those observed in ion traps. The collision-induced activation in ion traps is a slow heating mechanism, produced through multiple collisions with the trap bath gas, which favors lower-energy fragmentation pathways. High-energy collisions that are in the kiloelectron volt range such as those produced in MALDI TOF-TOF MS produce other product ions in addition to the types that have been discussed so far. Side-chain cleavage ions that are produced by a combination of backbone cleavage and a side-chain bond are observed in high-energy collisions and are denoted as *d*, *v*, and *w* ions. Figure 1.25 contains some illustrative structures of *d*-, *v*-, and *w*-type ions.

Immonium ions are produced through a combination of *a*-type and *y*-type cleavage that results in an internal fragment that contains a single side chain. These ions are designated by the one-letter code that corresponds to the amino acid. Immonium ions are not generally observed in ion trap product ion mass spectra but are in MALDI TOF-TOF product ion mass spectra. The structure of a general immonium ion is illustrated in Figure 1.26. Immonium ions are useful in acting as confirmation of residues suspected to be contained within the peptide backbone. Table 1.3 is a compilation of the amino acid residue

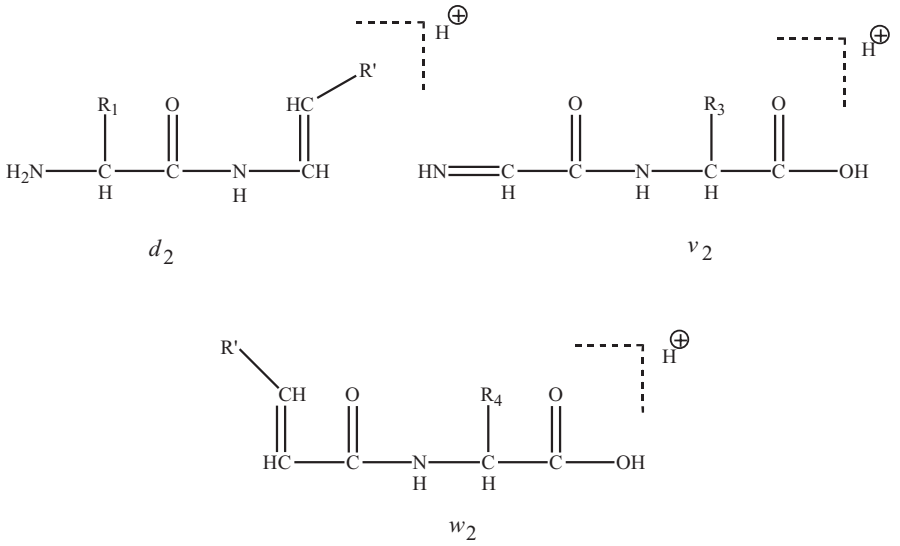

Figure 1.25. Structures of *d*-, *v*-, and *w*-type ions produced by a combination of backbone cleavage and a side chain bond observed in high-energy collision product ion spectra.

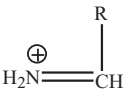

Figure 1.26. Structure of a general immonium ion.

information that is used in MS analysis of peptides. The table includes the amino acid residue's name, associated codes, residue mass, and immonium ion mass.

1.5.3.4 **De Novo *Sequencing.*** An example of *de novo* sequencing is illustrated in Figure 1.27. The product ion spectrum in Figure 1.27a is for a peptide composed of seven amino acid residues. The peptide product ion spectrum in Figure 1.27b is also composed of seven amino acid residues; however, the Ser residue (Ser, $C_3H_5NO_2$, 87.0320 amu) in Figure 1.27a has been replaced by a Thr residue (Thr, $C_4H_7NO_2$, 101.0477 amu) in Figure 1.27b. The product ion spectra are very similar, but a difference can be discerned with the b_5 ion and the y_3 ions where a shift of 14 Da is observed due to the difference in amino acid residue composition associated with Ser and Thr.

Though the sequencing of the amino acids contained within a peptide chain can be discerned by *de novo* MS as just illustrated, there is a

TABLE 1.3. Amino Acid Residue Names, Codes, Masses, and Immonium Ion *m/z* Values

Residue	One-Letter Code	Three-Letter Code	Residue Mass	Immonium Ion (m/z)
Alanine	A	Ala	71.04	
Arginine	R	Arg	156.10	129
Asparagine	N	Asn	114.04	87.09
Aspartic acid	D	Asp	115.03	88.04
Cysteine	C	Cys	103.01	76
Glutamic acid	E	Glu	129.04	102.06
Glutamine	Q	Gln	128.06	101.11
Glycine	G	Gly	57.02	30
Histidine	H	His	137.06	110.07
Isoleucine	I	Ile	113.08	86.1
Leucine	L	Leu	113.08	86.1
Lysine	K	Lys	128.09	101.11
Methionine	M	Met	131.04	104.05
Phenylalanine	F	Phe	147.07	120.08
Proline	P	Pro	97.05	70.07
Serine	S	Ser	87.03	60.04
Threonine	T	Thr	101.05	74.06
Tryptophan	W	Trp	186.08	159.09
Tyrosine	Y	Tyr	163.06	136.08
Valine	V	Val	99.07	72.08

problem associated with isomers and isobars. Isomers are species that have the same molecular formula but differ in their structural arrangement, while isobars are species with different molecular formulas that possess similar (or the same) molecular weights. For example, it is not possible to determine whether a particular peptide contains Leu (Leu, $C_6H_{11}NO$, 113.0841 amu) or its isomer Ile (Ile, $C_6H_{11}NO$, 113.0841 amu) both at a residue mass of 113.0841 amu. Furthermore, even though the remaining 18 amino acid residues each contain distinctive elemental compositions and thus distinct molecular masses, some combinations of residues will actually equate to identical elemental compositions. This produces an isobaric situation where different peptides will possess either very similar or identical sequence masses. If every single peptide amide bond cleavage is not represented within the product ion spectrum, then it is not possible to discern some of these possible combinations. The use of high-resolution/high-mass accuracy instrumentation such as the FTICR mass spectrometer or the Orbitrap can be used to

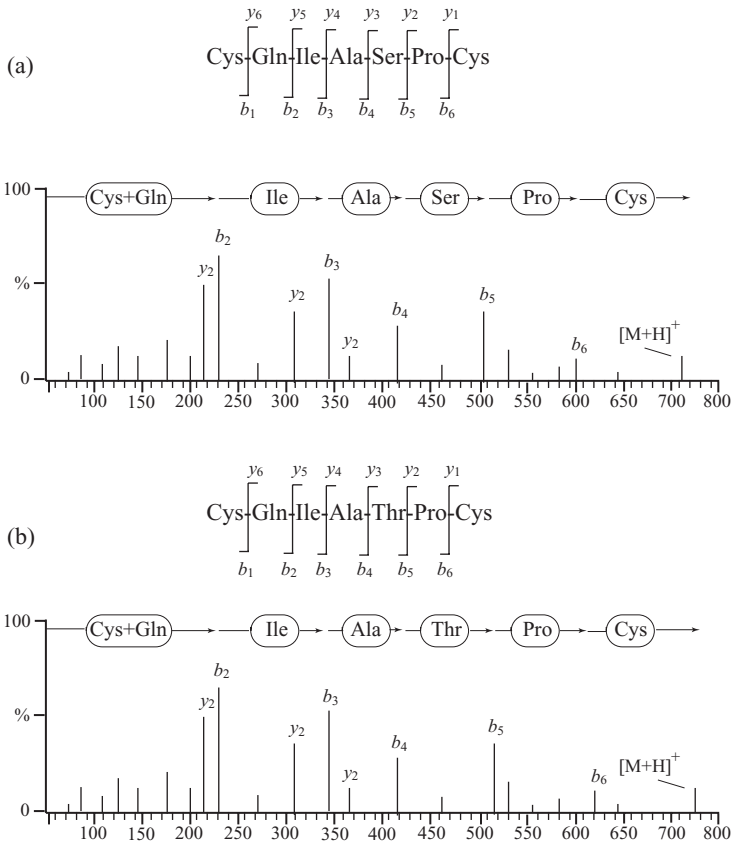

Figure 1.27. Example of *de novo* sequencing using product ion spectra collected by collision-induced dissociation mass spectrometry. (a) Peptide composed of seven amino acid residues. (b) Peptide composed of seven amino acid residues with the serine residue (Ser, $C_3H_5NO_2$, 87.0320 amu) replaced by a threonine residue (Thr, $C_4H_7NO_2$, 101.0477 amu). The product ion spectra are very similar, but a difference can be discerned with the b_5 ion and the y_3 ions where a shift of 14 Da is observed due to the difference in amino acid residue composition associated with serine and threonine.

help reduce this problem when complete *de novo* sequencing is not possible. Table 1.4 is a listing of some of the amino acid combinations that may arise that can contribute to unknown sequence determination when complete *de novo* sequencing is not being obtained. For a peptide with a mass of 800 Da, the differences in the table for, for example, the Gln versus Lys difference at 0.03638 Da would take a mass accuracy of better than 44 ppm to distinguish the two. For the Arg versus Gly + Val at 0.01124 Da, it would require a mass accuracy of 14 ppm to distinguish the two. For the FTICR mass spectrometers and the hybrid mass spectrometers such as the linear ion trap-Fourier transform (LTQ-FT)

TABLE 1.4. Examples of Combinations of Amino Acid Residues Where Isobaric Peptides Can Be Observed

Amino Acid Residue	Residue Mass (Da)	Δ Mass (Da)
Leucine	113.08406	
Isoleucine	113.08406	0
Glutamine	128.05858	
Glycine + alanine	128.05858 (57.02146 + 71.03711)	0
Asparagine	114.04293	
2 × glycine	114.04293 (2 × 57.02146)	0
Oxidized methionine	147.03540	
Phenylalanine	147.06841	0.03301
Glutamine	128.05858	
Lysine	128.09496	0.03638
Arginine	156.10111	
Glycine + valine	156.08987 (57.02146 + 99.06841)	0.01124
Asparagine	114.04293	
Ornithine	114.07931	0.03638
Leucine/isoleucine	113.08406	
Hydroxyproline	113.04768	0.03638
2 × valine	198.13682 (2 × 99.06841)	
Proline + threonine	198.10044 (97.05276 + 101.04768)	0.03638

or the LTQ-Orbitrap, this is readily achievable, but often, for ion traps this is not always achievable.

1.5.3.5 Electron Capture Dissociation (ECD). Other techniques such as ECD[51] and electron transfer dissociation (ETD) have also been used to alleviate the problem of isobaric amino acid combinations by giving complimentary product ions (such as *c* and *z* ions) that help to obtain complete sequence coverage. The technique of ECD tends to promote extensive fragmentation along the polypeptide backbone, producing *c*- and *z*-type ions while also preserving modifications such as glycosylation and phosphorylation. The general *z*-type ion that is shown in Figure 1.22 is different though from the *z*-type ion that is produced in ECD, which is a radical cation. Peptide cation-radicals are produced by passing or exposing the peptides, which are already multiply-protonated by ESI through low-energy electrons. The mixing of the protonated peptides with the low-energy electrons will result in exothermic ion–electron recombinations. There are a number of dissociations that can take place after the initial peptide cation-radical is formed. These include loss of ammonia, loss of H atoms, loss of side chain fragments, cleavage of disulfide bonds, and most importantly, peptide backbone cleavages. The *c*-type ion is produced through homolytic cleavage

at the N–C peptide bond, and charges are present in the amino-terminal fragment. The *z*-type ion is produced when charges are present in the carboxy-terminal fragment. The mechanism that has been given for the promotion of fragmentation of the peptides is due to electron attachment to the protonated sites of the peptide. The now cation radical intermediate that has formed will release a hydrogen atom. A nearby carbonyl group will capture the released hydrogen atom and the peptide will dissociate by cleavage of the adjacent N–C peptide bond. The mechanism for the production of an α-amide radical of the peptide C-terminus, a *z*-type ion, and the enolamine of the N-terminus portion of the peptide, a *c*-type ion, is illustrated in Figure 1.28.

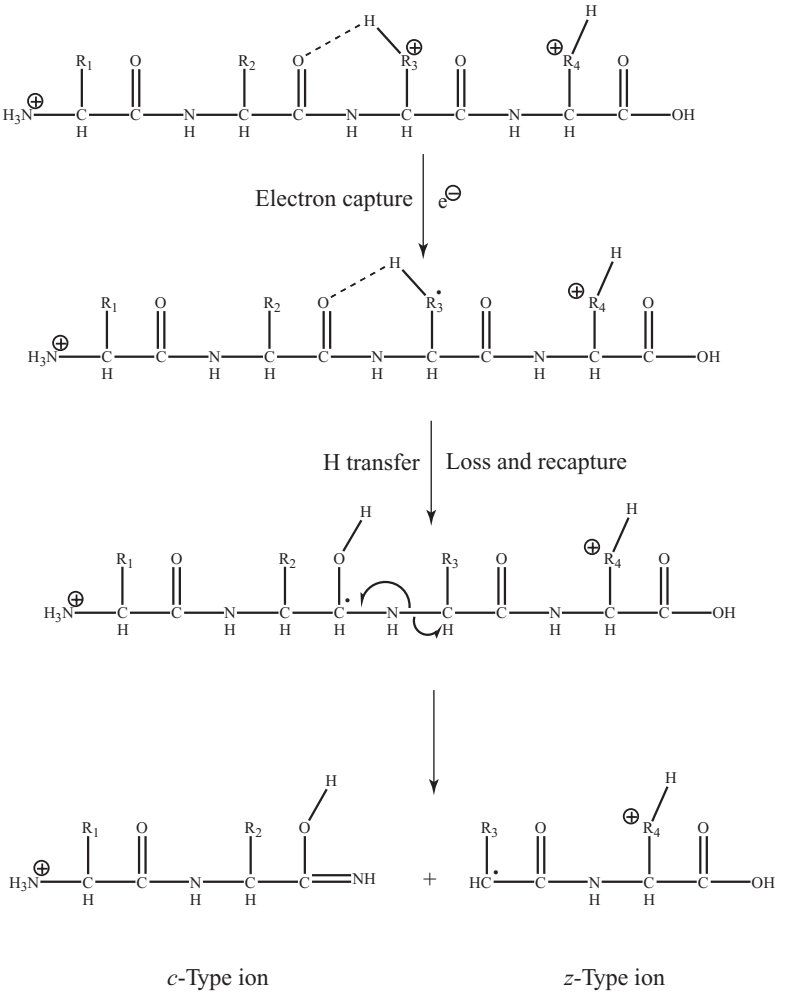

Figure 1.28. Mechanism for the production of *c*- and *z*-type ions observed in ECD.

1.5.4 Top-Down Proteomics: MS of Intact Proteins

1.5.4.1 Background. Measuring the whole, intact protein in the GP using mass spectrometric methodologies is known as "top-down" proteomics. Top-down proteomics measures the intact protein's mass followed by collision-induced dissociation of the whole protein, breaking it into smaller parts. A vital component of top-down proteomics is the accuracy in which the masses are measured. Often, high-resolution mass spectrometers such as the FTICR mass spectrometer are used to accurately measure the intact protein's mass and the product ions produced during collision-induced dissociation experiments. In early top-down experiments, though, this was not the case. Mass spectrometers such as the triple quadrupole coupled with electrospray were first used to measure intact proteins in the GP.[52,53] However, the triple quadrupole mass spectrometer does not allow the resolving of the isotopic distribution of the product ions being generated in the top-down approach. The use of FT-MS/MS was later reported with high enough resolution to resolve isotopic peaks.[54,55] An example of these early top-down experiments utilizing FT-MS is illustrated in Figure 1.29. Extensive initial, premass analysis sample preparation, such as cleanup, digestion, desalting, and enriching, all often incorporated in "bottom-up" proteomics, is not necessarily required in top-down approaches. The dynamic range in top-down proteomics can be limited by the number of analytes that can be present during analysis, but this is usually overcome by using some type of separation prior to introduction into the mass spectrometer. The separation of complex protein mixtures can be obtained using techniques such as RP-HPLC, gel electrophoresis, anion exchange chromatography, and capillary electrophoresis. Typically in bottom-up analysis, the digested protein peptides are <3 kDa and the complete description of the original, intact protein is not possible. With top-down analysis, there is often 100% coverage of the protein being analyzed. This allows the determination of the N- and C-termini, the exact location of modifications to the protein such as phosphorylation, and the confirmation of DNA-predicted sequences.

1.5.4.2 GP Basicity and Protein Charging. It is the process of ESI that allows the measurement of intact proteins with large molecular weight. As a rule of thumb, for each 1000 Da of the protein, there is associated one charge state. For example, a 30 kDa protein, as an approximation, will have a charge state of 30+. This brings the measured mass of the protein down into the range of many mass

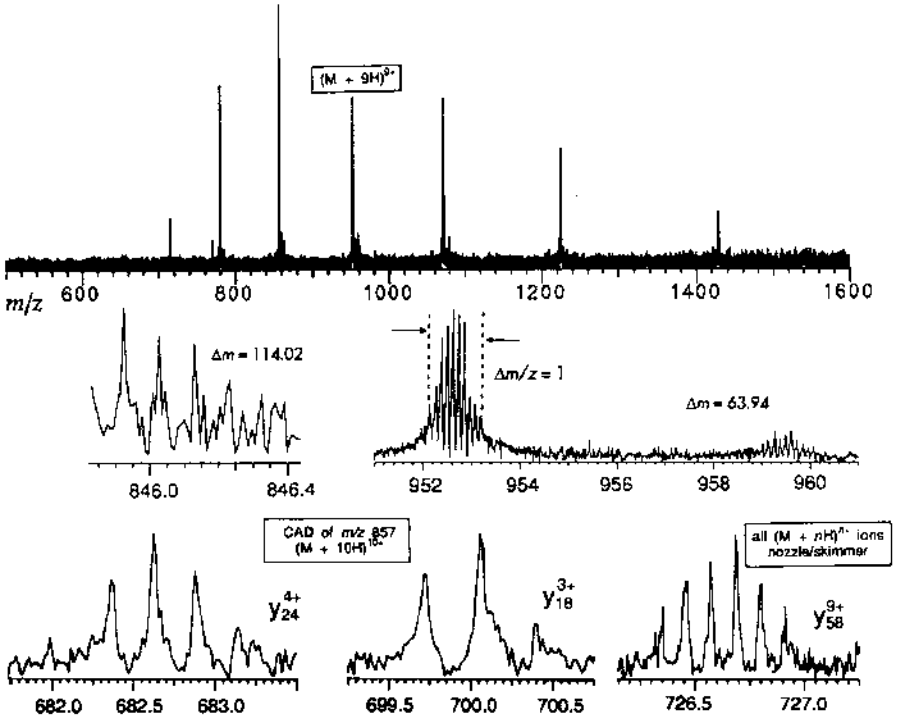

Figure 1.29. (Top) ESI mass spectrum of ubiquitin (sum of 10 scans, 64,000 data). (Center) Regions expanded to show the presence of impurities. (Bottom) MS/MS fragment ions from collisionally activated dissociation of $[M + 10H]^{10+}$ and from placing 200 V between the nozzle and skimmer of the ESI source. (Reprinted with permission from Loo, J.A., Quinn, J.P., Ryu, S.I., Henry, K.D., Senko, M.W., McLafferty, F.W. High-resolution tandem mass spectrometry of large biomolecules (electrospray ionization/polypeptide sequencing), *Proc. Natl. Acad. Sci. U.S.A.* 1992, 89, 286–289.)

spectrometers that typically scan between m/z 100 and m/z 4000 (m/z = 30 kDa/30+ = 1000 Th). Multiple charging of peptides and proteins is achieved during the ESI process due to the presence of amino acid residue basic sites. There is a limit to the number of charges that can be placed onto a peptide or protein during the ESI process as was demonstrated by Schnier et al. in 1995.[56] In their study, the apparent GP basicity as a function of charge state was measured using cytochrome c. A graphical plot of apparent GP basicity versus charge state is illustrated in Figure 1.30. The curves in Figure 1.30 have a negative trend (go down) as each charge state is increased. As a new charge state is added (increasing x-axis), the apparent GP basicity decreases

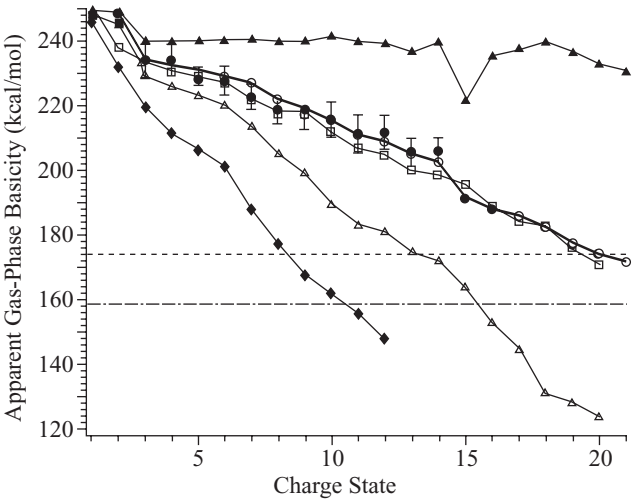

Figure 1.30. Apparent gas-phase basicity as a function of charge state of cytochrome c ions, measured (•); calculated, linear ($\epsilon_r = 1.0$, Δ; best fit $\epsilon_r = 2.0$, O); intrinsic, ▲; calculated, X-ray crystal structure ($\epsilon_r = 2.0$, ◆); calculated, α-helix ($\epsilon_r = 4.1$, □).The dashed line indicates gas-phase basicity (GB) of methanol (174.1 kcal/mol) and the dash-dot line indicates GB of water (159.0 kcal/mol). (Reprinted with permission from Schnier, D.S.; Gross, E.R.; Williams, E.R. *J. Am. Chem. Soc.* 1995, 117, 6747–6757. Copyright 1995 American Chemical Society.)

(*y*-axis). In the graph, the dashed line represents the GP basicity of methanol, which is included as a reference of a species present during the ESI process that can also accept a charge. The charging of large molecules during the ionization process is thought to follow the charged residue mechanism. In this ionization process, the solvent molecules evaporate from around the protein, leaving charges that associate to basic sites within the protein. What the intersection in the graph between the charging of cytochrome c and the GP basicity of methanol is demonstrating is that there is a limit to the amount of charges that can be placed on a species during ionization. At some point (intersection), it becomes thermodynamically favorable to put the next charge (proton) onto a methanol molecule than onto the protein. Here, a state has been reached where Coulombic repulsion between the charges and a loss of basic sites does not allow further charging. At this point, the maximum charge state has been reached for the protein molecule.

1.5.4.3 Calculation of Charge State and Molecular Weight. Another interesting feature of the electrospray charging of proteins is the ability to calculate the charge state and molecular weight of an unknown

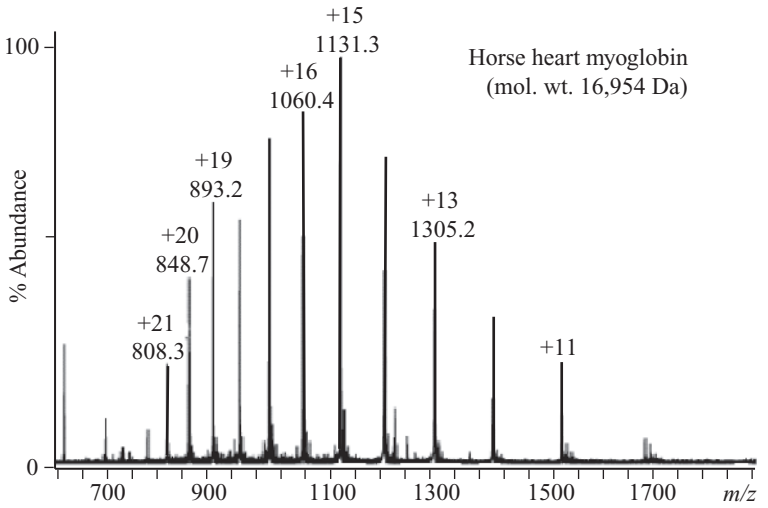

Figure 1.31. Electrospray mass spectrum of horse heart myoglobin at a molecular weight of 16,954 Da. The spectrum illustrates an envelope of peaks of different m/z values and charge states for the protein.

protein from its single-stage mass spectrum. Figure 1.31 illustrates the electrospray mass spectrum of horse heart myoglobin at a molecular weight of 16,954 Da. If this were an unknown protein species, we would only have the respective m/z values from the mass spectrum. Using the following two simultaneous equations with two unknowns, the charge states and the molecular weight of the unknown protein can be calculated using only the information obtained within the mass spectrum:

$$m/z = \frac{m + z(1.0079)}{z} \text{ higher } m/z \text{ peak from spectrum,} \tag{1.1}$$

$$m/z = \frac{m + (z+1)(1.0079)}{z+1} \text{ lower } m/z \text{ peak from spectrum.} \tag{1.2}$$

If we solve the higher mass/lower charge state peak (m/z = 1131.3 Th) for m by taking Equation 1.1 and solving for m, we obtain the following:

$$m = 1131.3z - 2.0079z, \tag{1.3}$$

$$m = 1130.2921z. \tag{1.4}$$

If we next substitute this mass into the lower mass/higher charge state (m/z = 1060.4 Th) in Equation 1.2, we can calculate the charge state of the m/z 1131.3 peak:

$$m/z = \frac{m + (z+1)(1.0079)}{z+1},$$

$$1060.4 = \frac{m + 1.0079z + 1.0079}{z+1}, \tag{1.5}$$

$$1060.4 = \frac{1130.2921m + 1.0079z + 1.0079}{z+1}, \tag{1.6}$$

$$z = 15. \tag{1.7}$$

Therefore, we have determined that the charge state of the m/z 1131.3 peak is +15. To calculate the molecular weight of the unknown protein, we can substitute the charge state value into Equation 1.1 and solve for m:

$$m/z = \frac{m + z(1.0079)}{z},$$

$$1131.3 = \frac{m + 15(1.0079)}{15}, \tag{1.8}$$

$$m = 16954. \tag{1.9}$$

This process is known as deconvolution where a distribution (mass spectral envelope) of a protein's m/z values and associated charge states are collapsed down to a single peak representing the molecular weight of the protein. The deconvolution of the horse heart myoglobin protein is illustrated in Figure 1.32. The width of the deconvoluted peak indicates the variability in the calculation of the molecular weight of the protein from the mass spectral peaks. While it is possible to deconvolute protein peaks by hand by solving two simultaneous equations with two unknowns, mass spectral computer software is typically used to perform this task.

1.5.4.4 Top-Down Protein Sequencing. In bottom-up proteomics, the mass spectral identification of proteins through sequence determination of separated peptides requires the isolation of a single peptide for fragmentation experiments. This is typically done by removing from an ion trap mass spectrometer all m/z species present except for one that

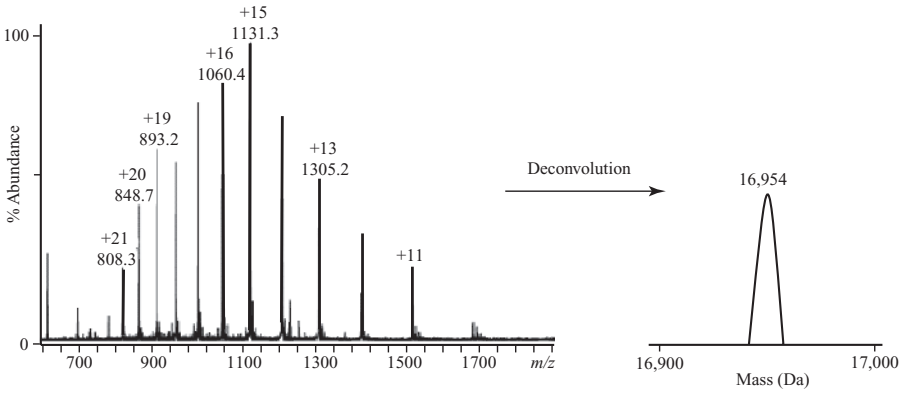

Figure 1.32. Deconvoluted, computer-generated spectrum of horse heart myoglobin at a molecular weight of 16,954 Da.

is of interest for fragmentation and subsequent sequencing. It has been demonstrated though that multiple proteins can be simultaneously fragmented and identified in top-down proteomics. In a study reported by Patrie et al.,[57] the authors used a hybrid mass spectrometer that coupled a quadrupole mass analyzer to an FTICR mass spectrometer that uses a 9.4 Tesla magnet. The instrumental design is illustrated in Figure 1.33. Prior to the FTICR-MS, there is a resolving quadrupole that can act as either a radio frequency (RF)-only ion guide or as a fully functional mass analyzer. Following this in the instrumental design is an accumulation octopole that was used to accumulate and store ions prior to introduction into the FTICR mass spectrometer. Nitrogen or helium gas at a pressure of approximately 1 mtorr was introduced into the accumulation octopole to help improve the accumulation. The FTICR cell located within the 9.4 Tesla magnet is an open-ended capacity coupled cell that is cylindrical and divided axially into five segments. At the end of the instrument (far right side) is a laser that is used for infrared multiphoton dissociation (IRMPD) experiments. For fragmentation experiments, collision-induced dissociation could be performed in the accumulation octopole, IRMPD could be performed within the ion cyclotron resonance (ICR) cell by irradiating the trapped species with the laser, or finally, the instrumental design also included ECD capabilities. A top-down experiment of a mixture of proteins collected as a fraction eluting from a reversed-phase liquid chromatography (RPLC) separation is illustrated in Figure 1.34. Figure 1.34a illustrates a broadband spectrum of the RPLC fraction where a very low response is observed for the proteins present. The same broadband spectrum is illustrated in Figure

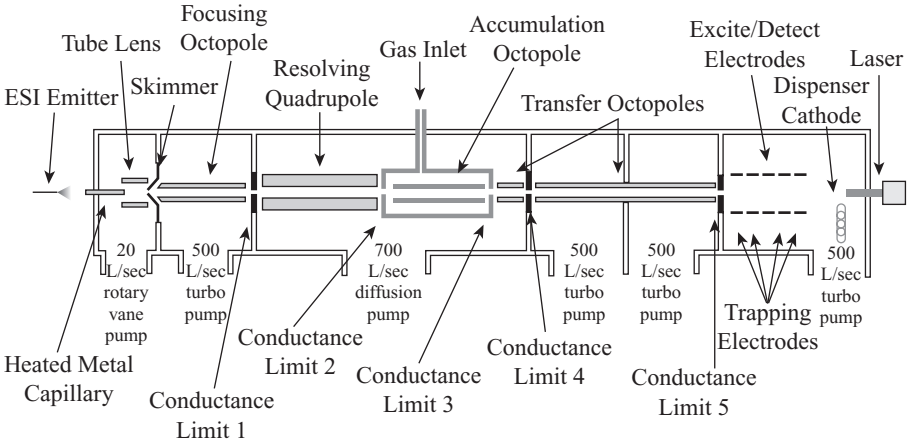

Figure 1.33. Schematic representation of the quadrupole/Fourier transform ion cyclotron resonance hybrid mass spectrometer for versatile MS/MS and improved dynamic range by means of *m/z*-selective ion accumulation external to the superconducting magnet bore. (Reprinted with permission. This article was published in *J Am Soc Mass Spectrom*, Patrie, S.M., Charlebois, J.P., Whipple, D., Kelleher, N.L., Hendrickson, C.L., Quinn, J.P., Marshall, A.G., Mukhopadhyay, B. Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa, 2004, 15, 1099–1108. Copyright Elsevier 2004.)

1.34b after using the accumulation octopole to increase the amount of sample that is being introduced into the FTICR cell and thus increase the sensitivity of the mass measurement of the seven proteins present. Figure 1.34c, d illustrates the fragmentation result of an IRMPD experiment of the seven proteins that were present. Of the seven proteins present in Figure 1.34b, three were identified by top-down proteomics, listed as X, the 19,431.8 Da protein MJ0543; O, the 20,511.3 Da protein MJ0471; and Z, the 17,263.0 Da protein MJ0472. Notice the large number of amino acid residues contained within the *b*- and *y*-type ions in Figure 1.34d and the associated high number of charges (e.g., the Oy_{63}^{8+} product ion that contains 63 amino acid residues and eight charges). This is quite different from the peptides that are normally observed in bottom-up proteomics where most peptides contain between 7 and 25 amino acid residues with mostly two charges, but three or four charges are also observed for the longer chain peptides.

1.5.5 Systems Biology and Bioinformatics

The application and use of MS as an analytical tool to the field of biology is obviously apparent. The amount of information given from

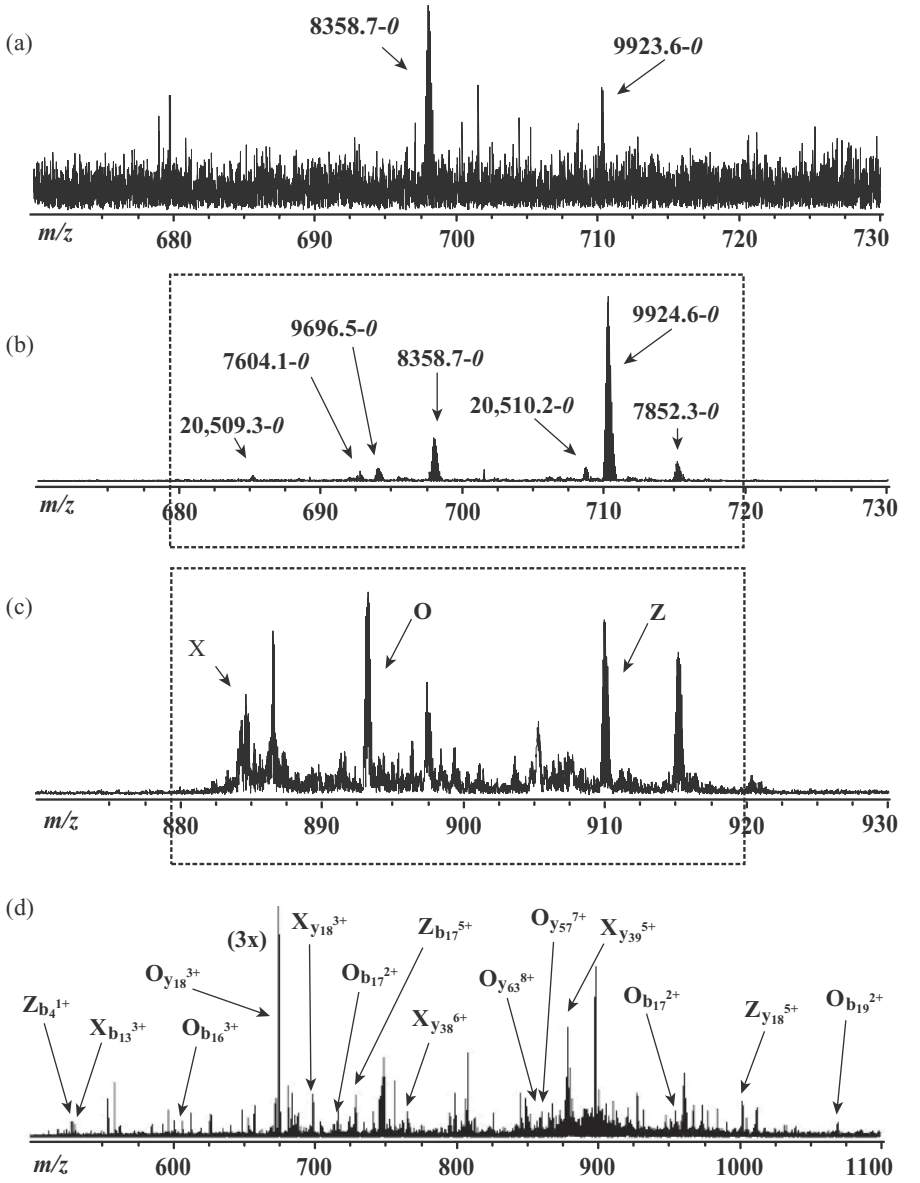

Figure 1.34. (a) Expansion of a 60 m/z section segment of the broadband spectrum from a *Methanococcus jannaschii* RPLC fraction (3.7-second scan time; 50 scans). (b) The same $\Delta(m/z) = 60$ segments after quadrupole-enhanced ion accumulation ($\Delta[m/z] = 40$ segment, 9.7-second scan time; 10 scans). (c) $\Delta(m/z) = 60$ segments from the same sample with subsequent IRMPD fragmentation of all intact proteins in parallel (shown in [d], 25 scans). Identified proteins are MJ0543 (X, Exptl. 19,431.8-*0* kDa, Theoret. 19,432.5-*0* kDa), MJ0471 (O, Exptl. 20,511.3-*0* kDa, Theoret. 20,511.1-*0*), and MJ0472 (asterisk, Exptl. 17.263.0-*0* kDa, Theoret. 17,263.6-*0* kDa). (Reprinted with permission. This article was published in *J Am Soc Mass Spectrom*, Patrie, S.M., Charlebois, J.P., Whipple, D., Kelleher, N.L., Hendrickson, C.L., Quinn, J.P., Marshall, A.G., Mukhopadhyay, B. Construction of a hybrid quadrupole/fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa, 2004, 15, 1099–1108. Copyright Elsevier 2004.)

mass spectrometric methodologies, that is, the molecular weight, the structure, and the amount (can be relative and/or absolute) of a particular biomolecule extracted from a biological matrix, is similar to the biological analytical approach traditionally used of "one gene" or "one protein" at a time study. The study of biological systems is now moving toward more encompassing analysis such as the sequencing of an organism's genome or proteome. The trend now is to study both the single components of a system in conjunction with the particular system's entire compliment of components. Of special interest is how the various components of the system interact with one another under normal conditions and some type of perturbed condition such as a diseased state or a change in the systems environment (e.g., lack of oxygen, food, water). Thus, systems biology is the study of the processes and complex biological organizational behavior using information from its molecular constituents.[58] This is quite broad in the sense that the biological organization may go to the level of tissue up to a population or even an ecosystem. The hierarchical levels of biological information were summed up by L. Hood at the Institute for Systems Biology (Seattle, WA) showing the progression from DNA to a complex organization as illustrated in Figure 1.35. The idea is to gather as much information about each component of a system to more accurately describe the system as a whole. This encompasses most of the disciplines of biology and incorporates analytical chemistry (MS) through the need to measure and identify individual species on the molecular level. When attempting to describe the behavior of a biological system, often, the reality of the system is that the behavior of the whole system is greater than what would be predicted from the sum of its parts.[59] The study of systems biology incorporates multiple analytical techniques and methodologies, which include MS, to study the components of a biological system (e.g., genes, proteins, metabolites) to better model their interactions.[60]

The experimental workflow involved in systems biology that center on mass spectrometric analysis is illustrated in Figure 1.36. Specifically, the metabolite analysis and the protein analysis, illustrated in the second step of the flow, are increasingly performed using highly efficient separation methodology such as nano-HPLC (1-D and 2-D) coupled with high-resolution MS such as FTICR-MS and LTQ-Orbitrap-MS. Much development has been done, and is still ongoing, in the processing and data extraction of the information obtained from high-peak capacity nano-HPLC-ESI MS. A tremendous amount of information is obtained from the experiments that need to be

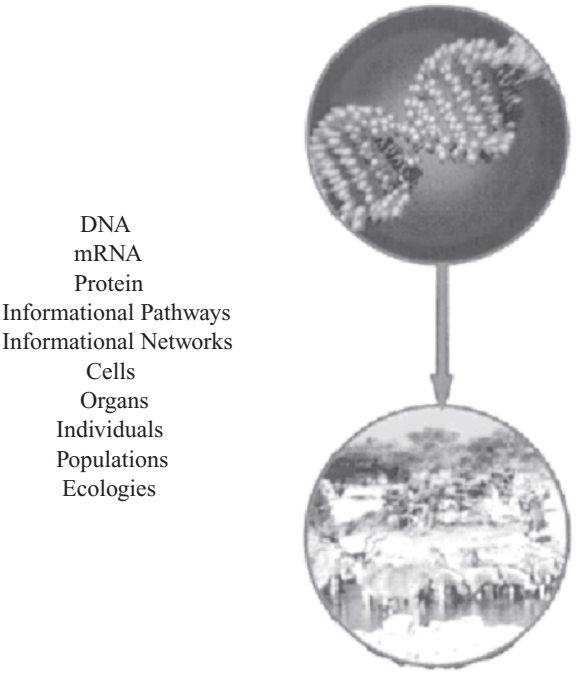

Figure 1.35. Hierarchical levels of biological information. (Reprinted with permission from Hood, L. *J. Proteome Res.* 2002, 1, 399–409. Copyright 2002 American Chemical Society.)

processed, validated, and statistically evaluated. Bioinformatics for proteomics has developed with great speed and complexity with many open source software available with statistical methods and filtering algorithms for proteomics data validation. Some examples at the present are Bioinformatics.org; sourceforge.net; Open Bioinformatics Foundation that features toolkits such as BioPerl, BioJava, and BioPython; and BioLinux, an optimized Linux operating system for Bioinformaticians.

Finally, the visualization of the data obtained from mass spectrometric analysis has also been developed significantly in the past few years. The ability to take different compliments of analyses and integrate them with the goal of correlation is quite daunting when hundreds to thousands of biomolecules are involved. Programs such as Cytoscape allow the visualization of molecular interaction networks. Figure 1.37 is a correlation network between data obtained from proteomics,

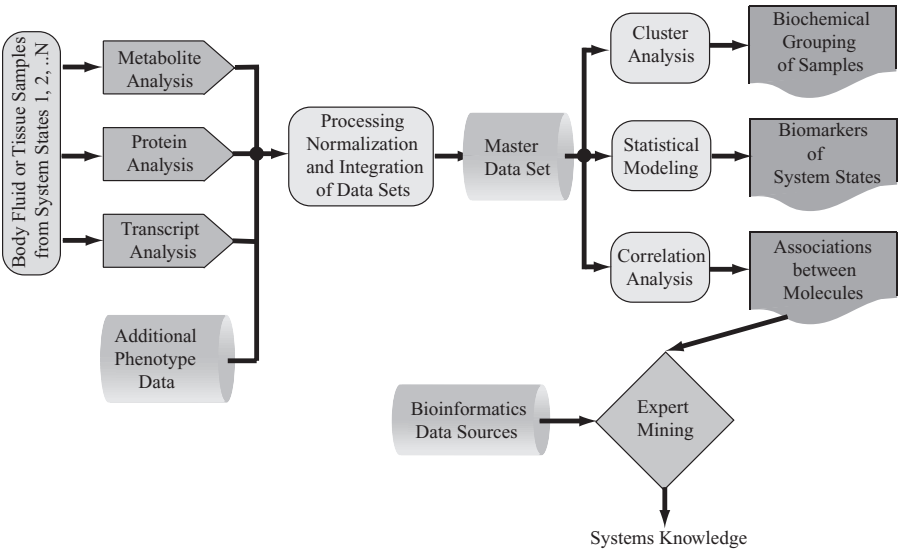

Figure 1.36. Systems biology workflow. Data are produced by different platforms (transcriptomics, proteomics, and metabolomics) followed by integration into a master data set. Different biostatistical strategies are pursued: clustering, modeling, and correlation analysis. Integration with extensive bioinformatics tools and expert biological knowledge is key to the creation of meaningful knowledge. (Reprinted with permission from van der Greef, J., Martin, S., Juhasz, P., Adourian, A., Plasterer, T., Verheij, E.R., McBurney, R.N. *J. Proteome Res.* 2007, 6, 1540–1559. Copyright 2007 American Chemical Society.)

metabolomics, and transcriptomics from liver tissue and plasma. The correlation network found biomarker analytes in the plasma that may be useful in monitoring processes occurring in the organ.

1.5.6 Biomarkers in Cancer

MS is having a substantial impact on systems biology-type studies such as those of biomarker discovery in cancer diagnostics. Figure 1.38 illustrates a scheme for the search for biomarkers in patient-derived samples through mass spectral analyses. From the patient, the figure lists nine types of patient-derived samples that include blood, urine, sputum, saliva, breath, tear fluid, nipple aspirate fluid, and cerebrospinal fluid. Also described in the figure are the various types of "omic" studies done using mass spectrometric techniques such as proteomics,

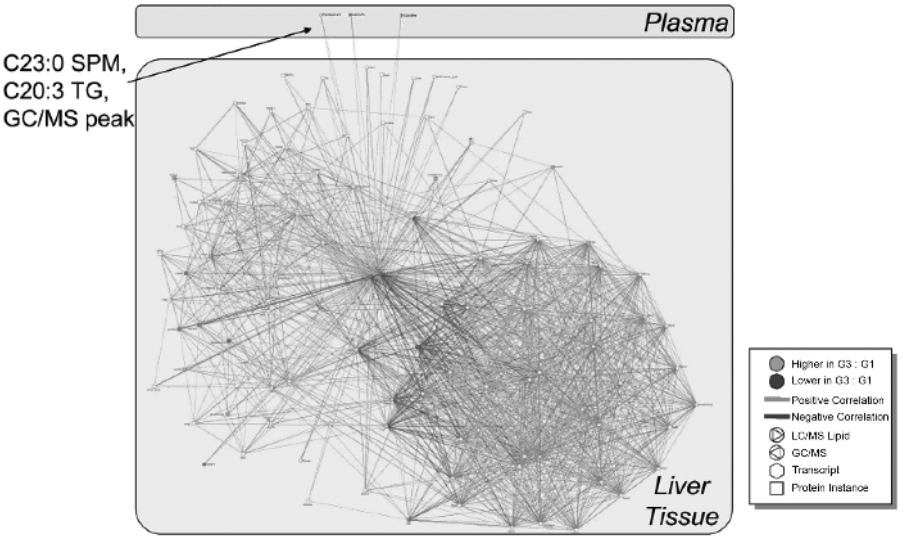

Figure 1.37. Correlation network of analytes across blood plasma (top of figure) and liver tissue (bottom of figure). Analytes include proteins, endogenous metabolites, and gene transcripts. Not only is structure evident among analytes profiled from liver tissue, but there are also a number of correlations to analytes profiled in plasma in this case. Such analytes can serve as useful circulating biomarkers for the tissue-based biochemical processes occurring in the organ. (Reprinted with permission from van der Greef, J., Martin, S., Juhasz, P., Adourian, A., Plasterer, T., Verheij, E.R., McBurney, R.N. *J. Proteome Res.* 2007, 6, 1540–1559. Copyright 2007 American Chemical Society.)

metabonomics, peptidomics, glycomics, phosphoproteomics, and lipidomics. Indicative of the progress or status of a disease, a biomarker is a biologically derived molecule in the body that is measured along with many other species present by the omics methods. Bioinformatics analysis is performed on the mass spectral data to determine the presence of potential biomarkers through expression studies and response differentials. Once identified, the biomarkers can be used for clinical diagnostics such as early detection before the onset of a serious disease such as cancer. This allows medical intervention that may have a substantial influence on the success of an early treatment and subsequent cure. Table 1.5 lists a number of identified potential cancer biomarkers that have been discovered through mass spectrometric analyses, primarily through proteomics and metabonomics.

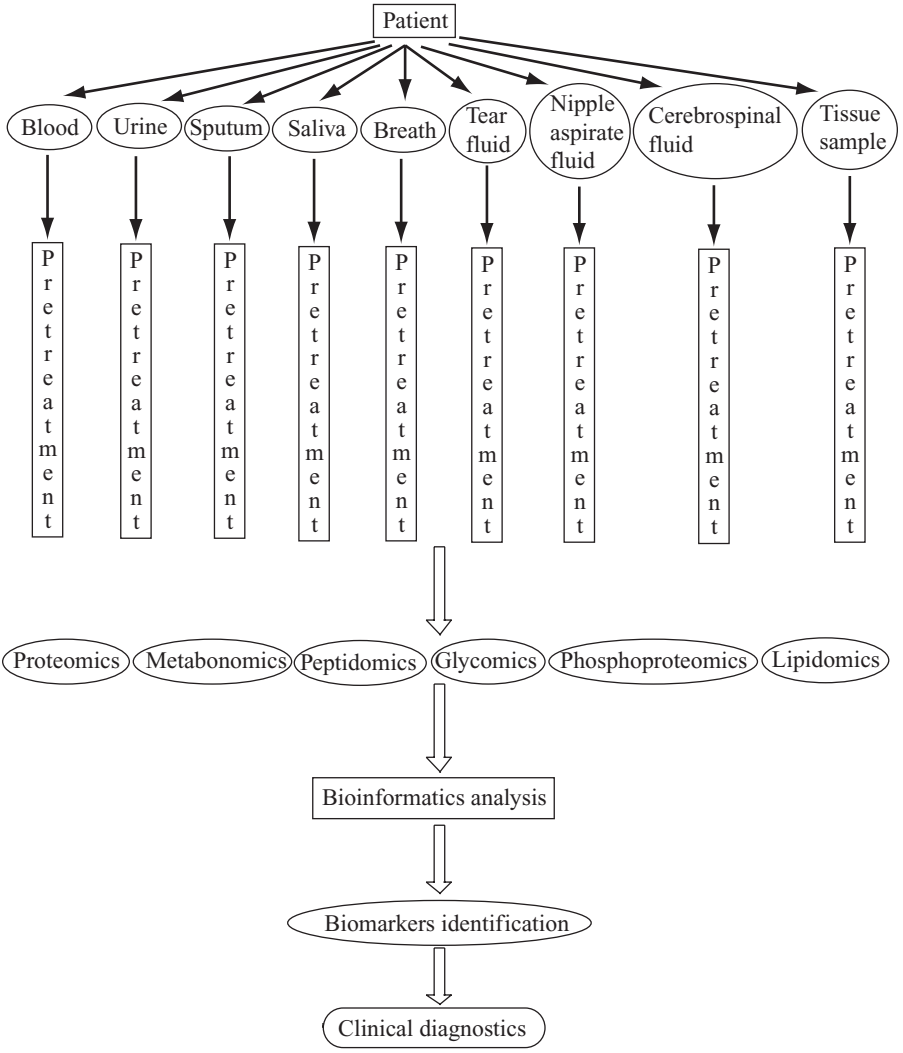

Figure 1.38. Scheme for mass spectrometry-based "omics" technologies in cancer diagnostics. Proteomics is the large-scale identification and functional characterization of all expressed proteins in a given cell or tissue, including all protein isoforms and modifications. Metabonomics is the quantitative measurement of metabolic responses of multicellular systems to pathophysiological stimuli or genetic modification. Peptidomics is the simultaneous visualization and identification of the whole peptidome of a cell or tissue, that is, all expressed peptides with their posttranslational modifications. Glycomics is to identify and study all the glycan molecules produced by an organism, encompassing all glycoconjugates (glycolipids, glycoproteins, lipopolysaccharides, peptidoglycans, and proteoglycans). Phosphoproteomics is the characterization of phosphorylation of proteins. Lipidomics is system-level analysis and characterization of lipids and their interacting partners. (Reprinted with permission of John Wiley & Sons, Inc. Zhang, X., Wei, D., Yap, Y., Li, L., Guo, S., and Chen, F. Mass spectrometry based "omics" technologies in cancer diagnostics. *Mass Spectrometry Reviews*, 2007, *26*, 403–431.)

TABLE 1.5. Potential Cancer Biomarkers Identified by Mass Spectrometry-Based "Omics" Technologies

Biomarkers	"omics" Platforms	MS Methods	Sample Source	Cancer Type
Apolipoprotein A1, inter-α-trypsin inhibitor Haptoglobin-a-subunit Transthyretin	Proteomics	SLDI-TOF	Serum	Ovarian
Vitamin D-binding protein	Proteomics	SELDI-TOF	Serum	Prostate
Stathmin (Op18), GRP 78 14-3-3 Isoforms, transthyretin Protein disulfide isomerase	Proteomics	ESI-MS	Tissue	Lung
Peroxiredoxin, enolase Protein disulfide isomerase HSP 70, α-1-antitrypsin	Proteomics	MALDI-TOF, LC-MS	Tissue	Breast
HSP 27	Proteomics	MALDI-TOF	Serum	Liver
Annexin I, cofilin, GST Superoxide dismutase Peroxiredoxin, enolase Protein disulfide isomerase	Proteomics	MALDI-TOF, ESI-MS, Q-TOF	Tissue	Colon
Neutrophil peptides 1-3	Proteomics	SELDI-TOF	Nipple aspirate fluid	Breast
PCa-24	Proteomics	MALDI-TOF	Tissue	Prostate
Alkanes, benzenes	Metabonomics	GC-MS	Breath	Lung
Decanes, heptanes	Metabonomics	GC-MS	Breath	Breast
Hexanal, heptanal	Metabonomics	LC-MS	Serum	Lung
Pseu, m1A, m1I	Metabonomics	HPLC, LC-MS	Urine	Liver

SLDI, soft laser desorption/ionization; SELDI, surface-enhanced laser desorption/ionization; GC, gas chromatography.

Reprinted with permission of John Wiley & Sons, Zhang, X., Wei, D., Yap, Y., Li, L., Guo, S., Chen, F. Mass spectrometry based "omics" technologies in cancer diagnostics. *Mass Spectrom. Rev.* 2007, 26, 403–431.

REFERENCES

1. Wold, F. *Annu. Rev. Biochem.* **1981**, *50*, 783–814.
2. Wilkins, M.R.; Gasteiger, E.; Gooley, A.A.; Herbert, B.R.; Molloy, M.P.; Binz, P.A.; Ou, K.; Sanchez, J.C.; Bairoch, A.; Williams, K.L.; Hochstrasser, D.F. *J. Mol. Biol.* **1999**, *280*, 645–657.
3. Whitehouse, C.M.; Dreyer, R.N.; Yamashita, M.; Fenn, J.B. *Anal. Chem.* **1985**, *57*, 675.
4. Fenn, J.B. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 524.
5. Dole, M.; Hines, R.L.; Mack, R.C.; Mobley, R.C.; Ferguson, L.D.; Alice, M.B. *J. Chem. Phys.* **1968**, *49*, 2240.
6. Kebarle, P.; Ho, Y. In *Electrospray Ionization Mass Spectrometry*, Cole, R.B., Ed. New York: Wiley, **1997**; p. 17.
7. Cole, R.B. *J. Mass Spectrom.* **2000**, *35*, 763–772.
8. Cech, N.B.; Enke, C.G. *Mass Spectrom. Rev.* **2001**, *20*, 362–387.
9. Gaskell, S.J. *J. Mass Spectrom.* **1997**, *32*, 677–688.
10. Cech, N.B.; Enke, C.G. *Anal. Chem.* **2000**, *72*, 2717–2723.
11. Sterner, J.L.; Johnston, M.V.; Nicol, G.R.; Ridge, D.P. *J. Mass Spectrom.* **2000**, *35*, 385–391.
12. Cech, N.B.; Enke, C.G. *Anal. Chem.* **2001**, *73*, 4632–4639.
13. Taylor, G.I. *Proc. R. Soc. Lond. A* **1964**, *280*, 383.
14. Cao, P.; Stults, J.T. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1600–1606.
15. Ho, Y.P.; Huang, P.C.; Deng, K.H. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 114–121.
16. Kocher, T.; Allmaier, G.; Wilm, M. *J. Mass Spectrom.* **2003**, *38*, 131–137.
17. Lorenz, S.A.; Maziarz, E.P.; Wood, T.D. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 795–804.
18. Daniel, J.M.; Friess, S.D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom.* **2002**, *216*, 1–27.
19. Wilm, M.S.; Mann, M. *Int. J. Mass Spectrom. Ion Process.* **1994**, *136*, 167–180.
20. Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1–8.
21. Caprioli, R.M.; Emmett, M.E.; Andren, P. Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, May 29–June 3, **1994**; p 754.
22. Qi, L.; Danielson, N.D. *J. Pharm. Biomed. Anal.* **2005**, *37*, 225–230.
23. Fernandez de la Mora, J.; Loscertales, I.G. *J. Fluid Mech.* **1994**, *260*, 155–184.
24. Pfiefer, R.J.; Hendricks, C.D., Jr. *AIAA J.* **1968**, *6*, 496–502.
25. Juraschek, R.; Dulcks, T.; Karas, M. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 300–308.

26. Schmidt, A.; Karas, M. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 492–500.
27. Li, Y.; Cole, R.B. *Anal. Chem.* **2003**, *75*, 5739–5746.
28. El-Faramawy, A.; Siu, K.W.M.; Thomson, B.A. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1702–1707.
29. McLuckey, S.A.; Van Berkel, G.J.; Glish, G.L. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60–70.
30. Cerny, R.L.; Gross, M.L.; Grotjahn, L. *Anal. Biochem.* **1986**, *156*, 424.
31. McLuckey, S.A.; Habibi-Goudarzi, S. *J. Am. Chem. Soc.* **1993**, *115*, 12085.
32. Morrison, R.T.; Boyd, R.N. *Organic Chemistry*, 5th ed. Boston: Allyn and Bacon, **1987**.
33. Schulz, G.E.; Schirmer, R.H. In *Principles of Protein Structure, Springer Advanced Texts in Chemistry*, Cantor, C.R., Ed. New York: Springer-Verlag, **1990**.
34. Franks, F. *Biophys. Chem.* **2002**, *96*, 117–127.
35. Edman, P. *Acta Chem. Scand.* **1950**, *4*, 283–293.
36. Shimonishi, Y.; Hong, Y.M.; Kitagishi, T.; Matsuo, H.; Katakuse, I. *Eur. J. Biochem.* **1980**, *112*, 251–264.
37. Morris, H.R.; Panico, M.; Barber, M.; Bordoli, R.S.; Sedgwick, R.D.; Tyler, A. *Biochem. Biophys. Res. Commun.* **1981**, *101*, 623–631.
38. Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*(6), 338–345.
39. Pappin, D.D.J.; Hojrup, P.; Bleasby, A.J. *Curr. Biol.* **1993**, *3*, 327–332.
40. Henzel, W.J.; Billeci, T.M.; Stults, J.T.; Wong, S.C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*(11), 5011–5015.
41. Kenrick, K.G.; Margolis, J. *Anal. Biochem.* **1970**, *33*, 204–207.
42. Gras, R.; Muller, M. *Curr. Opin. Mol. Ther.* **2001**, *3*(6), 526–532.
43. Wolters, D.A.; Washburn, M.P.; Yates, J.R. *Anal. Chem.* **2001**, *73*, 5683–5690.
44. Eng, J.K.; McCormack, A.L.; Yates, J., III *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
45. Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390–4399.
46. Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*(11), 601.
47. Yalcin, T.; Csizmadia, I.G.; Peterson, M.R.; Harrison, A.G. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 233–242.
48. Dongre, A.R.; Jones, J.L.; Somogyi, A.; Wysocki, V.H. *J. Am. Chem. Soc.* **1996**, *118*, 8365–8374.
49. McCormack, A.L.; Somogyi, A.; Dongre, A.R.; Wysocki, V.H. *Anal. Chem.* **1993**, *65*, 2859–2872.
50. Somogyri, A.; Wysocki, V.H.; Mayer, I. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 704–717.

51. Zubarev, R.A.; Kelleher, N.L.; McLafferty, F.W. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
52. Loo, J.A.; Edmonds, C.G.; Smith, R.D. *Science* **1990**, *248*, 201–204.
53. Feng, R.; Konishi, Y. *Anal. Chem.* **1993**, *65*, 645–649.
54. Loo, J.A.; Quinn, J.P.; Ryu, S.I.; Henry, K.D.; Senko, M.W.; McLafferty, F.W. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 286–289.
55. Senko, M.W.; Beu, S.C.; McLafferty, F.W. *Anal. Chem.* **1994**, *66*, 415–417.
56. Schnier, D.S.; Gross, E.R.; Williams, E.R. *J. Am. Chem. Soc.* **1995**, *117*, 6747–6757.
57. Patrie, S.M.; Charlebois, J.P.; Whipple, D.; Kelleher, N.L.; Hendrickson, C.L.; Quinn, J.P.; Marshall, A.G.; Mukhopadhyay, B. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1099–1108.
58. Kirschner, M.W. *Cell* **2005**, *121*, 503–504.
59. Srivastava, R.; Varner, J. *Biotechnol. Prog.* **2007**, *23*, 24–27.
60. Smith, J.C.; Lambert, J.P.; Elisma, F.; Figeys, D. *Anal. Chem.* **2007**, *79*, 4325–4344.