CHAPTER 7

Biological Mass Spectrometry

7.1 IONISATION OF BIOMOLECULES AND BIOPOLYMERS

The low volatility and polar character of biomolecules and biopolymers initially prevented their direct ionisation and analysis by mass spectrometry. These compounds could, at best, only be studied after derivatisation of their polar groups, through methylation and acetylation, or following their degradation by, for example, acid hydrolysis. This converts large biopolymers into manageable (and ionisable) smaller molecules. Even then, only low to moderate (~1000) molecular weight compounds could be introduced into a mass spectrometer in the form of gaseous ions.

This situation changed, first with the development of plasma desorption and fast atom bombardment, and subsequently with the introduction of the electrospray and matrix-assisted laser desorption ionisation techniques. ESI and MALDI are particularly proficient at ionising large biopolymers (to several hundred thousand Daltons) without any pre-treatment or degradation of the sample. These ionisation methods are highly complementary in terms of their performance and suitability to particular samples. As a consequence, most laboratories that study biological compounds by mass spectrometry possess at least two instruments, one with an ESI source and the other with a MALDI source. Alternatively an instrument that can support both ion sources interchangeably can be used.

Peptides, proteins, glycoproteins, glycoconjugates, glycolipids, lipids, oligonucleotides and moderately-sized nucleic acids can all be efficiently introduced into a mass spectrometer by virtue of the ESI and MALDI techniques. Their importance to the analysis of biological macromolecules was recognised in 2002 by the award of the Nobel Prize in Chemistry to those who contributed to their discovery, John Fenn and Koichi Tanaka.

7.2 PEPTIDES AND PROTEINS

7.2.1 Molecular Weight Analysis

Following Barber's demonstration that peptides and small proteins could be successfully ionized by FAB in 1980, the field of protein mass spectrometry has rapidly developed. Mass spectrometry is often the first approach employed to characterise protein samples and now can provide a great deal of structural information. This includes measuring the size of a protein, its complete amino acid sequence, the nature and site of post-translational modifications and even three-dimensional structural characteristics of a protein. Recent work has led to the use of mass spectrometry for studying protein interactions.

Molecular weights are routinely measured to an accuracy of better than 0.01%, or 1 Da at a molecular weight of 10 kDa. The use of ESI and MALDI coupled to high resolution mass spectrometers, such as FT-ICR instruments (Chapter 3), has enabled the molecular weights of proteins (and other biopolymers) to be measured to an accuracy of a few ppm. This is illustrated in Figure 7.1 for the protein chondroitinase of 112 kDa measured to an accuracy of 3 Da (27 ppm). The mass accuracies obtained in all cases are far superior to those from other methods, including gel electrophoresis.

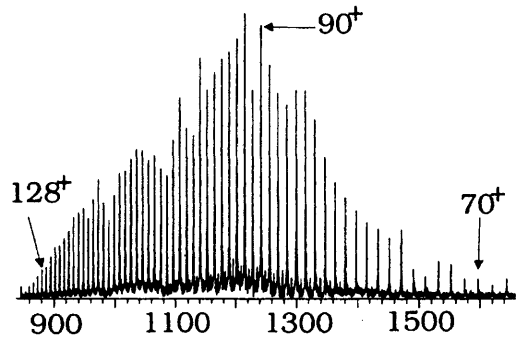

Figure 7.1 *The ESI FT-ICR mass spectrum of the protein chondroitinase of molecular weight 112 kDa measured to an accuracy of 1 Da (or 9ppm)* (Source: McLafferty *et al.*, *J. Am. Soc. Mass Spectrom.*, 1997, **8**, 380–383, Figure 1 – Part A)

Molecular weight measurements provide useful information in their own right, and may indicate that the protein isolated from a biological sample is different or in a modified form from that expected. Heterogeneous proteins and protein mixtures are often encountered in mass spectral data even when an investigator might believe the sample to be

pure. Mass spectrometry can also analyse a complex mixture of many different proteins in a single analysis and with high sample throughput, both features that are important to proteomic discoveries discussed later in this chapter.

A number of web-based algorithms have been developed to search protein databases with molecular weight information. Yet these databases contain a large number of proteins that coincidentally or due to structural similarities can share very similar, or even identical, molecular weights. As such it is often not possible to use the molecular weight of a protein alone to unequivocally identify it. For this reason, databases can be searched using a range of available information about a protein including its biological source, *pI*, molecular weight or features of its structure.

7.2.2 Mass Mapping

To assist with the identification of a protein that may already be known or display similar sequence homology to a protein that appears in a database, a *peptide mass map* can be of use. Here the protein is digested with a site-specific protease such as trypsin (which cleaves proteins on the C-terminal side of arginine and lysine residues) and the peptide products are analysed collectively by mass spectrometry (Figure 7.2). The molecular weights of these proteolytic peptides, in addition to that for the intact protein, are searched against the mass of the theoretical peptide products generated by the same enzyme for all proteins in the database. The protein(s) with the closest match based on the number and closeness of the masses matched appears in the output of the algorithm with the highest score.

A few important points must be made in relation to these searches. First, a match in the mass of a peptide with a hypothetical fragment of a known protein does not alone prove that the peptide has an identical sequence. Peptides identical in mass may coincidentally possess the same mass even when they have quite different sequences. Furthermore, some amino acid residues are indistinguishable by mass (see Appendix 8 for glutamic acid and lysine, leucine and isoleucine) and could be interchanged within a protein with no change to its molecular weight.

A second important issue concerns the level of protein coverage represented by the map. Ideally, the mass spectrum should contain ion signals for all peptides across the entire sequence of the protein. In practice, due to ionisation and detection efficiencies and the ease with which some sites are cleaved by enzymes over others, the map reflects only part of the total protein. Coverage levels vary depending on the complexity of the sample,

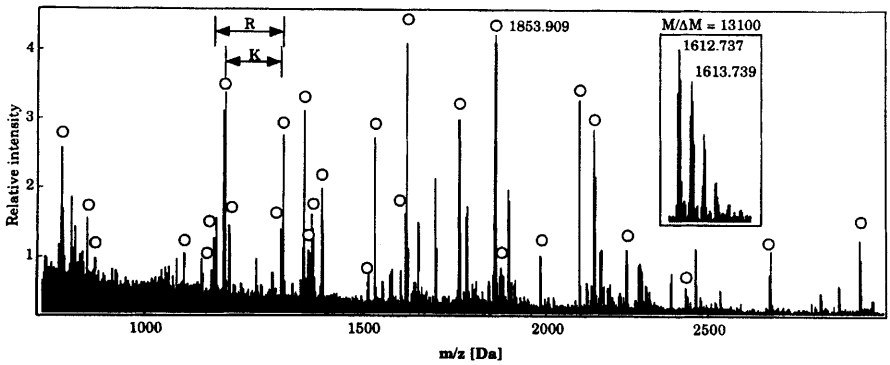

Figure 7.2 *Identification of yeast protein ILV5 from a MALDI mass map by database searching. Ions whose measured m/z are within 50 ppm of those calculated are indicated by circles.*
(Source: A. Shevchenko, O.N. Jensen, A.V. Podtelejnikov, F. Sagliocco, M. Wilm O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie and M. Mann, Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels, *Proc. Natl. Acad. Sci. USA*, 1996, **93(25)**, 14440–14445)

the nature of the proteins and the type of mass spectrometer. Thus a match of masses for peptides that only represent a portion of the total protein with those theoretically generated for a database entry leads to the possibility that they differ within other regions.

A third issue to consider when identifying proteins is mass accuracy. Irrespective of the type of mass spectrometer and measurement, some mass error is inevitable and the assignment of a protein to that of a database entry inherits these errors. This last consideration is becoming less important given that the molecular weights of peptides can be routinely measured with high accuracies. Mass measurements accurate to a few ppm are not uncommon and can be obtained even on time-of-flight instruments employing ion mirrors and time-lag focusing where appropriate mass calibration procedures are employed.

A number of web-based algorithms that use peptide mass maps to search protein databases are publicly available. These include Mascot at Matrix Science and Mass Mapper in the UK, Peptide Search at the European Molecular Biology Laboratory (EMBL), and Protein Prospector and PROWL developed in the United States. In many cases, mirror sites have been established at other laboratories to assist with data transfer throughout the world. Uniform resource locator (URL) addresses for these sites appear in Appendix 11.

An example of the output from the ProFound algorithm at the PROWL web site based on a search of the Swiss-Prot database for some

of the peptide mass map data (*m/z* 1547.8, 1612.7 and 2145.1) from Figure 7.2 is shown in Figure 7.3. A mass error of 0.05 Da was selected and the cysteine residues were deemed to be unmodified. A protein molecular weight range of 20–70 kDa was chosen but no *pI* range selected.

ProFound - Search Result Summary

Version 4.10.5
The Rockefeller University Edition

Protein Candidates for search BCA7C9DF-0434-7A2DBB76 [88041 sequences searched]

Rank	Probability	Est'd Z	Protein Information and Sequence Analyse Tools (T)	%	*pI*	kDa	
1	2.0e-001	0.03	T gi|6686325|sp|P71018|PLSX_BACSU FATTY ACID/PHOSPHOLIPID SYNTHESIS PROTEIN PLSX	10	5.6	35.75	®
2	2.0e-001	0.03	T gi|113783|sp|P06547|AMYB_BACCI BETA-AMYLASE PRECURSOR (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE)	6	6.3	62.88	®
3	1.8e-001	0.02	T gi|401194|sp|P31015|TNA2_SYMTH TRYPTOPHANASE 2 (L-TRYPTOPHAN INDOLE-LYASE 2) (TNASE 2)	8	5.8	50.59	®
4	1.3e-001	-	T gi|120716|sp|P09316|G3P_ZYMMO GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)	10	6.3	36.08	®
5	1.0e-001	-	T gi|124371|sp|P06168|ILV5_YEAST KETOL-ACID REDUCTOISOMERASE, MITOCHONDRIAL PRECURSOR (ACETOHYDROXY-ACID REDUCTOISOMERASE) (ALPHA-KETO-BETA-HYDROXYLACIL REDUCTOISOMERASE)	8	9.2	44.35	®
6	9.4e-002	-	T gi|1703244|sp|P53448|ALFC_CARAU FRUCTOSE-BISPHOSPHATE ALDOLASE C (BRAIN-TYPE ALDOLASE)	9	6.4	39.46	®
7	2.4e-002	-	T gi|134102|sp|P08823|RUBA_WHEAT RUBISCO SUBUNIT BINDING-PROTEIN ALPHA SUBUNIT PRECURSOR (60 KD CHAPERONIN ALPHA SUBUNIT) (CPN-60 ALPHA)	6	4.8	57.50	®
8	1.9e-003	-	T gi|1169045|sp|P43374|COX2_DEKBR CYTOCHROME C OXIDASE POLYPEPTIDE II	8	4.3	28.26	®
9	1.9e-003	-	T gi|1169041|sp|P43371|COX2_BRECU CYTOCHROME C OXIDASE POLYPEPTIDE II	8	4.3	28.28	®
10	1.8e-003	-	T gi|3123193|sp|P08897|COGS_HYPLI COLLAGENASE PRECURSOR (HYPODERMIN C) (HC)	6	4.6	28.56	®

Figure 7.3 *The ten highest-ranking entries output by the PROFOUND algorithm from a search of the SWISS-PROT database with three m/z values (m/z 1,547.8, 1,612.7 and 2,145.1) of tryptic peptide ions generated from a yeast protein ILV5*

The yeast protein IVL5 (entry P06168) appears as the fifth entry in the output. The use of a larger set of *m/z* values of ions from the mass map allows this protein to be identified with a greater confidence.

7.2.3 Peptide and Protein Sequencing

A protein can only be unequivocally identified when its entire amino acid sequence has been determined. Peptide and protein sequencing can now be accomplished solely within the confines of a mass spectrometer, or by using chemical and enzymatic approaches in conjunction with mass spectrometric analysis.

7.2.3.1 Chemical and Enzymatic Sequencing In these approaches, the peptide or protein under investigation must be in a purified form. Other contaminating compounds in the sample can seriously compromise the

analysis. The peptide or protein is treated with a chemical or enzyme to cleave amino acid residues from the N or C-terminus in a stepwise manner. An aliquot of the sample is removed at a series of time points and the collections combined. A mass spectrum is then recorded of the combined reaction products. The difference in mass of the products should correspond to the residue masses of the amino acids (Appendix 8) representing the molecular weight of an amino acid less 18 u for a molecule of water released in each step. This procedure has been described as *ladder sequencing* with the sequence of a peptide or the protein termini read directly from the mass spectrum. A partial C-terminal sequence (WCND) for an epitopic peptide of hen lysozyme has been determined based on its partial digestion with carboxypeptidase Y after reduction and alkylation of its cystein residues with ethyl pyridine (C*) (Figure 7.4).

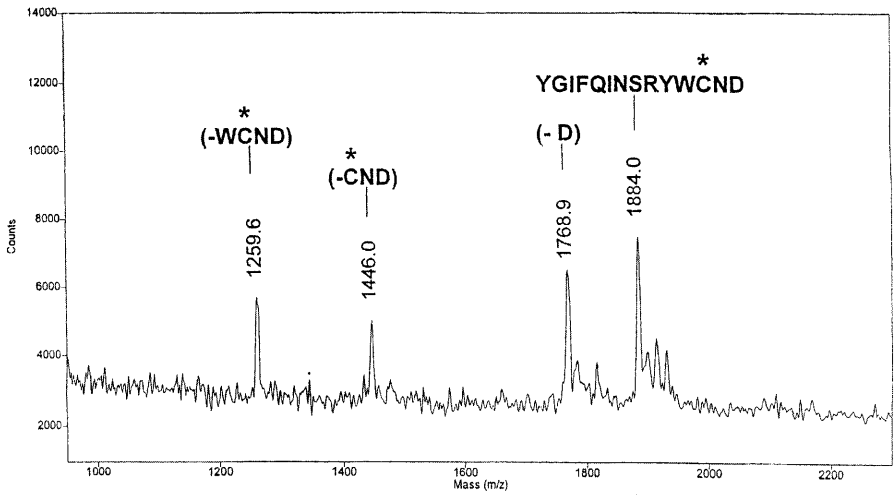

Figure 7.4 *MALDI mass spectrum of the products of limited proteolysis of an epitopic peptide of hen lysozyme with carboxypeptidase Y*
(Source: J.G. Kiselar and K.M. Downard, *Anal. Chem.*, 1999, **71**, 1792–1801, Figure 2)

Chemical methods employ the first Edman degradation reaction in which the N-terminus of the protein is converted to a phenylisothiocyanate (PITC) derivate. The modified N-terminal amino acid residue is then cleaved with trifluoroacetic acid and the process repeated. This approach is not suitable for the sequencing of N-terminally blocked proteins or peptides as the initial PITC derivative cannot be formed. The enzymatic approach alternatively makes use of amino or carboxypeptidases to cleave amino acids from the N and C-termini respectively.

In practice, the efficiency of a chemical or enzymatic cleavage

decreases with the length of the polypeptide chain. Therefore, such an approach is capable of generating a complete amino acid sequence of a peptide, but not for a protein. To obtain the complete sequence for a protein, tandem mass spectrometry can be employed.

7.2.3.2 Tandem Mass Spectrometric Sequencing The sequencing of proteins by tandem mass spectrometry follows many years of study into the dissociation pathways of peptides under collisional and other ion activation conditions. Peptides have been shown to fragment along predictable pathways, for the most part involving the cleavage of the peptide backbone.

The common fragment ions for a dissociated peptide are summarised in Figure 7.5 where a nomenclature has been adopted to identify those fragments that contain the charge in the N-terminal portion (a_n, b_n, c_n, d_n) or in the C-terminal portion (x_n, y_n, z_n, v_n, w_n) of the peptide. Lower case letters are preferred so that they are not confused with nomenclature used for amino acids in their single-letter code. Many fragmentation processes involve the additional transfer of hydrogen atoms and protons but these are usually ignored for the purposes of labelling peaks in the MS/MS spectra of peptides to minimise the level of annotation. The numeral subscript (n) denotes the number of amino acid residues from either the N or C-terminus to the cleavage site. For example, a b_4 ion is formed by cleavage of the amide N–C bond at the 4th residue from the N-terminus. The fragment ions (d_n, v_n, w_n) that are formed from the cleavage of both the backbone and a sidechain group are produced only at high collision energies (keV) where precursor ions are accelerated from the ion source at kV potentials. These fragments are only observed in tandem experiments performed on magnetic sector and time-of-flight-based instruments. Although at first sight they appear to complicate a tandem MS/MS spectrum, these side-chain specific fragment ions are useful to distinguish between isobaric residues such as leucine and isoleucine that exhibit unique side chain losses.

Ideally a tandem MS/MS spectrum will contain only one series of ions in which the *m/z* value differences between each successive fragment corresponds to the mass of an amino acid residue (Appendix 8). In practice, a number of different fragment ion types can be produced simultaneously from discrete precursor ions due to the distinct structure of amino acids and the nature and energetics of the dissociation event. As certain bonds are more easily broken, and some product ions are more stable than others, the energy transferred and distributed throughout a peptide ion during activation influences its fragmentation. These factors are most pronounced in high energy (keV) dissociation experiments

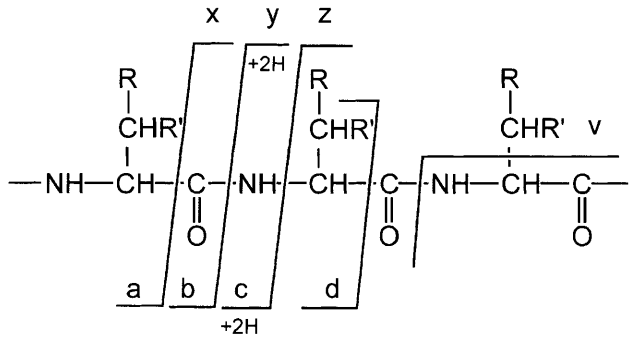

Figure 7.5 *Common fragment ions formed upon the dissociation of peptide or protein ions. Fragment ions containing the N-terminal portion of the peptide or protein are denoted a, b, c or d, those containing the C-terminal portion are denoted x, y, z, v or w. The d, v and w ions are only formed in high (keV) energy dissociation experiments*

where the nature of the fragments formed is strongly driven by the location of basic amino acid residues (particularly arginine and lysine). Low energy (eV) dissociation experiments, in contrast, generally give rise to mostly b_n and y_n type fragments regardless of the peptide sequence (see Figure 7.6).

Proteins can be sequenced following their treatment with a site-specific protease. MS/MS spectra are then acquired for each of the proteolytic peptides without the need for their purification. This establishes the sequence of segments of the protein but does not determine the order in which the segments appear in the molecule. A second protease of different specificity is used to generate a complementary set of peptides whose mass-to-charge ratios alone or in conjunction with their sequences derived from earlier MS/MS experiments allows the entire protein sequence to be assembled.

Recent advances have overcome the need to digest a protein in order to obtain sequence data. Using a FT-ICR mass spectrometer, McLafferty and colleagues have performed tandem MS/MS experiments on the multiply-charged ions of intact proteins using electron-capture dissociation. This so-called "top-down" approach has led to the production of fragment ions that cover almost all of the protein's sequence (see Figure 4.3).

The great challenge rests with interpreting the single MS/MS spectrum of the protein ions to derive the sequence. Although many of the fragment ions support more than one charge, the values for these can be measured based on the difference in the mass-to-charge ratios within their isotope distributions which are easily resolved on the FT-ICR

instrument. However, since many of the fragments represent a large segment of the protein, the probability that their ions contain no ^{13}C atoms becomes exceedingly small. Thus many of the isotope distributions contain little to no detectable levels of ions containing no ^{13}C (the monoisotopic or ^{12}C-only ion peaks). It is therefore necessary to measure the mass of the fragment based on a ^{13}C-containing ion peak. The number of ^{13}C atoms in the ion must be known in order for this mass measurement to be reliable. This has been approached by comparing the ion intensities within the resolved isotopic distributions with those generated theoretically. The closest match between an experimental and theoretically calculated profile allows the ^{13}C in the ion to be assigned and the mass of the fragment to thus be derived.

These top-down experiments are an impressive demonstration of the performance of an FT-ICR mass spectrometer. The use of an electrospray ion source provides an efficient means to introduce proteins directly into the mass spectrometer. The high resolution and ion storage capabilities of the instrument coupled with electron-capture dissociation allow the protein ions to be efficiently dissociated and the fragments resolved.

7.2.3.3 Interpretation of MS/MS Spectra of Peptides Although the methods described in this section have been developed to interpret the tandem (MS/MS) mass spectra of peptides, they can be extended to sequence proteins. Both manual and computer-assisted approaches are now in use to interpret the tandem mass spectra of peptides. The assisted approaches range from algorithms that can generate probable peptide sequences from the MS/MS data to those that attempt to identify the peptide, and the protein from which it may have been derived, by comparing the MS/MS spectral profile with a set of hypothetically-generated MS/MS spectra for all proteolytic peptides with the same mass across proteins of a database. Computer-based methods have considerably aided in the interpretation of MS/MS spectra, but it is important to note that they are fallible and programs are known to assign incorrect sequences on some occasions.

It is useful then to be practiced in interpreting MS/MS spectra of peptides by manual means. One approach to do so is illustrated for the data shown in Figure 7.6 recorded under low energy collision conditions. All mass-to-charge ratios represent monoisotopic (^{12}C only) values. The MS/MS spectrum exhibits a series of fragments from m/z 175 to 1,511 in addition to a doubly-protonated precursor ion at m/z 813.2. The monoisotopic mass of the peptide is then 1,624.4 (or 813.2×2 (to correct for the charge $z = 2$) $- 2$ (for the mass of the protons attached)). Note also

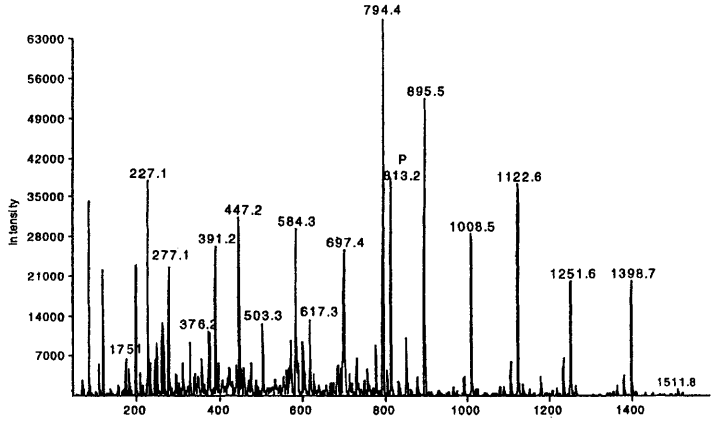

Figure 7.6 *Low energy CID tandem MS/MS spectrum of the doubly-protonated precursor (P) ions (m/z 813.2) of peptide XFENXTPXHANSR. X denotes either leucine or isoleucine*
(adapted from J.R. Chapman, ed., *Peptide and Protein Analysis by Mass Spectrometry*, Ch. 6, Fig. 6, Humana Press, NJ, 1996, p. 95)

that all of the fragment ions with $m/z > 813$ must be singly-charged since their mass is less than that of the intact peptide.

A flow chart (Figure 7.7) can be constructed where amino acid residue mass (Appendix 8) differences are searched for between one fragment ion and the next, beginning with the one with the highest m/z value. The lack of a mass match ends a branch, and another possible association must be considered. Starting at the fragment ion with the highest m/z of 1,511.8, the subtraction of 113.1 (corresponding to either leucine or isoleucine) arrives at m/z 1,398.7. Subtracting a further 147.1 units (consistent with the residue mass of phenylalanine), leads to a product ion at m/z of 1,251.6. Repeating the process further, and ignoring the precursor ion signal, the sequence of amino acids can be read in single letter code as XFENXTP where X = I or L. Two ions appear below the ion at m/z 697.4 with m/z values of 617.3 and 584.3. The first of these corresponds to a mass loss of 81 u that is inconsistent with an amino acid residue. The second corresponds to a mass loss of 113 u consistent with leucine or isoleucine. Thus we derive the partial sequence XFENXTPX.

It is now necessary to consider the fragment ions of low m/z. By repeating the subtractive process further, the sequence can be extended to XFENXTPXHANS down to the ion at m/z 175.1. This remaining mass is associated with at least one amino acid residue. If the ion at m/z 175 is a b_1 ion, the N-terminal residue must have a mass of (175 – 1) (for the N-terminal hydrogen atom) or 174. Since this value is not consistent with the mass of an amino acid residue (Appendix 8), the ion

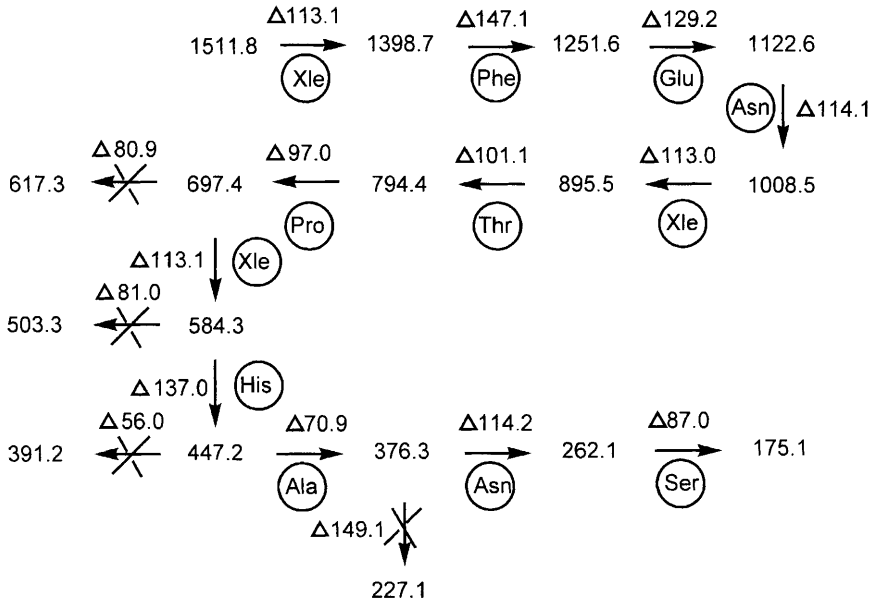

Figure 7.7 *Interpretation of the low energy CID tandem MS/MS spectrum shown in Figure 7.6*

must be considered to be a y_1 ion. The C-terminal residue must then have a mass of ((175 – 17) (for the mass of the C-terminal HO group) – 2) (for the mass of the two protons attached to the ion) or 156. This mass is consistent with that of arginine. For the interpretation to be correct, all ions must be of the y_n-series and thus the direction of the sequence from the N to C-terminus is XFENXTPXHANSR. It is not possible to distinguish leucine and isoleucine residues from MS/MS spectra recorded under low-energy conditions. These residues can only be distinguished in high-energy experiments where side chain (d_n, w_n) fragments are produced.

As illustrated above, the interpretation of tandem mass spectral data for peptides involves locating fragment ions that are separated by approximately 100 u (or strictly 57 to 186 u) consistent with the residue masses of the amino acids. Mass loss differences greater than 186 may indicate that the ions are not from the same series or type, or that one or several ions of a fragment ion series do not appear in the spectrum. For example, a spectrum may exhibit a b_4 and b_6 ion, but not a b_5 ion associated with cleavage of the 5th residue from the N-terminus at the amide bond. This b ion may not form due to energetic or stability issues. Under these circumstances, it is sometimes still possible to theorise as to the identity of the missing sequence based on the mass of dipeptides. It is not

possible, however, without additional information to assign the order of these two amino acid residues.

7.2.3.4 Detection of Mutants and Post-Translational Modifications by MS/MS It is immediately apparent that MS/MS spectra can also be used to detect the presence of amino acid substitutions within homologous proteins (mutants) and post-translational modifications of amino acid side chains. Any such alterations will lead to a change in the *m/z* of all fragment ions that contain these amino acid residues. Fragment ions that do not contain the substituted residue or modification will appear at the same *m/z* ratio as that for the unmodified proteolytic peptide. However, post-translational modifications are often incomplete such that peptides that contain these modified residues are present at low levels in a protein digest versus their unmodified counterparts. Furthermore, some post-translational modifications (such as phosphorylation) can adversely impact the ionisation efficiency of peptides.

A number of tandem mass spectral approaches have been implemented to assist with the detection and analysis of these modified peptides and their protein counterparts. Where phosphorylated peptides are ionised and detected in the negative ion mode, a series of characteristic ions, $H_2PO_4^-$ (*m/z* 97), PO_3^- (*m/z* 79) and PO_2^- (*m/z* 63) at low *m/z* ratios are detected. Alternately, a precursor ion scan (Section 4.3.2.2) to identify all peptides that dissociate to form these fragment ions can be employed. A neutral loss scan (Section 4.3.2.3) can also be of use to detect phosphopeptides due to the characteristic loss of 98 *u* associated with H_3PO_4.

7.2.4 Protein Structure and Folding

Beyond the size and sequence of a protein, mass spectrometry can provide insights into a protein's secondary and tertiary structure. The ability to detect differences in a protein's conformational state came to light following the development of the ESI technique. Early studies of proteins showed that the charge state distribution of ions for a protein in its native state was centred at a higher *m/z* than that for the same protein in a denatured state. This was shown both as a function of pH, and more convincingly following reduction of a protein's disulphide bonds (Figure 7.8). The relationship between a protein's conformational state in the gas phase (in the absence of solvent) and that in solution is the subject of current investigation and debate. To probe the solution state of a protein, a number of indirect approaches that utilise mass spectrometry have been developed. Foremost among these is the use of hydrogen/deuterium exchange.

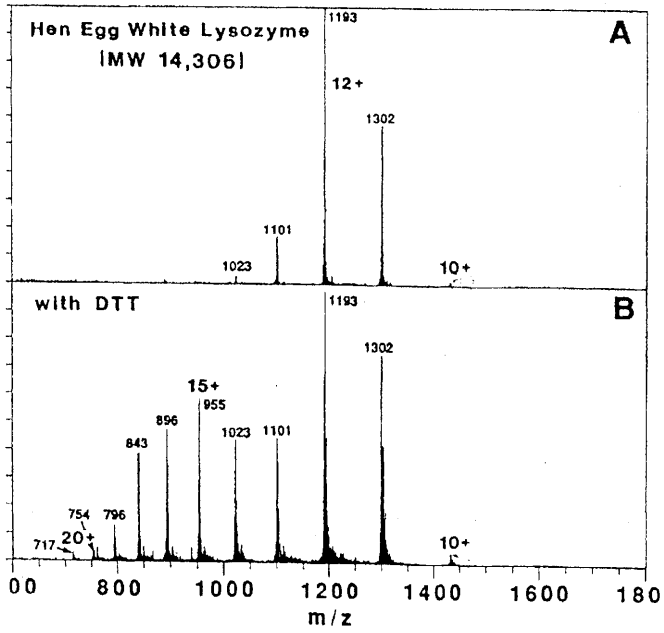

Figure 7.8 *ESI mass spectra of hen egg white lysozyme in 5% acetic acid (A) and following the addition of dithiothreitol (B)*
(Source: J.A. Loo, C.E. Edmonds, H.R. Udseth and R.D. Smith, *Anal. Chem.*, 1990, **62**, 693–698, Figure 2)

7.2.4.1 Hydrogen Exchange Mass Spectrometry In these experiments, some of the hydrogen atoms in proteins are exchanged with deuterium or tritium by dissolving proteins in an isotopically-enriched solvent (such as deuterium oxide, D_2O). The exchange of hydrogen occurs at different rates that are dependent in part upon the accessibility of each atom to the bulk solvent. Hydrogen atoms covalently bonded to carbon undergo isotopic exchange so slowly that this exchange is not observed. Hydrogen atoms of the hydroxyl, sulphydryl, amine and carboxylic acid groups of amino acid side chains exchange very rapidly (less than a second) and their rates of exchange are indistinguishable. Hydrogen of the amide backbone, however, undergo isotopic exchange at different rates ($k \sim 5$ to 0.05 s^{-1}) that can be measured and compared. Since each amino acid, with the exception of proline, has one amide-hydrogen located in the protein backbone, hydrogen exchange levels and rates can be measured along the entire length of the protein molecule. As these backbone amide hydrogen atoms participate in formation of secondary structural elements (such as alpha helices and beta sheets), the exchange rates are affected markedly by the structure and stability of the protein. At pH 7,

the half-lives for isotopic exchange of amide hydrogens within a protein may be as short as seconds or as long as several months.

The basis of the experiment is outlined in Figure 7.9. First a protein in solution at physiological pH (or some other pH of interest) is allowed to undergo hydrogen atom exchange with deuterium oxide. The reaction is then quenched by lowering the pH to 2.5 by the addition of acid and lowering the temperature to 0 °C. Under these conditions, the fast-exchanging hydrogen of the amino acid side chains are left unlabelled, while the slow-exchanging amide hydrogen atoms remain either substituted with deuterium or unexchanged. The entire protein is then analysed by mass spectrometry where the total level of deuterium incorporated is measured based on the increase in its molecular weight. For each deuterium atom that replaces hydrogen, the molecular weight of the protein increases by 1 Da. To determine the site of deuterium incorporation, a portion of the quenched sample is digested with the non-specific protease pepsin. This protease cleaves the protein efficiently at low temperature at residues across the entire protein sequence. The peptide segments are then analysed by mass spectrometry and their molecular weights measured. The level of deuterium incorporated into each peptide segment provides a way in which to measure the accessibility of the amide hydrogen to the bulk solvent across the entire protein in its original structure. Those peptides containing more deuterium represent a region of the protein backbone that is more accessible to solvent. Those peptide segments that contain minimal to no deuterium are interpreted as buried or shielded from solvent within the protein structure. In this way, the structure of a protein can be explored by mass spectrometry under a variety of solution conditions.

By extension, this approach has been used to study the interaction of proteins with other molecules. Amide hydrogen in regions of the protein within the binding site will undergo isotopic exchange at a slower rate when shielded from the bulk solvent by the interacting molecule. When such an association does not occur, the same region will exchange hydrogen with the solvent more rapidly. Thus the location of the binding site can be determined.

One difficulty with hydrogen exchange experiments arises during the analysis of the protein and peptide segments. It is critical that once the reaction is quenched or stopped that the protein does not undergo reverse exchange prior to or during its mass spectral analysis. In practice, the samples are subjected to moisture during the ionisation process that can result in deuterium loss. Therefore some care in performing the mass spectral analysis is required to ensure that the level of deuterium in

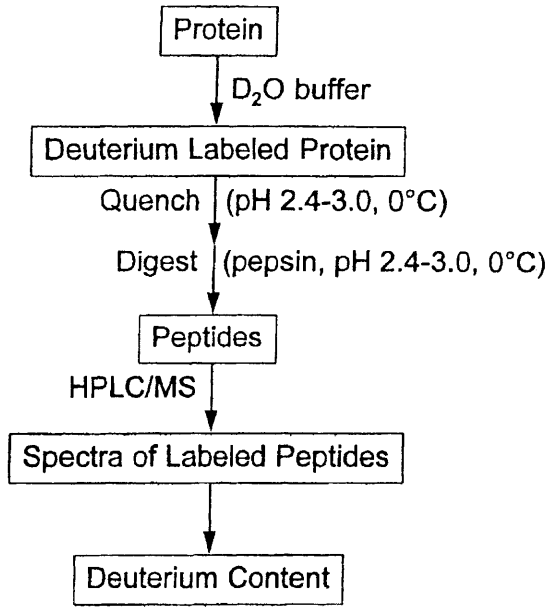

Figure 7.9 *Typical procedure employed in hydrogen exchange mass spectrometry experiments to study protein structure*
(Source: Z. Zhang and D.L. Smith, *Protein Science*, 1993, **2**, 522–531)

regions of the protein are not underestimated. One way to overcome this is to employ a reverse-exchange strategy. Here all the amide hydrogens within a protein are first completely exchanged with deuterium. This is usually achieved by denaturing the protein. The level of deuterium incorporation is then verified by mass spectrometry. The protein is then returned to its native state (usually by pH or temperature adjustment) and the reverse-exchange of deuterium with hydrogen monitored in the same manner as that described above.

Hydrogen/deuterium exchange has also been used to investigate the structure of proteins and their transient intermediate states *in vacuo* following the trapping of their ions for extended periods. These studies allow the conformational characteristics and dynamics of gaseous ions to be explored and compared with solution state observations. At least six different intermediate states for the protein, cytochrome c, have been characterised in such gas-phase experiments.

7.2.4.2 Ion Mobility Mass Spectrometry Mass spectrometry can also be employed to study the conformational characteristics of gas phase ions by measuring their mobility as they pass through an inert buffer gas. In these *ion mobility measurements*, protein ions have drift times that depend upon their average cross-section and hence conformation.

Different drift times have been recorded for the ions of proteins in their native (oxidised or disulphide-bridged) and denatured (reduced) forms.

7.2.4.3 Radical-Based Studies of Protein Structure A recent development that has been employed to study the structure and dynamics of proteins and their interactions by mass spectrometry involves their reaction with radicals. Although the reaction of radicals with proteins *in vivo* leads to their structural degradation and aggregation through cross-linking, it has been found that proteins can undergo limited oxidation without structural change when the reaction times are kept very short (several milliseconds).

Furthermore, the degree to which oxidation occurs is highly dependent on the accessibility of amino acid side chains to the bulk solvent. By measuring the site and degree of oxidation at amino acid markers throughout the protein by mass spectrometry, a protein's structure can be verified or even predicted.

The approach has a number of advantages over hydrogen exchange experiments in that the experiments can be performed extremely rapidly, the radical-induced oxidation reactions are irreversible, and the reaction timescale is sufficiently short to allow some protein conformational changes to be followed at both a global and local level (Figure 7.10). The approach has also been applied to study the dynamics of protein folding and the interactions of proteins with other molecules.

7.2.5 Protein Complexes and Assemblies

Beyond the indirect approaches described above, the development of ESI-MS for the study of proteins gave rise to early observations in which ions corresponding to intact protein complexes were sometimes detected within the mass spectrometer. Depending on the solution conditions under which the sample is introduced, and those of the mass spectrometer itself, these complexes can reflect either solution-state or non-specific associations. The pH of the solution, temperatures within the ion source, and the degree to which ions are accelerated within and as they leave the ion source, are all of importance for the detection of gas phase protein complexes.

The nature of these gas phase protein complexes is a matter of immediate question, since their molecular weights reflect that they typically have no molecules of solvent attached. Electrostatic interactions between charged groups rather than hydrophobic interactions between the component ions and the solvent are believed to play a more

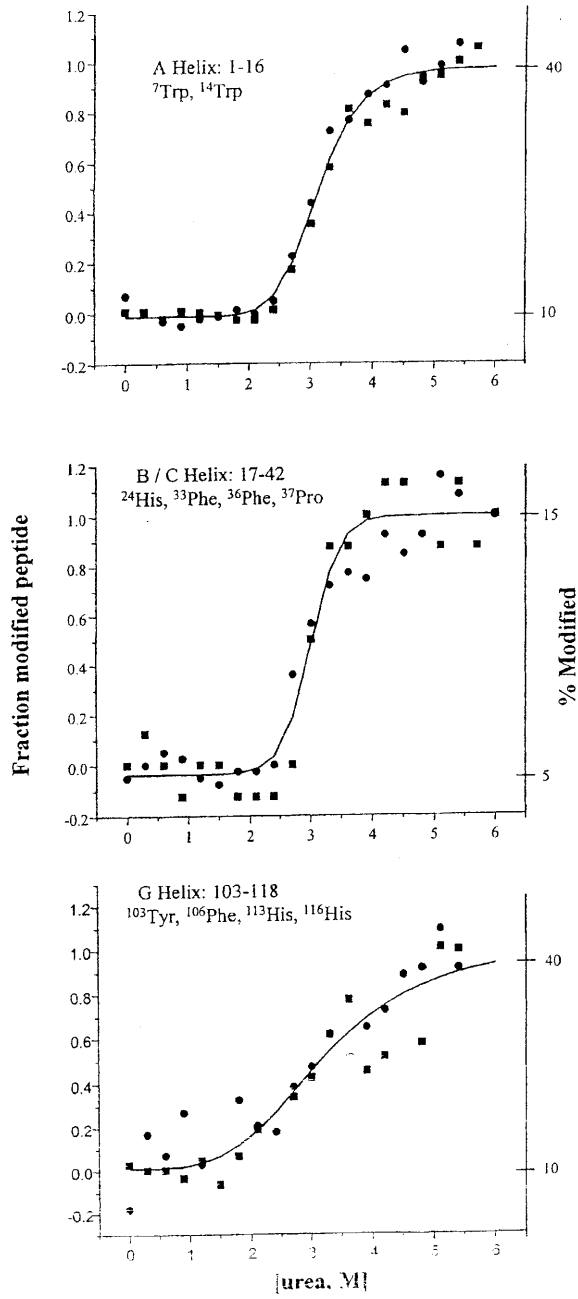

Figure 7.10 *Unfolding profiles within segments of apomyoglobin obtained by limited oxidation of the protein with hydroxyl radicals on millisecond timescales at increasing concentrations of urea*
(Source: S.D. Maleknia and K.M. Downard, *Eur. J. Biochem.*, 2001, **268**, 5578–5588, Figure 6)

important role in their stability. It cannot therefore be assumed, even when a known solution-based protein complex is observed in the gas phase, that the complex maintains the same structural features. Nonetheless, a growing body of data now suggests that relationships between solution state and gas phase protein complexes *do* exist in some systems, and that their direct detection by mass spectrometry may provide a useful means by which to identify such associations in the first instance. This has important ramifications for studies in proteomics, an ultimate description of which requires all protein associations within a cell to be identified.

Although the majority of proteins, and other macromolecular complexes, have been detected using ESI mass spectrometry (Figure 7.11), some complexes and aggregates have also been detected using the MALDI approach. This is even more surprising since samples are introduced into the mass spectrometer from a solid surface onto which the analyte is added to a high concentration of an organic matrix. Such conditions are expected to dissociate any protein or macromolecular complex prior to the ionisation event. A number of protein and other macromolecular complexes, however, have been both preserved and detected by MALDI mass spectrometry. Among them are immune complexes between protein antigens and monoclonal antibodies (Figure 7.12).

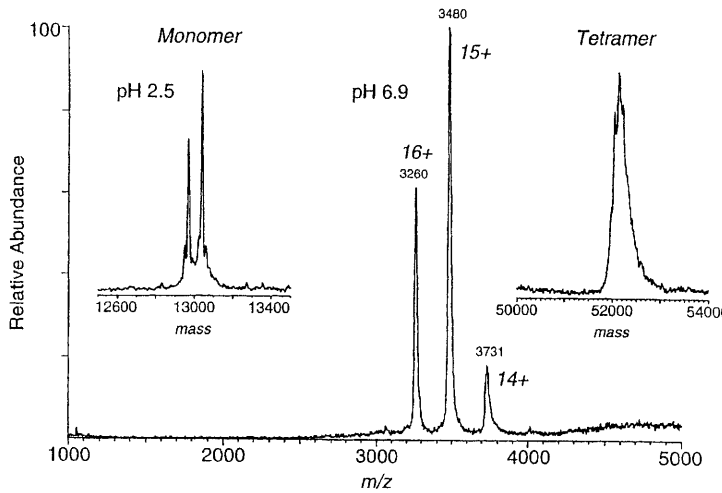

Figure 7.11 *ESI mass spectrum of streptavidin in 10 mM ammonium acetate (pH 6.9) at a concentration of 5 μm. Inserts show deconvoluted (by molecular weight) mass spectra at pH 2.5 and 6.9.*
(Source: J.A. Loo and K.A. Sannes-Lowery, in *Mass Spectrometry of Biological Materials*, 2nd edn, B.S. Larsen and C.N. McEwen (ed), Marcel Dekker, New York 1998, p. 358, Figure 5)

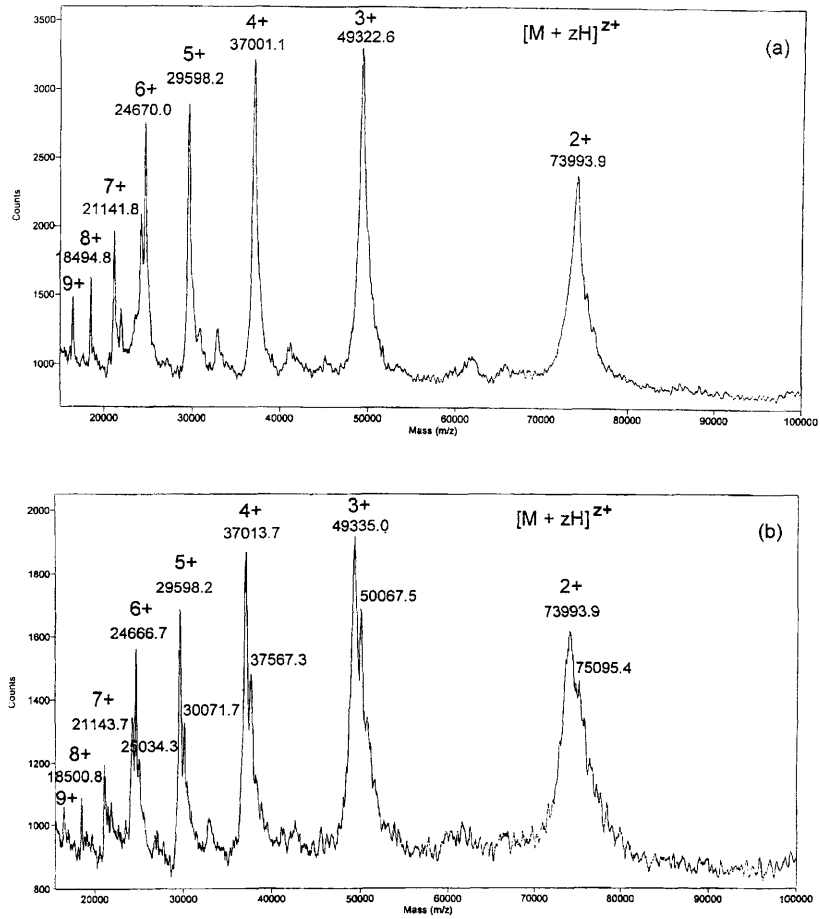

Figure 7.12 *MALDI mass spectra of the tryptic digest of the viral proteins from a type A influenza strain (a) before, and (b) after interaction with a monoclonal antibody raised to one of the proteins*
(Source: J.G. Kiselar and K.M. Downard, *J. Am. Soc. Mass Spectrom.*, 2000, **11**, 746–750, Figure 1)

An antigenic peptide that represents a surface domain of the hemagglutinin antigen of a type A influenza strain is found to preferentially bind to the antibody. Figure 7.12 shows the MALDI mass spectrum for the intact antibody before and after treatment with a mixture of peptides containing the hemagglutinin epitope. The molecular weight change that is evident in the additional ion signals associated with the antibody-peptide complex corresponds to the peptide antigen (MW 2210 Da).

A computer algorithm, known as COMPLX, has been recently developed which enables bimolecular protein and other macromolecular complexes to be identified in an automated manner from both ESI and MALDI mass spectra of the type shown in Figure 7.12. The program is of particular value for spectra that exhibit many hundreds of ion signals such that a manual interpretation of the data would be difficult or impossible. The ability to preserve protein complexes within a mass spectrometer in some circumstances is of potential value for identifying such associations in biological extracts.

Beyond such direct observations, a number of indirect approaches have made use of mass spectrometry to study protein complexes. Among these are limited proteolysis of protein complexes to release non-binding domains, hydrogen exchange studies, and the use of microfilters, chromatographic and native electrophoretic methods to separate or isolate large macromolecular complexes for their further characterisation by mass spectrometry.

7.2.6 Proteomics

The development of mass spectrometry for the study of proteins has led to its central role in *proteomics*. Proteomics, or proteome analysis, involves the identification and characterisation of the entire complement of proteins expressed in a single cell or tissue at any given point in time. *Functional proteomics* seeks to identify the protein components that are unique to diseased cells or tissue, those produced only in response to a genetic abnormality, or those of importance to a particular biological event or process.

Two primary methods are used to partially resolve the protein complement of a cell or tissue prior to mass spectrometric detection. The first of these is multi-dimensional chromatography in which proteins are partitioned according to their ionic character, molecular identity, size, and hydrophilicity using ion exchange, affinity, molecular exclusion and reverse-phase chromatography. These partitioned proteins, or their proteolytic products, are then introduced directly into an ESI-based mass spectrometer where they are characterised by molecular weight and sequence.

A second alternative strategy utilises two-dimensional gel electrophoresis to separate and display the proteins. Proteins isolated from the cell or tissue are separated in 2D gel electrophoresis in the first dimension according to their charge and in the second dimension by size. The proteins are typically digested in-gel to release their proteolytic peptides. These components are analysed by mass spectrometry in both MS and

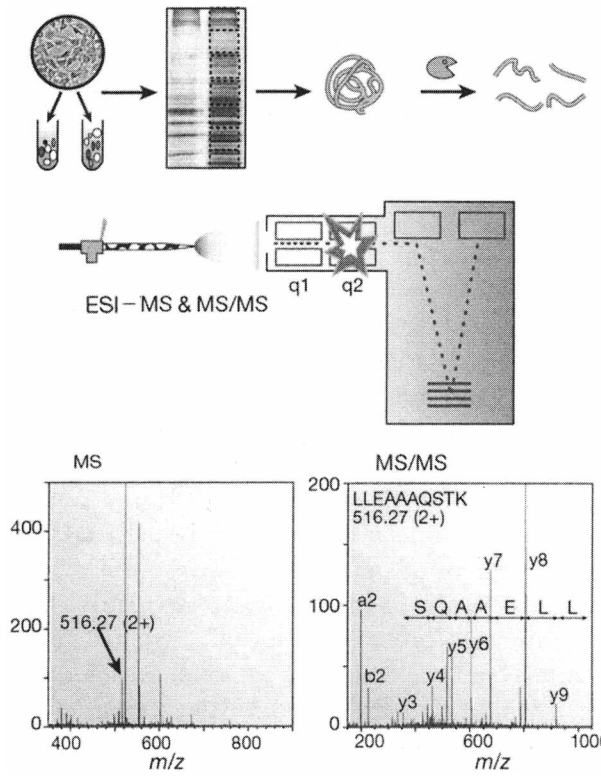

Figure 7.13 *A typical mass spectrometry (MS)-based proteomics experiment. Proteins recovered from cells are partially separated by SDS-PAGE, excised or digested in-gel, and the proteolytic peptides analysed on a tandem mass spectrometer. A Q-TOF hybrid tandem mass spectrometer is shown here (see Chapter 4, Section 4.7)*
(Source: adapted from R. Aebersold and M. Mann, Mass spectrometry-based proteomics, *Nature*, 2003, **422**, 198–207)

MS/MS experiments (Figure 7.13). Either an ESI or MALDI-based mass spectrometer can be used in these experiments.

Proteins are identified by mass spectrometry in the same manner as for single proteins. However, the use of protein and nucleotide databases are of particular importance to these investigations in order that previously characterised proteins, or their homologues, are rapidly identified. Mass map data as well as tandem MS/MS spectra of proteolytic peptides can be used to search a database for a known protein. Where tandem mass spectra are used, algorithms such as SEQUEST are capable of comparing a fragment ion profile with those for any hypothetical proteolytic peptide obtained from a known protein's sequence.

A major challenge confronted in proteomics is the management of large amounts of diverse data. Proteome analysis requires that data collected for each protein from within a cell or tissue be collated in terms of its recovery, electrophoretic or chromatographic profile, biochemical treatment, mass spectral appearance and bioinformatics discovery. Advanced computer-based bioinformatics software and systems are used for this purpose.

The field of proteomics has rapidly advanced due to the high performance and continual development of today's mass spectrometers. The discovery of new, previously uncharacterised proteins is continuing at a rapid rate. Beyond their identification at the molecular weight and sequence level, the quantitation of protein components in cellular lysates and biological extracts is also a core goal of proteomics. This represents a greater analytical challenge but one that has been addressed in part based on chromatographic detector responses, the image analysis of stained proteins on two-dimensional gels, and by mass spectrometry.

The chemical treatment of samples isolated from normal and diseased cellular extracts has been used to quantitate the relative levels of protein in each sample. In one protocol, an isotopically-enriched tag is reacted with the cysteine residue side-chains of one sample, while the same unlabelled reagent is reacted with those of the second. The two samples are then mixed and affinity chromatography is used to recover the tagged proteins. Mass spectra are then recorded for the recovered proteins where two ion signals (for the labelled and unlabelled forms) are mass resolved and detected. The relative area for these ion signals provides a quantititative measure of the levels of protein in each sample (Figure 7.14). Disadvantages of such an approach include the need to couple the reagant to proteins in each sample with equal efficiency and the ability to recover the proteins by affinity chromatography.

A particularly powerful application of MALDI-MS in proteomics is that involving a direct analysis of the spatial organisation (to a resolution of some 50 μm) of peptides and proteins in mammalian tissue sections. Caprioli and colleagues have demonstrated that it is possible to laser ablate compounds directly from a tissue section, or a blot of the tissue slice, to produce ions that span the range from m/z 1,000 to 100,000. Two-dimensional maps can then be reconstructed based on the intensity of these ions to provide, to a first approximation, the relative levels of certain molecules within the tissue (Figure 7.15). This mass spectrometric "imaging" of proteins and other components from tissue is still in its infancy but if well-developed would provide a powerful and rapid technique for use in both research and clinical settings.

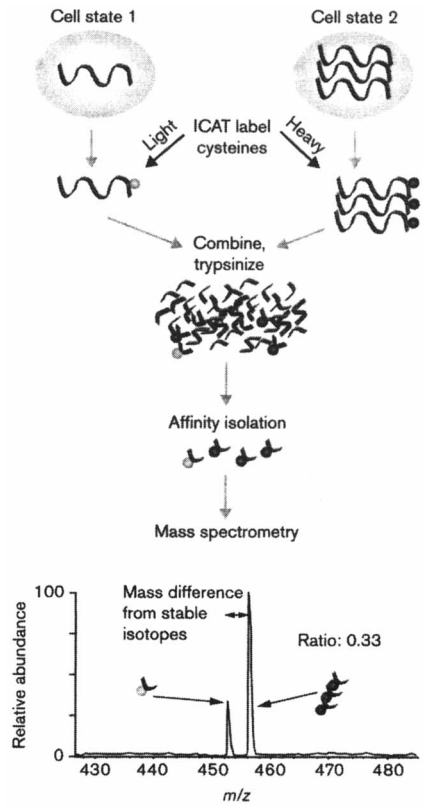

Figure 7.14 *The isotope-coded affinity tag (ICAT) approach to quantitate the relative levels of proteins from cellular extracts by mass spectrometry*
(Source: adapted from S.P. Gygi, R. Aebersold and M. Mann, Mass spectrometry and proteomics, *Curr. Opinion in Chem. Biol.*, 2000, Vol. 4, 489–494, Figure 2)

In a similar manner, MALDI mass spectrometry has been used to obtain protein profiles of unfractionated microorganisms including viruses, bacterial and fungal cells, and spores. The positive and negative MALDI mass spectra of proteins desorbed directly from the bacteria *Helicobacter pylori*, where 26995 intact cells were introduced into the mass spectrometer, are shown in Figure 7.16.

Such spectra have also been generated using laser ablation mass spectrometry where airborne microorganisms are introduced into the mass spectrometer as an aerosol. Microorganisms may also be identified, based on an analysis of the proteolytic peptides generated from their proteins in a mass map experiment.

A future goal of proteomics is to characterise the association of proteins *en masse* in cells and tissues. This will allow studies of

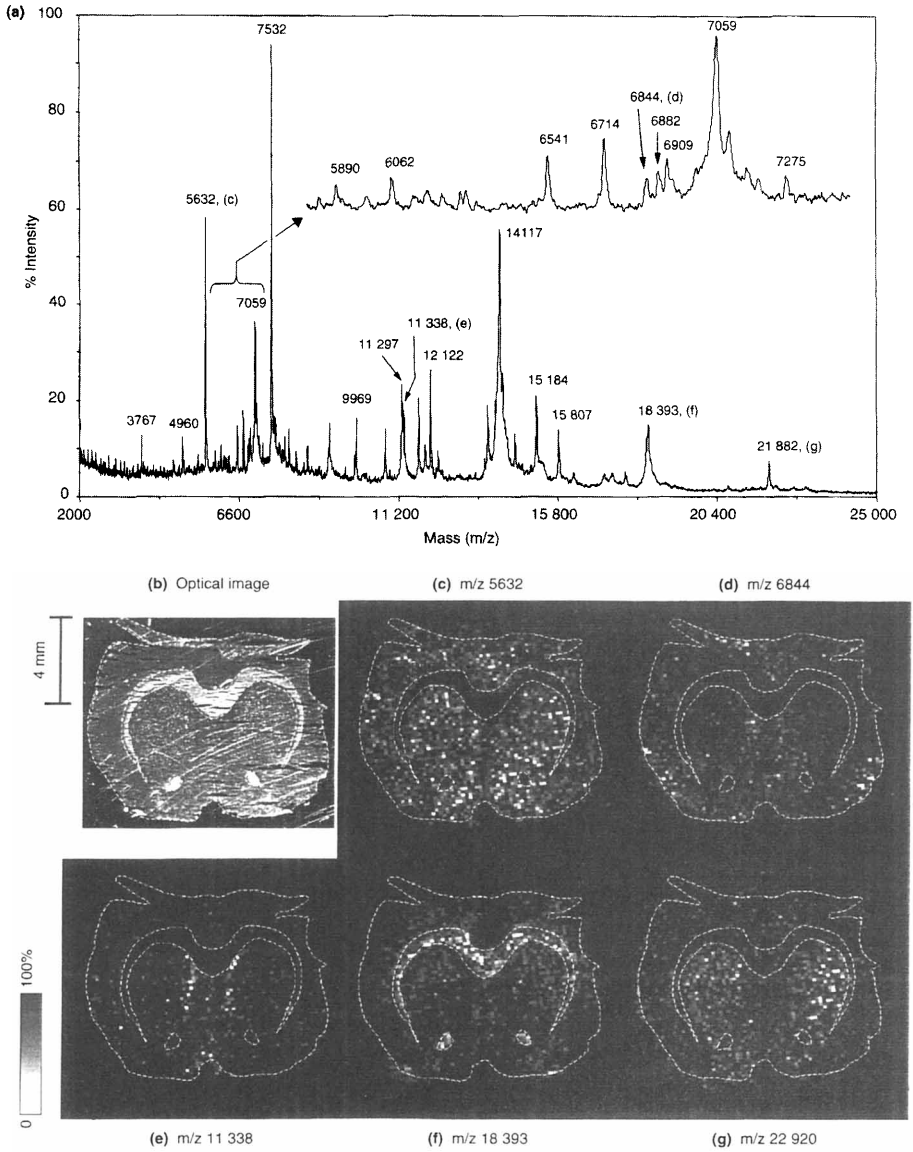

Figure 7.15 *MALDI-MS imaging of a transveral section of rat brain. (a) Survey profile taken randomly across the section, (b) optical image of the section before matrix application, (c)–(g) ion density maps obtained at different m/z values* (Source: P. Chaurand, S.A. Schwartz and R.M. Caprioli, Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections, *Current Opinion in Chemical Biology*, 2002, **6(5)**, 676–681, Figure 3)

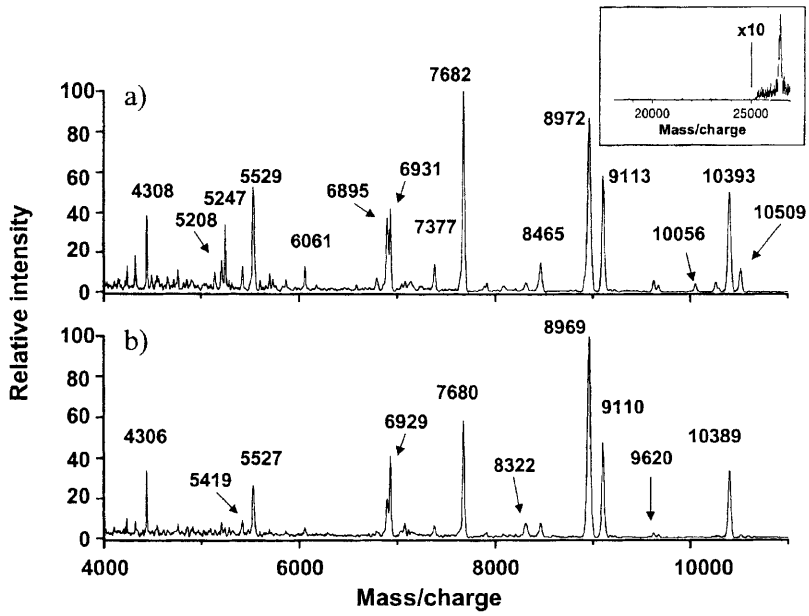

Figure 7.16 *Positive and negative MALDI mass spectra of proteins desorbed directly from the bacteria* Helicobacter pylori
(Source: P.A. Demirev, J.S. Lin, F.J. Pineda and C. Fenselau, Bioinformatics and mass spectrometry for microorganism identification: Proteome-wide post-translational modifications and database search algorithms for characterization of intact *H. pylori*, *Anal. Chem.*, 2001, **73(19)**, 4566–4573, Figure 1)

protein function to be followed on a global scale. Current genetic and biochemical methods are sure to be supplemented by the use of mass spectrometry for these endeavours, particularly since mass spectrometry is already in widespread use for the study of protein structure and interactions.

The use of MALDI mass spectrometry in this regard has been shown to be able to survey the structure and antigenic identity of the influenza virus. This has been achieved without the need to immobilise either the viral antigens or their cross-interacting antibodies. It has been found that the interaction between an antibody and a specific region (epitope) of a protein antigen can be preserved on a MALDI surface from which all non-binding peptides, generated after proteolysis, can be preferentially ionised. A careful comparison of the MALDI mass spectrum of the proteolysis products of an unreacted antigen mixture (the control) versus that of the antibody-containing mixture enables binding domains to be identified and characterised. Such spectra for the tryptic digest of all four viral antigens from a type A influenza strain are shown in Figure 7.17. A measure of the ion abundances in both spectra enable an

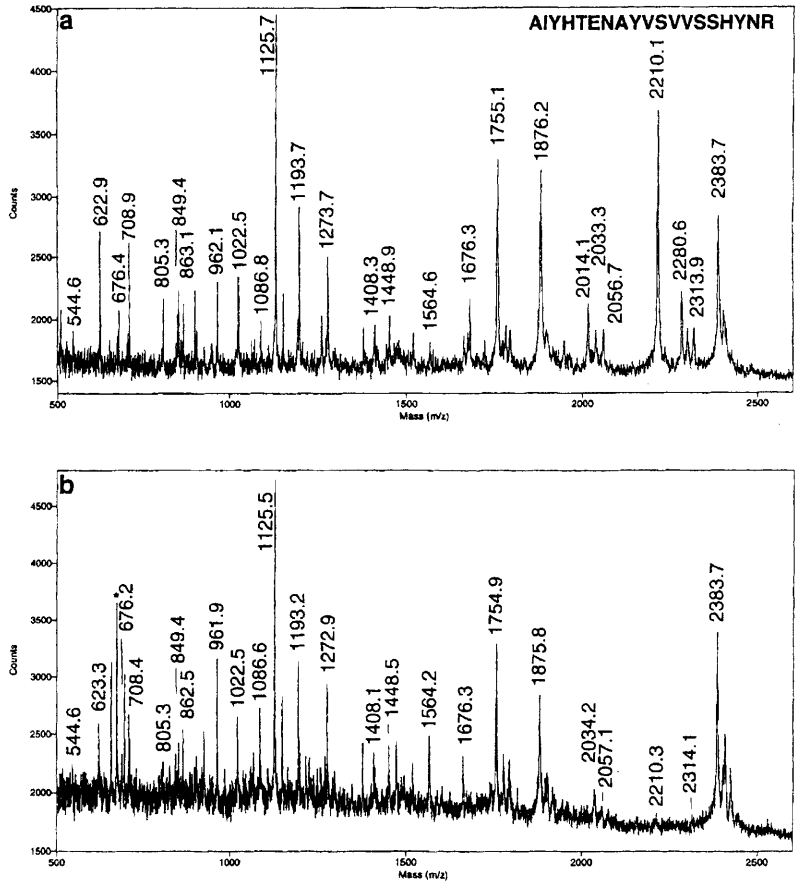

Figure 7.17 *MALDI mass spectra of the tryptic digest of the viral proteins from a type A influenza strain (a) before and (b) after reaction with monoclonal antibody. The peak labelled with an asterisk is attributed to a matrix cluster ion* (Source: J.G. Kiselar and K.M. Downard, *Biochemistry*, 1999, **38**, 14185–14191, Figure 2)

antibody-binding domain (*m/z* 2,210.1) to be identified within a single (hemagglutinin) antigen based upon its reduced intensity.

7.3 OLIGONUCLEOTIDES AND NUCLEIC ACIDS

Like peptides, oligonucleotides are linear polymers comprised of nucleoside monomers linked by a phosphodiester group. Conveniently, these polymers are composed of just four different natural monomer units: adenosine (A), guanosine (G), cytidine (C) and uridine (U) in the case of ribonucleic acids (RNA), and the deoxy forms of A,G,C and thymidine

(T) in the case of deoxyribonucleic acids (DNA) (see Appendix 9). This would appear to make the identification and sequencing of nucleic acids by mass spectrometry more straightforward than comparably sized proteins. In reality, however, oligonucleotides and nucleic acids are more fragile within a mass spectrometer than proteins and prone to degradation at the phosphodiester linking group. The larger mass of each of the mononucleotides also impacts the ionisation efficiency of oligonucleotides and nucleic acids and the ability to induce their dissociation in tandem mass spectrometry experiments. Nonetheless, mass spectrometry has been applied to the molecular weight and sequence analysis of oligonucleotides and nucleic acids.

7.3.1 Identification of Modified Nucleosides

Post-transcriptional and other cellular processes result in a wide range of structural modifications occurring at the purine and pyrimidine bases of RNA. Mass spectrometry has played a significant role in the identification of these modified bases, originally through the enzymatic hydrolysis of RNA. Such hydrolysates are analysed by either LC-ESI-MS, or GC-MS in the form of their volatile trimethylsilylated derivatives. The mass difference between the nucleoside and natural forms provides information concerning the type of modification(s). Tandem (MS/MS) mass spectrometry of the modified nucleosides can also be employed to enable the site of the modifications to be determined.

Although both RNA and DNA segments in excess of 100 bases have been successfully detected within mass spectrometers, they are typically ionised with less efficiency than proteins of a comparable size. This is associated with the fact that the negatively charged phosphodiester group shows both a propensity to dissociate and to adduct alkali metal cations such as Na^+ and K^+. These traits adversely impact the study of the low amounts of DNA and RNA that is usually available. As a consequence, mass spectrometry is yet to compete with some other analytical approaches for the detection and characterisation of nucleic acids. It does, however, offer a highly complementary approach and has proved useful for certain applications.

7.3.2 Sequencing of Oligonucleotides by Tandem Mass Spectrometry

High-resolution mass spectrometry can be used to verify the sequence of a synthetic or isolated RNA and DNA segment. To avoid contributions to their mass from metal ions, samples are desalted prior to analysis.

Like proteins, RNA and DNA can be digested into smaller oligonucleotides that are more amenable to analysis by the use of restriction endonucleases. These segments and synthetic oligonucleotides can be sequenced in the same manner as described for peptides (see Section 7.2.3.2). Tandem mass spectrometry has been applied to sequence simple short-chain oligonucleotides and a nomenclature has been proposed to describe the observed fragments (Figure 7.18).

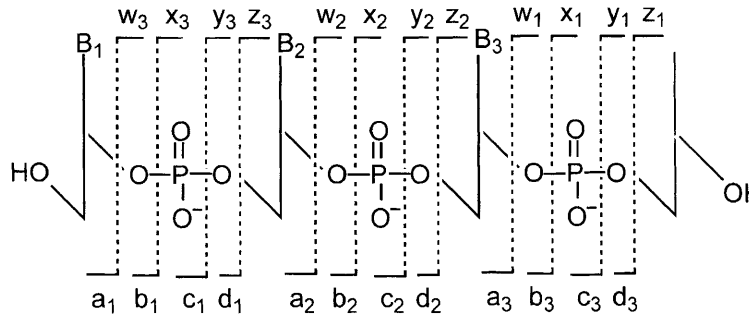

Figure 7.18 *Nomenclature to designate fragment ions detected upon the dissociation of oligonucleotides in a mass spectrometer*

The ESI tandem CID spectrum of oligonucleotide d(GCTGGCATCCGT) recorded in the negative ion mode is shown in Figure 7.19. Although MS/MS experiments are useful for particular applications, they fail to compete with automated high-throughput sequencing by means of the polymerase chain reaction (PCR).

7.4 OLIGOSACCHARIDES AND GLYCOCONJUGATES

Oligosaccharides and glycoconjugates are composed of monosaccharides such as glucose, hexose and fructose linked through glycosidic bonds. Determining the structure of these compounds is far more difficult than is the case for peptides and oligonucleotides. This is due to the isomeric nature of some monosaccharides and the fact that these monomers can couple in a multitude of ways to produce highly branched structures.

Like oligonucleotides, oligosaccharides and glyconjugates also ionise less efficiently than their equivalently sized peptide and protein counterparts. It is often necessary to analyse oligosaccharides in the negative ion mode due to the propensity of free hydroxyl groups to support a negative charge. Alkali metal ions often present in a sample may coordinate to these groups resulting in an unpredictable increase in the mass of the compound. To prevent this, glyconjugates can be methylated, acetylated or otherwise derivatised to facilitate their ionisation as positive ions

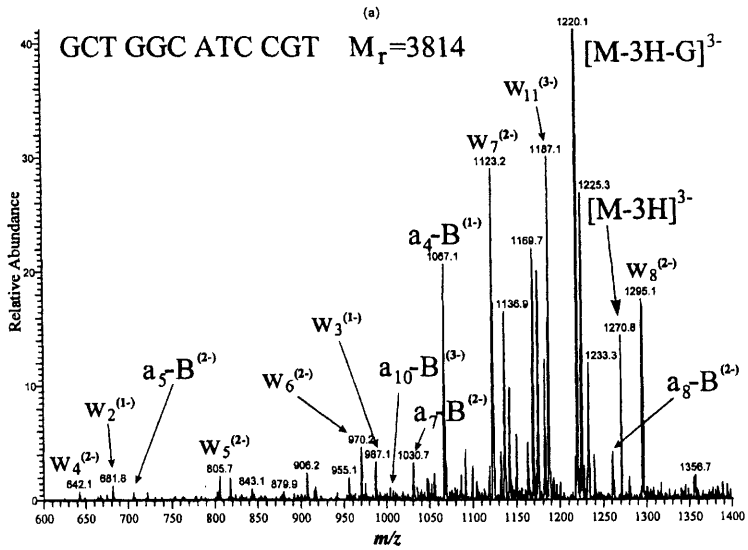

Figure 7.19 *ESI tandem mass spectrum of the* $[M - 3H]^{3-}$ *ion of oligonucleotide d(GCTGGCATCCGT) from which its sequence can be derived*
(Source: R.H. Griffey, M.J. Greig, H.J. Gaus, K. Liu, D. Monteith, M. Winniman, L.L. Cummins, Characterisation of oligonucleotide metabolism in vivo via liquid chromatography/electrospray mass spectrometry with a quadrupole ion trap mass spectrometer, *J. Mass Spectrom.*, 1997, **32**, 305–313.)

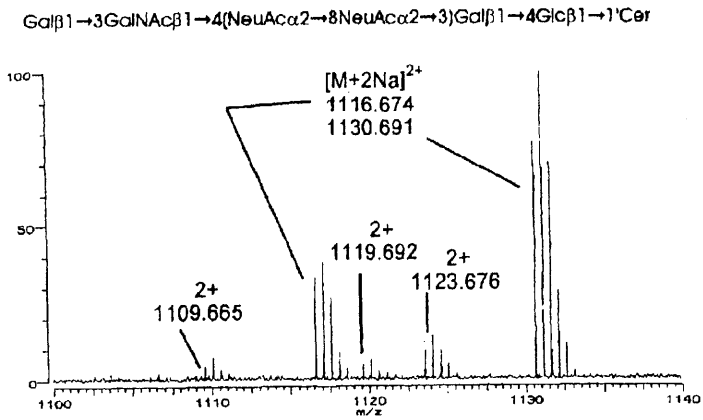

Figure 7.20 *Molecular ion region of the ESI FT-ICR mass spectrum of permethylated ganglioside* G_{D1b}
(Source: C.E. Costello, Bioanalytic applications of mass spectrometry, *Current Opinion in Biotechnology*, 1999, **10(1)**, 22–28, Figure 2a)

(Figure 7.20). This method of analysis proved mandatory prior to the development of the newer desorption ionisation methods (such as FAB and MALDI) in order to volatilise oligosaccharides.

Heterogeneity is often encountered in glyconjugate samples where individual components differ in molecular weight by that of a monosaccharide (Figure 7.21). A molecular weight profile of the sample provides an immediate indication as to its complexity and may identify potential heterogeneity among the components. Hydrolysis of the sample prior to analysis can allow the component monosaccharides to be identified though care must be taken to avoid modifying groups being removed in this process. Glycosidases are also used to evaluate such samples as described later in section 7.4.2.

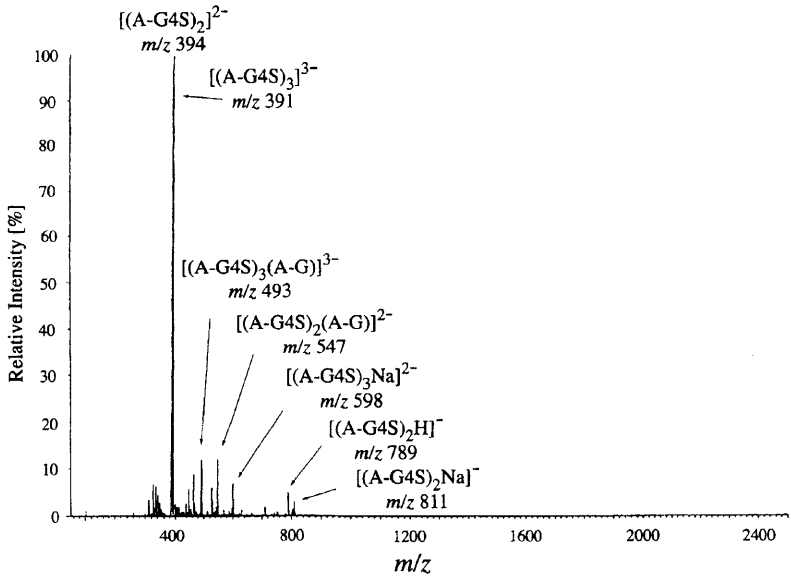

Figure 7.21 *ESI mass spectrum of the enzymic hydrolysate of kappa-carrageenan containing ions associated with the tetrasaccharide (A-G4S)$_2$ and hexasaccharide (A-G4S)$_3$*
(Source: D. Ekeberg, S.H. Knutsen and M. Sletmoen, Negative-ion electrospray ionisation-mass spectrometry (ESI-MS) as a tool for analysing structural heterogeneity in kappa-carrageenan oligosaccharides, *Carboyhydrate Research*, 2001, **334(1)**, 49–59, Figure 4)

7.4.1 Sequencing of Oligosaccharides by Tandem Mass Spectrometry

The dissociation of oligosaccharides within a tandem mass spectrometer results primarily in the cleavage of the glycosidic bonds. A nomenclature

has been proposed to designate the possible fragment ions formed by the dissociation about glycosidic bonds in which charge is retained at either the reducing or non-reducing terminus resulting in the production of B, C, Y and Z ions (Figure 7.22). A numbered subscript identifies the number of monomeric units toward the termini where the B and C ions contain the non-reducing terminus and Y and Z ions contain the reducing terminus. The Greek letters (α, β, *etc.*) are used to denote the branch position by a subscript. Two additional fragments A and X are a result of ring cleavage. These ions are designated with numerical superscripts to denote the bonds broken within the ring. For instance, the ion nomenclature $^{1,3}A_{2\alpha}$ denotes a fragment formed from the cleavage of the first and third bonds of the monosaccharide unit (in a clockwise direction from the O-C bond, denoted bond zero) at the first two sugars from the non-reducing terminus along the first (α) branch.

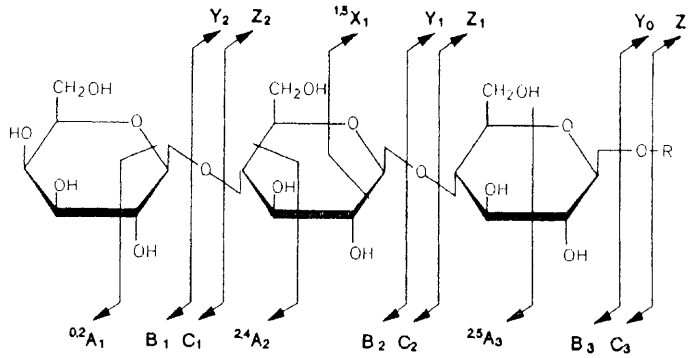

Figure 7.22 *Nomenclature to designate the fragments ions formed by the dissociation of oligosaccharides in a mass spectrometer*
(Source: J. Vath and C.E. Costello, in *Methods in Enzymology*, McCloskey (ed), Academic Press, New York, 1990 Vol. 193, Ch. 40, p. 743, Figure 2)

As for peptides and oligonucleotides, the sequence and structure of oligosaccharides can be assembled from the *m/z* values of their fragment ions. Appendix 10 shows the mass increments for common monosaccharide units, representing the molecular weight of a monosaccharide less 18 *u* for a molecule of water. To illustrate this process, the tandem mass spectrum of a pentasaccharide recorded in the negative ion mode is shown in Figure 7.23.

Fragment ions that result from glycosidic bond cleavages are evident in this spectrum in addition to those formed by ring cleavage (A and X ions). The appearance of abundant $^{2,4}A_3$ and $Y_{3\alpha}$ ions at *m/z* 545 and 586 enables the branched nature of the structure to be deduced.

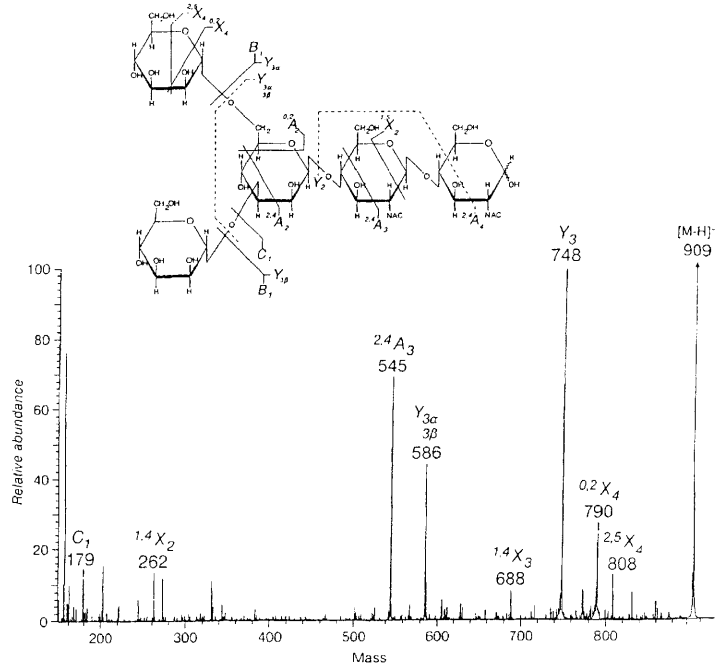

Figure 7.23 *Negative ion CID tandem mass spectrum of a heptasaccharide from N-linked glycans*
(Source: D.L. Gillece-Castro and A.L. Burlingame, in *Methods in Enzymology*, McCloskey (ed), Academic Press, New York, Vol. 193, Ch. 37, p. 689, Figure 3)

7.4.2 Exoglycosidase Digestion

An alternate strategy to tandem mass spectrometry is to digest the oligosaccharide with a series of exoglycosidases and record the molecular weights of the resulting products. A number of commercial exoglycosidases are available that show a range of specificities for oligosaccharide cleavage. Sialidases, galactosidases and mannosidases can be employed to cleave glyosidic bonds at specific monosaccharides thus releasing these components. Figure 7.24 represents in cartoon form how glycosidases can be used to distinguish between alternate biantennary and triantennary structures by MS analysis after stepwise digestion.

7.4.3 Derivatisation Approaches: Oxidative and Reductive Cleavage to Identify Branching

There are many occasions where it is desirable, even necessary, to derivatise an oligosaccharide prior to its analysis. This is conducted

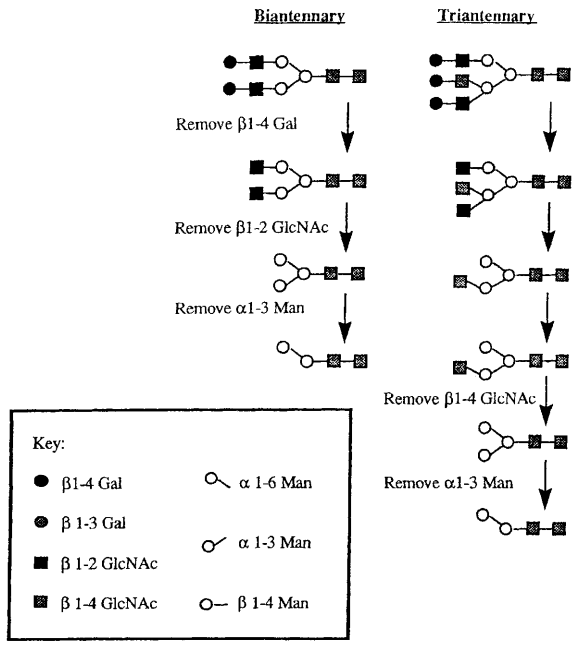

Figure 7.24 *Exoglycosidase digestion used to determine the structure of branched oligosaccharides*
(Source: adapted from R. Orlando and Y. Yang, in *Mass Spectrometry of Biological Materials*, 2nd edn, B.S. Larsen and C.N. McEwen (ed), Marcel Dekker, New York, USA, 1998, Ch.9, p. 236, Figure 11)

in order to improve the volatility of the molecule, increase the yield of parent ions, direct its fragmentation in a more predicatable manner or assist with the investigation of particular structural features. Methylation and acetylation reactions are among the most common derivatisation reactions employed. These reactions, however, show little specificity for functional groups within the sugar and as such are generally used to completely derivatise all such groups (known as permethylation and peracetylation) to improve the molecule's ionisation efficiency in the positive ion mode.

Periodate oxidation provides a more selective strategy for probing structural elements within an oligosaccharide. The reaction involves the cleavage of C-C bonds and allows for the branching pattern within O-linked oligosaccharides to be determined. This is illustrated in Figure 7.25 where a 1,4-linked and 1,6-linked hexose unit undergoes ring opening following oxidation and reduction to yield monosaccharide-units of different mass (208 and 164 u respectively) after methylation.

The same chemistries can be used to determine the anomeric configuration of glycosidic bonds by mass spectrometry. It is possible to

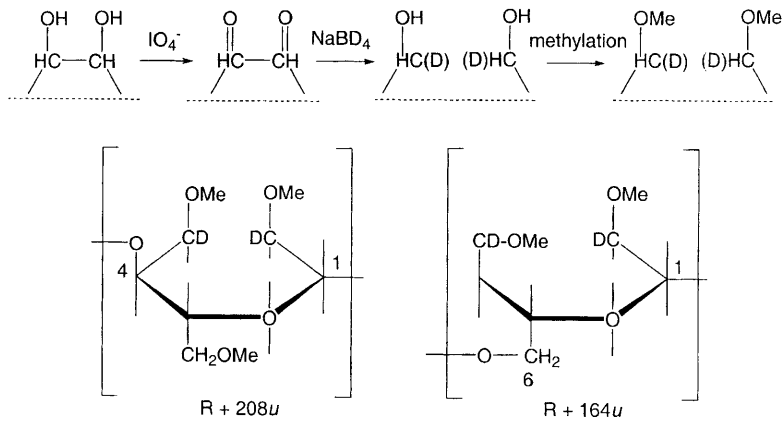

Figure 7.25 *Identification of 1,4-linked and 1,6-linked hexose after ring opening by periodate oxidation, reduction and methylation*

selectively oxidise the β-anomer of hexose, over its α counterpart, leading to a mass shift being detected in this region of the oligosaccharide (Figure 7.26).

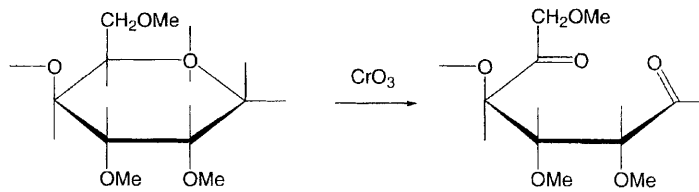

Figure 7.26 *Oxidisation of the β-anomer of hexose, that occurs selectively over its α counterpart, can be identified by a mass shift*

FURTHER READING

G. Siuzdak, *Mass Spectrometry for Biotechnology*, Academic Press, 1996.

K.M. Downard Advances in protein analysis and sequencing by mass spectrometry, *New Adv. Anal. Chem.*, 2000, **P2**, 1–30.

A.G. Marshall, C.L. Hendrickson and S.D.H. Shi, Scaling MS Plateaus with FTICR MS, *Anal. Chem.*, 2002, **74**, 252A–259A.

W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley and C. Watanabe, Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 5011–5015.

J.R. Yates III, S. Speicher, P.R. Griffin and T. Hunkapiller, Peptide mass maps: A highly informative approach to protein identification, *Anal. Biochem.*, 1993, **214**, 397.

M. Mann, P. Hojrup and P. Roepstorff, Use of mass spectrometric molecular weight information to identify proteins in sequence databases, *Biol. Mass Spectrom.*, 1993, **22**, 338–345.

K. Biemann and I.A. Papayannopoulos, Amino acid sequencing of proteins, *Acc. Chem. Res.*, 1994, **27(11)**, 370–378.

I.A. Papayannopoulos, The interpretation of collision-induced dissociation tandem mass spectra of peptides, *Mass Spectrom. Rev*, 1995, **14(1)**, 49–73.

J.R. Engen, D.L. Smith, Analysis of proteins with hydrogen exchange and mass spectrometry, *Anal. Chem.*, 2001, **73**, 256A-265A.

S.D. Maleknia, K.M. Downard, Radical approaches to probe protein structure, folding, and interactions by mass spectrometry, *Mass Spectrom. Rev.*, 2001, **20(6)**, 388–401.

J.A. Loo, Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom. Rev.*, 1997, **16(1)**, 1–23.

J.G. Kiselar and K.M. Downard, Antigenic Surveillance of the Influenza Virus by Mass Spectrometry, *Biochemistry*, 1999, **38(43)**, 14185–14191.

M.R. Wilkins, K.L. Williams, R.D. Appel and D.F. Hochstrasser, *Proteome Research: New Frontiers in Functional Genomics*, Springer Verlag, Berlin, Germany, 1997.

J.A. McCloskey, A.B. Whitehill, J. Rozenski, F. Qiu and P.F. Crain, New techniques for the rapid characterization of oligonucleotides by mass spectrometry, *Nucleosides & Nucleotides*, 1999, **18**, 1549–1553.

B. Reinhold, V. Reinhold and C.E. Costello, Carbohydrate molecular weight profiling, sequence, linkage and branching data: ESI-MS and CID, *Anal. Chem.*, 1995, **67**, 1772–1784.

K.H. Khoo and A. Dell, Assignment of anomeric configurations of pyranose sugars in oligosaccharides using a sensitive FAB-MS strategy, *Glycobiology*, 1990, **1** 83–911.