ELSEVIER

Journal of Chromatography A, 705 (1995) 21–45

JOURNAL OF CHROMATOGRAPHY A

Review

Protein mass spectrometry: applications to analytical biotechnology

D.N. Nguyen*, G.W. Becker, R.M. Riggin

Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA

Abstract

The advent of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and electrospray ionization mass spectrometry (ESI-MS) in the last 5 years has greatly enhanced the area of protein mass spectrometry. This paper presents an overview of the applications of protein mass spectrometry in the area of analytical biotechnology, particularly as related to biopharmaceutical research and development. These applications include the determination of protein molecular mass, peptide mapping, peptide sequencing, ligand binding, determination of disulfide bonds, active site characterization of enzymes, protein self-association and protein folding/higher order structural characterization.

Contents

1. Introduction 22
2. Modes of mass spectrometry 22
   2.1. Fast atom bombardment mass spectrometry (FAB-MS) 22
   2.2. Plasma desorption mass spectrometry (PD-MS) 24
   2.3. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) 25
   2.4. Electrospray ionization mass spectrometry (ESI-MS) 26
3. Protein applications 28
   3.1. Molecular mass determination 28
   3.2. Peptide mapping/MS 29
      3.2.1. Confirmation of sequence 30
      3.2.2. Post-translational modifications 31
      3.2.3. Protein degradation products 33
      3.2.4. Protein metabolites 34
      3.2.5. Disulfide bond determination 34
      3.2.6. Ligand binding 35
      3.2.7. Enzyme active sites 35
      3.2.8. Peptide mass maps 36
   3.3. Sequencing 36
      3.3.1. Tandem mass spectrometry 36
      3.3.2. Post-source decay MALDI-MS 38

* Corresponding author.

0021-9673/95/$29.00
SSDI 0021-9673(94)01256-3

3.3.3. Carboxypeptidases 38
3.4. Non-covalent interactions 39
3.5. Protein folding/higher order structural characterization 41
3.5.1. Deuterium exchange 41
3.5.2. Proteolysis 42
3.5.3. Charge state 42
3.5.4. Cross-linking reagents 42
References 43

1. Introduction

Various forms of mass spectrometry having the ability to analyze protein structures have been introduced over the past 10–15 years. These techniques have served a central role in the advancement of biotechnology during that time period. The advent of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and electrospray ionization mass spectrometry (ESI-MS) in the last 5 years has greatly enhanced the area of protein mass spectrometry. In many areas of research and development (e.g. industrial biotechnology) these techniques are now indispensable tools for examining protein and peptide structures.

Protein mass spectrometry has been the subject of several extensive reviews [1–4], and we therefore do not intend to include a comprehensive review of the field in this paper. Instead, we intend to present an overview of the applications of protein mass spectrometry in the area of analytical biotechnology, particularly as related to biopharmaceutical research and development. Ideally, this examination will prove a useful guide to the general types of protein structural studies which can be addressed by mass spectrometry.

The paper consists of two major sections. In the first section the fundamental aspects of the various modes of protein mass spectrometry are only briefly described since they have been discussed in great detail elsewhere. In the second section various application areas in which mass spectrometry of proteins serves a useful role are discussed. These applications include the determination of protein molecular mass, peptide mapping, peptide sequencing, ligand binding, determination of disulfide bonds, active site characterization of enzymes, protein self-association and protein folding/higher order structural characterization. Each of these applications represents an important research area where many tools can be used, but in which mass spectrometry is becoming increasingly significant.

2. Modes of mass spectrometry

The development and application of mass spectrometry to the analysis of organic molecules began in the late 1950s and early 1960s. During this time period, electron impact (EI) ionization was the only practical ionization technique [5]. The development of chemical ionization (CI) in the late 1960s provided a complementary method for ionization of organic molecules in the gaseous phase [6]. A disadvantage of both methods is that the sample analytes must be in the vapor phase before ionization can occur, often requiring heat for vaporization. Because of their size, involatility and thermal lability, proteins are generally not candidates for study using EI or CI methods.

2.1. Fast atom bombardment mass spectrometry (FAB-MS)

The role of mass spectrometry in protein characterization underwent explosive growth in 1981 when Barber and co-workers [7,8] introduced fast atom bombardment mass spectrometry. The application of FAB-MS allowed scientists to obtain molecular masses of peptides and small proteins routinely and, sometimes, limited structural information was attainable.

In a typical FAB-MS analysis, the sample is dissolved in a non-volatile, viscous matrix such

as glycerol, thioglycerol or a mixture of dithiothreitol and dithioerythritol. The sample matrix mixture, usually about 1 μl, is then applied to a sample probe tip and directly inserted into the mass spectrometer. The sample matrix is bombarded with a high-energy (about 6–8 keV) beam of xenon atoms (Fig. 1).

The mechanism of sample desorption and ionization is not completely understood and is still being investigated. It is likely that a combination of the following mechanisms is responsible for ionization. There are currently two equally favorable mechanisms in describing the ionization process. The first requires that the ions are preformed in the matrix through protonation and complexation with metal ions (Na^+, K^+) [9]. The second theory proposes that the sample is desorbed as neutral molecules which undergo gas-phase ionization in the high-pressure region directly above the matrix–vacuum interface [10]. In practice, the mechanism that predominates the ionization process depends on many factors such as the presence of metal ions in the sample and/or matrix, the concentration of the sample in the matrix and the pK_a of the analyte molecules. In many instances, sample molecules which have extensive hydrogen bonding with the matrix tend to form a homogeneous solution with no concentration of the analyte in a particular locus within the droplet. On the other hand, hydrophobic analytes tend to concentrate on the surface of the droplet. Thus, signal suppression of hydrophilic peptides is often observed in FAB-MS during the analysis of a peptide mixture [11]. This limits one from relying on the FAB mass spectrum of a mixture to determine the concentrations of the analytes present in the sample unless internal standards are carefully applied. Although FAB-MS can be used to analyze a mixture of peptides, practical consideration must be taken. The sample mixture must be free of salt and buffer, hence it should be purified by HPLC or other suitable chromatographic techniques. Peptides should be stored in polypropylene vials to prevent complexation with metal ions. Complete suppression of the analyte signal may be observed if the sample is contaminated with a high concentration of alkaline salts or buffer. When there is a need to concentrate a peptide solution, it must not proceed to dryness since irreversible adsorption to the tube walls is likely. It is also good practice to avoid contact with any glass surface during sample preparation.

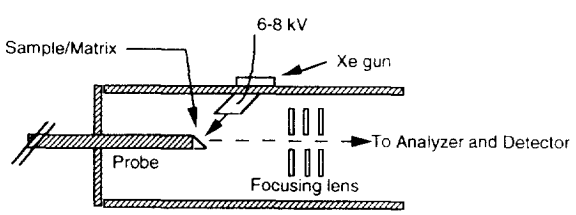


Fig. 1. Ionization source for FAB-MS.

Several drawbacks associated with FAB-MS should also be pointed out. The most obvious drawback is the presence of background peaks in the mass spectrum. The background signals are present throughout the mass spectrum, but it is particularly problematic in the low-molecular mass region (below 300 u). Simply switching to a different matrix allows one to discriminate the background from the analyte signals. A different drawback results directly from the nucleophilicity of the matrix components. For example, dithiothreitol could undergo an S_N2 reaction with samples containing good leaving groups such as halides. An experimental indication that the analyte is reacting with the matrix can be obtained by analyzing the sample using different matrices. If different peak patterns are obtained for different matrices, then it is likely the sample has undergone chemical reactions with the matrix. Other chemical reactions which could occur in the matrix include the reduction of disulfide bonds, the exchange of sulfur by oxygen atoms, ring opening of lactones and hydrolysis [12].

A related technique, continuous-flow FAB (CF-FAB), overcomes some of these problems. In CF-FAB, the sample is introduced continuously by a flow of volatile solvents such as water, methanol and acetonitrile. The use of these volatile solvents eliminates much of the matrix-derived background, thereby lowering the detection limit. CF-FAB also provides mechanical mixing of the sample and ultimately decreases the ion suppressing effect [13]. This technique

allows the analysis of a large number of aqueous samples and can be used to monitor the purity of a synthetic peptide.

Perhaps the greatest drawback of FAB and CF-FAB is the inability to generate good signal intensity with proteins of M_r greater than about 15 000, owing to involatility and thermal instability of proteins, thereby limiting the general applicability of this technique to large proteins. The use of liquid secondary ionization mass spectrometry (LSIMS), where the xenon beam in FAB-MS is replaced with a stream of ions such as Ar^+ or Cs^+, has shown some success with proteins in the M_r range 15 000–20 000. However, these results are far from routine and, in practice, FAB works best in the analysis of proteins with molecular masses of no more than about 5000.

2.2. Plasma desorption mass spectrometry (PD-MS)

The first breakthrough in attempts to solve the problems associated with protein mass spectrometry was the development of field desorption mass spectrometry [14]. In 1974, Friedman and co-workers proposed that rapid heating of a sample could lead to preferential desorption of surface molecules [15]. At the same time, Macfarlane and co-workers discovered that when high-energy fission fragments from ^{252}Cf irradiated thin films of arginine and cysteine, intact molecular ions were observed [16]. Thus, a new ionization technique, plasma desorption, was born. Today, PD-MS is used primarily for molecular mass determination of proteins. This technique provides very little fragmentation, and therefore little structural information can be obtained. The sensitivity of this technique can be impressive with examples of spectra obtained on proteins of M_r about 45 000 (ovalbumin) at the picomole level [17]. Perhaps one of the most useful applications of this technique is its use for peptide mapping by in situ enzymatic digestion [18]. This procedure is normally performed with the same nitrocellulose-bound sample as already used for molecular mass determination. An enzyme solution is applied and after an appropriate time the digestion is terminated by removal of buffer by spin-drying the target. The target containing the digested peptide fragments can then be analyzed by PD-MS.

PD-MS has been a particularly attractive method because of its operational simplicity and high reliability. A schematic diagram of a PD mass spectrometer is shown in Fig. 2. PD-MS is based on spontaneous fission of ^{252}Cf which produces a pair of nuclear fragments such as $^{144}Cs^{20+}$ and $^{106}Tc^{22+}$. These fission fragments, with MeV energy, move in opposite directions. One of these fragments can be oriented to hit the sample foil and ionize the sample molecules. The sample foil containing the adsorbed protein, often referred to as the target, is then mounted on a wheel assembly which can accommodate several samples at one time. The target wheel is positioned on the axis of the cylindrical tube called the flight tube. The ions are accelerated by a grid maintained at ground potential. The main function of the fission fragment detector is to record the time of the fission. The masses of the ions are calculated from the time of a fission event and the time it takes for an ion to reach the detector. The time an ion takes to traverse the flight tube is dependent on its mass/charge ratio. Both detectors produce an electronic pulse when an ion is detected.

In principle, a time-of-flight (TOF) mass spectrometer has no upper mass limit. However, there are several factors which limit the range of samples that can be analyzed by PD-MS. It is clear that the energy density developed by the

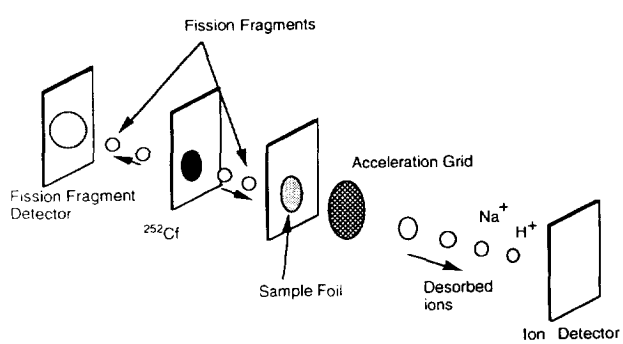


Fig. 2. Schematic diagram of plasma desorption mass spectrometer.

fission of ^{252}Cf is too high and that most of the protein molecules undergo pyrolysis. This problem is very apparent when the molecular mass of the protein approaches 50 000. In this mass range, very few intact molecular ions are detected. Since most of the plasma desorption ion sources are coupled to a TOF analyzer (typical resolution of about 1000), the mass accuracy is far from desirable. This, perhaps, results from the large kinetic energy distribution the ion carries with it [19]. Broadening of the peak to the extent that the isotopic pattern cannot be resolved is a common phenomenon in PD-MS. Therefore, the mass value derived from a PD-MS is often the isotopically averaged value.

2.3. *Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)*

The development of modern laser technology has provided a means of directing a large amount of energy into a sample, leading to desorption of intact molecules rather than thermal decomposition. Matrix-assisted laser desorption ionization was introduced by Tanaka et al. [20] and Karas and Hillenkamp [21]. Two ranges of laser wavelengths, the far-infrared and the far-ultraviolet, are used for desorption of sample molecules. The most commonly used wavelength is between 266 and 366 nm, which is generated from a neodymium/yttrium–aluminum–garnet (Nd:YAG) laser. The reason for choosing this wavelength is that compounds containing π-bonds can be electronically excited. With regard to mass spectrometry, lasers offer two important benefits. First, lasers provide the capability to pulse from a continuous wave down to femtosecond (10^{-15} s). Most laser mass spectrometers use pulses of 100 ns or less to prevent pyrolysis of the proteins. Second, laser beams can be focused to submicrometer diameters, allowing the operator to control the laser beam precisely with an appropriate microscope. The amount of energy deposited in the sample is dependent on laser irradiance or intensity (W cm $^{-2}$), the pulse width and the absorptivity of the sample. Increasing the laser intensity will lead to an increase in desorption; however, for large proteins too high an intensity will certainly lead to excessive fragmentation.

In MALDI-MS, the sample molecules are mixed with a suitable matrix (1:10 000 ratio). The commonly used matrices are 3-methoxy-4-hydroxycinnamic acid (ferulic acid) or 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) dissolved in a mixture of acetonitrile and 0.1% trifluoroacetic acid (TFA). The final concentration of the matrix is 5–10 g/l. The matrix–analyte mixture is deposited on a silver support and allowed to crystallize by slow evaporation of the matrix solvents. Proper sample preparation is essential in obtaining a good spectrum. Co-crystallization of sample and matrix is critical and will depend on the physical properties of both sample and matrix. The presence of TFA has been found to help achieve the co-crystallization. As mentioned previously, the ability to focus the laser beam precisely on the target allows the accumulation of successive spectra from the same and/or different areas. A simple schematic diagram of a MALDI instrument is shown in Fig. 3. The impact of a laser pulse on the target desorbs a large number of ions, which can cause a saturation effect in the detector. The high yields of low-mass ions saturate the detector, which results in lower efficiency for the high-mass ions because the detector cannot recover fast enough. The solution to this problem is the installation of an ion deflector which deflects the low-mass ions. The detector is also located slightly off-axis from the mass spectrometer. This combination reduces about 80% of the low-mass ions. The low-mass ions dominate the M_r 0–500

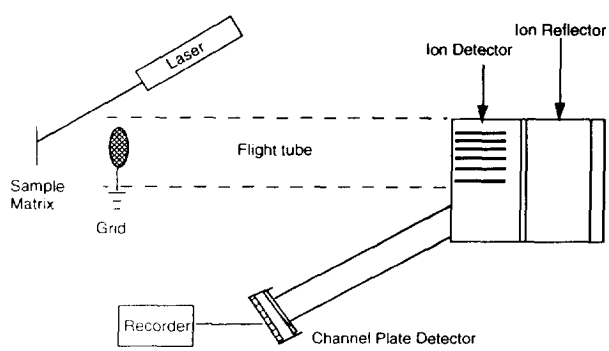


Fig. 3. Schematic diagram of MALDI mass spectrometer.

range and can extend up to 1000. However, small noise background can be seen across the mass spectrum. The low-level noise background is probably the result of fragmentation due to ions colliding with the surfaces in the spectrometer.

The resolution of a MALDI mass spectrometer is much less than desirable, in part due to the use of a TOF mass analyzer. An interesting, but not surprising, fact is that the resolution for low-mass ions is much better than that for high-mass ions. This effect is the result of the spread of the initial kinetic energy of the ions, and also the existence of non-resolved adduct ions [22]. The low resolution of high-mass ions results in uncertainty in mass determination. The uncertainty is about 10–50 u at M_r 10 000 and 100–500 u for proteins of M_r 100 000.

Despite the poor resolution, sensitivity is a major advantage of MALDI-MS. A good signal-to-noise spectrum can be obtained with a few laser pulses from a single spot on the sample target. Since successive spectra can be obtained from a single spot, it can be concluded that this technique is non-destructive. In practice, the actual amount of sample consumed is in the femtomole range. Therefore, for samples available only in limited amounts, the sample can be recovered after the analysis. It should be pointed out that picomole amounts of sample are normally required to form uniform matrix–analyte crystals for the analysis.

A related technique which is still at an early developmental stage is continuous-flow MALDI (CF-MALDI). In CF-MALDI, the sample is introduced to the mass spectrometer via a probe with a liquid matrix (aqueous TFA, ethanol, ethylene glycol, and 3-nitrobenzyl alcohol). The laser beam is set on the opposite side of the probe with the flight tube perpendicular to the probe. In this present configuration, mass spectra of large peptides with molecular masses above 10 000 can be obtained [23–26].

2.4. Electrospray ionization mass spectrometry (ESI-MS)

The electrospray ionization process was originally described by Dole et al. [27] in their studies on synthetic and natural polymers of molecular mass in excess of 100 000. However, it was not until 10 years later that electrospray resurfaced as a major technique in mass spectrometry. The application of electrospray to mass spectrometry was largely the result of two different groups almost simultaneously. Yamashita and Fenn [28] coupled atmospheric pressure electrospray to a quadrupole mass spectrometer and Alexandrov et al. [29] coupled it to a magnetic sector mass spectrometer. A few years after the demonstration of the fundamental aspects of electrospray, Fenn and co-workers demonstrated the ability of electrospray to analyze high molecular mass samples such as polyethylene glycol (M_r 17 500) bearing a net charge of up to +23 [30]. It is this unique feature of producing ions with multiple charges that allows electrospray to be used in the analysis of large proteins in a mass analyzer with limited mass range ($m/z < 2500$ u) with an accuracy of better than 0.01% [31,32].

Electrospray is produced by applying a high electrical field to a relatively small flow of liquid from a capillary tube. The electric field causes the liquid surface to be highly charged and a spray of charged liquid droplets forms at the end of the capillary tube. The polarity of the charged droplets can be controlled by the applied polarity on the capillary. The mechanism by which the molecular ions are formed from the charged droplets is not fully understood. Iribane and Thomson [33] proposed the field-assisted ion evaporation model. In this model, the molecular ion formation occurs when the field strength at the surface of the droplet reaches a critical value due to the evaporation of the solvent. Röllgen and co-workers proposed a different mechanism in which the disintegration of the charged droplet occurs via the Rayleigh jet mechanism that leads to very small charged droplets [34]. The bare molecular ions are formed when complete evaporation of the solvent from these micro-droplets occurs. Abbas and Latham [35] showed in a very elegant experiment that the droplets evaporate at a constant rate until the Rayleigh limit (1.1×10^5 elemental charges) is reached. The droplet then abruptly loses 20–25% of its charge and mass. This process is presumed to continue until a bare molecular ion is formed. A

particle of 1 μm requires about sixteen disintegrations to reach the ion evaporation limit. At this point, molecular ion production from the droplet can theoretically occur.

In the development of the electrospray mass spectrometer, transmitting the ions produced at atmospheric pressure into the vacuum chamber of the mass spectrometer presented a tremendous challenge. The major problem is the large amount of condensable vapor produced from the spray nozzle. This problem was overcome by applying a counterflowing curtain gas (20–70°C) to that of the spray and by nebulization. Nitrogen is used for counterflowing curtain gas. The curtain gas is delivered in the opposite direction to the flow of sample. This curtain of dry nitrogen serves to exclude large droplets and particles and to decluster the ions. The nebulization gas is usually air or nitrogen. It is delivered in the same direction of the flow of sample by means of a metal sleeve which wraps around the capillary. Chait and co-workers showed that when the capillary nozzle was operated at around 85°C, counterflowing gas was not necessary [36]. However, thermal destruction of labile peptides and proteins can result from operation at elevated temperature. Fig. 4 shows a schematic diagram of an electrospray source and the formation of molecular ions from charged droplets. These multiply charged ions result from the attachment of protons and/or metal ions (Na^+ or K^+) to the basic and acidic sites on the molecules, respectively. For example, the ε-amino group of lysine can be protonated whereas the side-chain carboxyl group of aspartic acid can carry a sodium ion adduct. Thus, the electrospray mass spectrum shows a distribution of these multiply charged ions. Proteins with a molecular mass of more than 130 000, such as the dimer of bovine serum albumin, have been successfully analyzed by electrospray mass spectrometry [37].

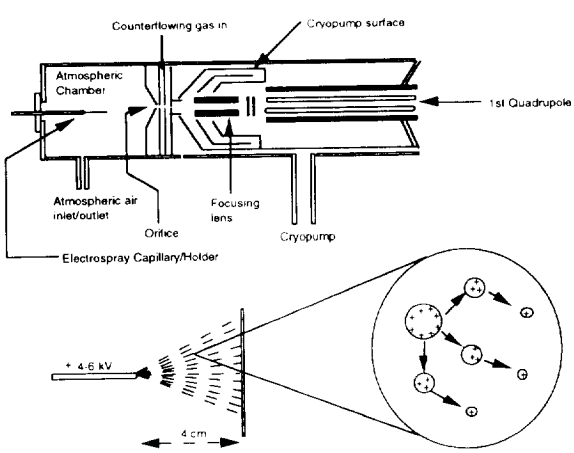


Fig. 4. Electrospray ionization source (above) and formation of molecular ions (below).

The determination of molecular mass from an electrospray mass spectrum is straightforward given the following assumptions. First, the adjacent peaks represent species differing by only one charge. Second, the charge is due to protonation (or some other known ionic species) of the molecular ion. Any two peaks are sufficient to determine the molecular mass. The relationship between the molecular mass (M_r) and the multiply charged ion (M_1) with its charge (Z_1) is described, for the case where the charge is due to protonation, by the equation

$$M_1 Z_1 = M_r + 1.0079\, Z_1$$

$$M_2 Z_2 = M_r + 1.0079\, Z_2$$

where $M_2 > M_1$. By solving the above equations, the charge of M_1 can be calculated using the following equation ($Z_2 = Z_1 - 1$):

$$Z_1 = (M_2 - 1.0079)/(M_2 - M_1)$$

To date, electrospray has been most widely applied using quadrupole mass spectrometers. Other workers have demonstrated that electrospray can be coupled to other mass analyzers such as magnetic sector [29], time-of-flight [38] and Fourier transform ion cyclotron resonance [39] types. Electrospray MS is gaining popularity among protein chemists because of its simplicity and versatility. The fact that electrospray MS can be coupled directly to HPLC allows the routine use of LC–MS. Applications such as peptide mapping, which could take hours or days using FAB-MS, can now be done with ESI-MS in a few hours.

Unlike FAB-MS, where both positive and negative ions are formed, only positive or negative ions are formed in ESI-MS. The charge of the droplets can be controlled by the polarity of

the applied electrical field. In most instances, peptides and proteins work well in the positive-ion mode. However, for molecules that contain many negative charges, such as oligonucleotides, the sample should be analyzed by negative-ion ESI-MS [40]. The negative ion ESI mass spectrum is characterized by an envelope of multiply charged molecular ions with the form $(M - nH)^{n-}$. Negative-ion ESI-MS generally has lower sensitivity than positive-ion ESI-MS. The decrease in sensitivity is the result of substitution of sodium for hydrogen ions. As the molecular mass of the oligonucleotide increases, the number of phosphate–sodium adducts also increases. This results in a broadening of the peaks and lowering of the measurable current for a given ion. Thus, for samples which are subject to alkali metal attachment, the measured molecular mass will often be overestimated.

3. Protein applications

3.1. Molecular mass determination

The molecular mass of a protein is an important parameter in the biochemical characterization of that protein. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) is a universal technique in protein molecular mass determination. However, the accuracy of SDS-PAGE, ranging from a few per cent for a well behaved globular protein to about 30% for a heavily glycosylated protein, makes the estimation of protein molecular masses extremely tenuous.

With the recent advances in ionization technology described in the previous section, mass spectrometry has become a useful tool for protein molecular mass determination with excellent accuracy. Techniques such as MALDI and ESI allow the determination at low-picomole levels of protein molecular masses exceeding 100 000. Table 1 shows the typical operating characteristics of commercially available instruments.

In the determination of the molecular mass of a protein, it should be remembered that the mass spectrometer will detect all isotopes of every element in the molecule. At high enough resolution, the ionic species consisting of different combinations of isotopes will be resolved, but at lower resolution, a single symmetrical peak will be obtained. In ESI-MS and MALDI-MS, centroiding the peak will afford the average molecular mass of the proteins. For protein analysis, it is not necessary to carry out the analysis at high resolution because the monoisotopic ion becomes undetectable, and even the most abundant peak contains a large combination of isotopes which requires a very high resolution mass spectrometer to achieve accurate analysis. Also, higher resolution is obtained at the expense of sensitivity.

The accuracy of ESI-MS is about 0.01%, which is very good compared with SDS-PAGE. Although ESI-MS can measure molecular masses in excess of 100 000, the analysis can be very complicated. This is because ESI-MS requires

Table 1
Operating characteristics of commercial instruments

Ionization technique	Mass range	Analyzer	Resolution	Total sample required for determination
ESI	150 000[a]	Quadrupole	2000	Picomoles
MALDI	>250 000	Time-of-flight[b]	1000	Picomoles
Plasma desorption	45 000	Time-of-flight	1000	Nanomoles
FAB/LSIMS	5000 15 000 (LSIMS)	Magnetic sector	40 000	Nanomoles

[a] Assuming the proteins can be highly protonated.
[b] Theoretically, the mass range is unlimited.

that the adjacent peaks be resolved from each other, otherwise the charge state cannot be determined. This is probably the main reason why ESI-MS fails in analyzing glycoproteins where heterogeneity from the carbohydrate moiety can give rise to a large number of overlapping peaks. In many cases MALDI-MS can be used successfully to determine the molecular mass of glycoproteins, since fewer charge states are observed compared with ESI-MS.

3.2. Peptide mapping/MS

Peptide mapping is a technique whereby a protein sample is digested either enzymatically or chemically, and the resulting peptides are separated and analyzed. This approach has been used for years for the determination of protein primary sequence. Overlapping sets of peptide fragments were generated using different enzymes or chemical agents and the peptides were separated and sequenced using standard Edman degradation reactions. In 1981, Morris et al. [41] used peptide mapping and FAB-MS to examine or screen protein digests or degradation products. Nowadays, peptide mapping used in conjunction with mass spectrometry (we shall refer to it here as peptide mapping/MS), is used in the following situations: (1) confirmation of a protein sequence, especially of proteins produced through recombinant DNA techniques, (2) detection and identification of post-translational modifications, (3) identification of protein degradation products, (4) identification of protein metabolites, (5) disulfide bond assignments, (6) ligand binding and (7) characterization of enzyme active sites. Each of these applications will be considered individually in the following sections. A general procedure for peptide mapping is shown in Fig. 5. This scheme shows all the options that are available; however, in practice all of these options are seldom used and the ultimate goal of the peptide mapping experiment determines the experimental strategy chosen. Briefly, the protein can be subjected to Edman sequencing and carboxypeptidase digestion to determine the N-terminal sequence and the C-terminal residue, respectively. The failure to

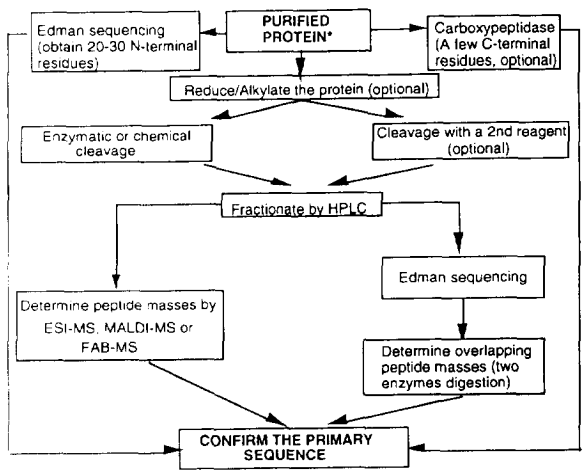


Fig. 5. Procedure for peptide mapping.

generate results from Edman sequencing suggests that the N-terminus of the protein is modified, most commonly by acetylation or formylation. The protein can then be reduced with DTT and alkylated with iodoacetic acid, 4-vinylpyridine or another alkylating agent. The reduced/alkylated protein is then subjected to chemical or enzymatic digestion. In some instances it may be appropriate to omit the reduction/alkylation step (e.g., for proteins containing only a few disulfide bonds or when assigning the disulfide bonds). The protein can be digested with two different enzymes in order to generate overlapping peptide fragments for mass spectral analysis. This option is used when prior information on the protein sequence is not available and it is necessary to align the peptides in the final sequence. The digested fragments can be analyzed directly or be fractionated by HPLC if the mixture is too complex. If desired, the digested fragments can be purified by HPLC and subjected to Edman sequencing. The results from Edman sequencing and mass spectral analysis can be combined to deduce the primary structure of the protein. When the N-terminus of a digested fragment (20 residues or less) is blocked by an acetyl group, the blocking group can be removed by acylaminoacyl-peptide hydrolase [42–44]. Digestion of proteins can be carried

Table 2
Chemical reagents and enzymes for digestion of proteins

Reagents/enzymes	Site of cleavage
1. Trypsin	C-terminus of Arg and Lys
2. *S. aureus* V8, pH 8	C-terminus of Glu and Asp
3. *S. aureus* V8, pH 4	C-terminus of Glu
4. Chymotrypsin	C-terminus of Phe, Tyr, Trp, Leu and Met
5. Endoprotease Lys-C	C-terminus of Lys
6. Endoprotease Asp-N	N-terminus of Asp
7. Clostripain	C-terminus of Arg
8. Pepsin	C-terminus of Phe, Met, Leu and Trp
9. Thermolysin	N-terminus of Leu, Ile, Val, Phe, Met and Ala
10. CNBr	C-terminus of Met
11. Hydroxylamine	Asn–Gly bond
12. Dilute acetic acid	Asp–Pro bond
13. Cysteine cyanylation	N-terminus of Cys
14. NCS	C-terminus of Trp
15. Iodosobenzoic acid	C-terminus of Trp

out chemically or enzymatically. Table 2 lists the common chemical reagents and enzymes that are used in peptide mapping.

As mentioned previously, certain ion suppression effects were observed in FAB-MS with hydrophilic peptides. This effect is virtually nonexistent when peptide mapping is carried out with PD-MS or ESI-MS. The capability of coupling electrospray to HPLC allows the analysis of the digestion fragments without any purification or fractionation. With ESI-MS, full sequencing information is often obtained by performing collision-induced dissociation tandem mass spectrometry (MS–MS). The application of the MS–MS technique in sequencing will be discussed later.

3.2.1. *Confirmation of sequence*

With the proliferation of proteins being produced by recombinant DNA techniques using a variety of expression systems, it is prudent to confirm that the protein sequence obtained is in agreement with that predicted from the DNA sequence. Peptide mapping/MS is an ideal procedure for performing this confirmation of structure in a rapid fashion with minimal consumption of protein sample. The molecular masses of tryptic peptides from the predicted sequence are compared with the molecular masses actually determined from the peptide mapping/MS experiment. This strategy was first suggested in 1984 by Gibson and Biemann [45], who used it to confirm and correct regions from the amino acid sequences of three large proteins, glutaminyl- and glycyl-tRNA synthetase from *Escherichia coli* and methionyl-tRNA synthetase from yeast. Since that time there have been several examples of this same approach applied to other proteins [46–49]. In the past, peptide mapping was a laborious process which included optimization of the HPLC conditions, fractionation of all the digested peptide fragments and analysis of all fractions by either amino acid analysis, peptide sequencing or mass spectrometry or all three. With modern technologies such as ESI-MS and MALDI-MS, peptide mapping/MS of a digested peptide mixture can be performed in a much shorter time.

MALDI-MS has proven useful for peptide mapping/MS and there are several advantages in using these techniques. Since MALDI-MS can tolerate the presence of buffer salts and other impurities, a mixture of peptides obtained from, for example, an enzymatic digest can be analyzed without prior purification. This technique allows a mass spectrum containing ion masses for all peptide fragments to be obtained in only a few minutes after completion of the digest.

Disadvantages of this technique are generally related to the matrix in which the sample is analyzed. The analytes must be uniformly distributed throughout the matrix crystals to ensure that all peptides can be desorbed. The choice of matrix is critical. Each of the common matrices discriminates against mixture components in ways that are sample and matrix specific, thereby necessitating that data be acquired using more than one matrix. Background signals are a problem with any matrix and with some the signals are present up to M_r 1000. Many peptide mixtures resulting from fragmentation of a protein contain peptides with molecular masses well below 1000. These problems are being slowly overcome by the development of new matrices and newer procedures for sample preparation. Stults et al. [50] evaluated the effects of carbohydrate-containing matrices for MALDI-MS analysis of in situ digests of proteins from two-dimensional gels and found fucose–2,5-dihydroxybenzoic acid (DHB) to be the best. Several crystallization methods for MALDI-MS have also been reported, which include vacuum crystallization [51] and stressed matrix crystallization [52,53].

ESI-MS has one major advantage over MALDI-MS, namely the ability to couple an HPLC system directly to the mass spectrometer and perform LC–MS. Application of ESI-MS in peptide mapping/MS offers a rapid means to confirm the primary sequence of proteins. In general, the digestion mixture is separated by RP-HPLC. The output from HPLC is split; a small amount of effluent goes to the mass spectrometer and the rest goes to a fraction collector or waste. Using this set-up the mass of a particular HPLC peak can be assigned readily. The same peak can be collected for Edman sequencing. The results from Edman sequencing determine the N-terminus of the peptide, whereas, the obtained mass determines the C-terminus of the peptide. This procedure can be applied for all HPLC peaks to confirm the primary sequence of proteins. The disadvantage of using ESI-MS in peptide mapping is the time necessary for LC analysis, which could range from 0.5 to 2 h. Fig. 6 shows an example of an LC–MS analysis of the trypsin digest of carboxyamidomethylated hGH. By examining the peaks in the TIC, one can assign the masses to most of the HPLC peaks. However, glycopeptides often show poor response in ESI-MS.

3.2.2. Post-translational modifications

A large number of proteins undergo post-translational modification after their biosynthesis. In some cases, post-translational modifications are critical for the transport and bioactivity of proteins. Post-translational modifications are often not detected in Edman sequencing because the modified residue is either not detected or is destroyed under the harsh chemical environment. Mass spectrometry is probably the best method for the detection of post-translational modifications. Krishna and Wold [54] compiled a useful list of post-translational modifications of proteins. Some of the common post-translational modifications include acylation of the N-terminus, phosphorylation of serine, threonine and tyrosine, oxidation of methionine, glycosylation of serine and asparagine, cleavage of N- and/or C-terminal residues by proteolytic enzymes and formation of C-terminal amides. For simple modifications, such as acetylation or phosphorylation, the analysis is simple and can be carried out using the normal peptide mapping procedure [55]. Caprioli et al. [56] developed a procedure to generate a "sequence-ordered" map from a protease digest by using the overlap information produced from a time-course protease digest of a protein. This technique is based on the fact that the rate of cleavage can differ widely from site to site and that all sites are not equally accessible. Clostripain was used as a limited proteolytic enzyme because it hydrolyzes only at the C-terminal side of arginyl residues. Since not all arginines are accessible to the enzyme, a time-course digestion would produce peptide fragments with large overlapping regions. One other approach is to perform the digestion in oxygen-18-enriched water; only the C-terminal fragment will be unlabeled [57].

Perhaps the most common modification of proteins is glycosylation. Analysis of glycoproteins by mass spectrometry is complicated by the

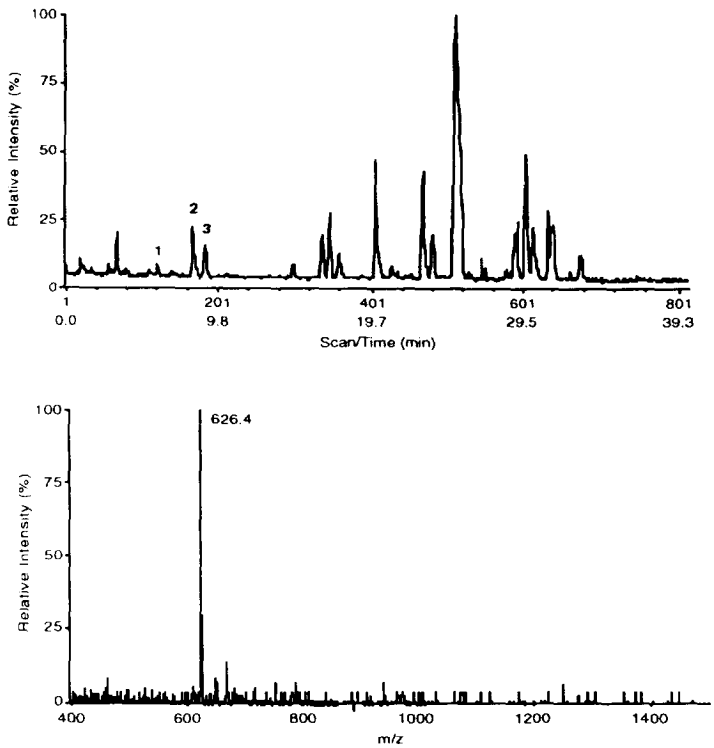


Fig. 6. LC–MS of a trypsin digest of carboxyamidomethylated hGH using ESI-MS. Peak 1 contains an ion at m/z 626.4 which matches the LEDGSPR (T_{14}) fragment of the protein. Every peak in the TIC chromatogram can be analyzed in the same manner to confirm the primary sequence of the protein.

heterogeneity of the carbohydrate moiety. The most reliable method of determining the molecular mass of a heavily glycosylated protein is MALDI-TOF mass spectrometry; the complexity derived from the multiple charge states present in ESI-MS generally renders it impractical for this purpose. However, when the glycoprotein is digested into smaller fragments, FAB, MALDI and ESI mass spectrometry can be successfully applied to the analysis. A classical approach to glycoprotein analysis is to deglycosylate completely using hydrazinolysis, isolate and purify the polysaccharides and then perform structural characterization. A major drawback of this technique is that sites of glycosylation on the protein cannot be determined because of the destruction of the protein. Application of mass spectral techniques to glycoprotein analysis has provided a means of characterizing these glycoproteins efficiently. Sequence analysis of oligosaccharides is outside the scope of this paper, although Sutton et al. [58] have utilized MALDI-MS to identify the site of glycosylation and to obtain the sequence of the carbohydrate moiety. The initial problem with an unknown protein is to determine whether it is glycosylated. One approach has been to apply a highly sensitive lectin screen where a range of lectins must be employed to ensure that all of the common oligosaccharides will be detected [59]. Once a protein is confirmed to be a glycoprotein, it is then fragmented chemically or enzymatically. Hawke et al. [60] have reported that, by using an affinity column, most of the glycopeptides could be separated. The remaining fractions were analyzed by HPLC and screened again for glycopeptides. The isolated glycopeptides were then subjected to mass spectral analysis. The determi-

nation of N-linked oligosaccharides can be performed readily with a "pure" oligopeptide. Digestion of the oligopeptide with peptide-N-glycosidase F (PNGase F) in 50% ^{18}O-labeled water affords a pair of signals (MH^+ and $MH^+ + 2$) due to the partial incorporation of oxygen-18 into the β-carboxyl group of aspartic acid [61]. Unfortunately, there is no known enzyme which can hydrolyze O-linked oligosaccharides. Carr et al. [2] have proposed an elegant procedure using LC–ESI-MS to identify selectively N- and O-linked oligosaccharides in glycoproteins. The glycopeptides are identified using a carbohydrate marker ion. This is done by stepping the "orifice potential" to enhance the low-m/z fragments, in particular m/z 204. This experiment allows one to confirm the presence of a glycoprotein and to localize the glycopeptide fragments in the enzymatic digested chromatogram. By performing a "precursor scan" of m/z 204, glycopeptides can be selectively detected in the presence of other peptides. It should be noted that the spectrum generated using this procedure corresponds to all parent ions that have decomposed to yield m/z 204. O-Linked glycopeptides can be differentiated from N-linked glycopeptides by analyzing a sample that has been treated with PNGase F. Sites of glycosylation can be obtained by MS–MS analysis of the glycopeptides. Edman sequencing can also be used to determine the site of glycosylation (absence of signal where the residue is glycosylated). A schematic diagram of a strategy for glycoprotein analysis is shown in Fig. 7.

One should remember that carbohydrates offer a relatively small number of protonation sites. This point is very important when ESI-MS is used for glycoprotein analysis. For example, if a large oligosaccharide is attached to a small peptide, there might not be enough charge to bring the ions into the mass range of the quadrupole analyzer. In this case, MALDI-MS is a good alternative method.

3.2.3. Protein degradation products

Since almost all degradation products of proteins result in a change in molecular mass compared with the original protein, mass spectrometry can be used to detect such degradation products, in principle. However, mass determination of the intact protein does not yield information as to the location of the chemical modification. Further, for large proteins (M_r greater than approximately 20 000), mass spectrometry generally does not offer sufficient resolution to detect modifications of only 1 u (e.g., as a result of deamidation). Hence the most common use of mass spectrometry with regard to characterization of protein degradation products is the combined use of peptide mapping and mass spectrometry. Examples of this approach include the characterization of deamidated and methionine sulfoxide forms of human growth hormone [63], hydroxylamine cleavage products in insulin-like growth factor I [64] and deamidated forms of tumor necrosis factor [65]. In each of these examples the approach used was to generate peptide fragments (usually by trypsin digestion), isolate those peptide fragments which differed from the fragments arising from the non-degraded protein and characterize the modified peptide fragments by FAB or ESI (usually employing MS–MS to confirm the sequence modification which had occurred). In most cases, N-terminal analysis by Edman degradation is also employed to provide corroborating evidence.

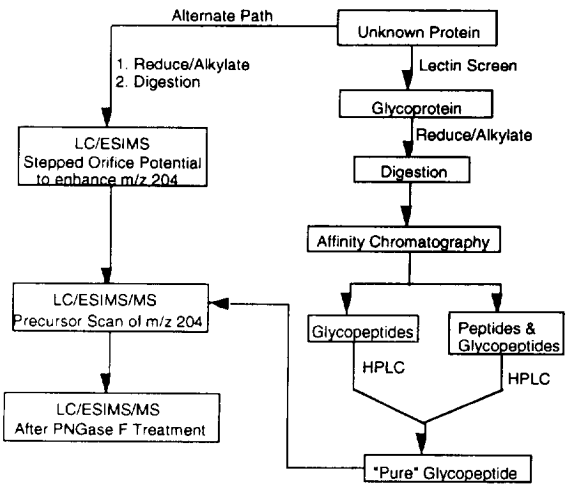


Fig. 7. Strategy for glycoprotein analysis (see Refs. [60] and [62]).

3.2.4. Protein metabolites

Wroblewski and co-workers have used ESI-MS to study the in vitro metabolism of human growth hormone (hGH) [66] and the in vivo metabolism of des(64,65)-human proinsulin [67]. In these examples the protein is proteolyzed, not by the addition of exogenous proteases as in a typical peptide mapping experiment, but by the action of endogenous proteases. An abundance of information concerning the activities involved, the metabolic products and potential metabolic pathways can be obtained from such experiments.

Human growth hormone was incubated with preparations of rat thyroid gland; the products were isolated by reversed-phase HPLC and characterized by a combination of ESI-MS and N-terminal Edman sequencing. The predominant activity that acted on hGH was found to be a chymotrypsin-like serine protease, biochemically similar to rat mast cell protease-I, with cleavages occurring exclusively at Tyr–Phe–Leu–Xaa bonds. The presence of a carboxypeptidase activity was indicated by the detection of metabolites that were truncated by a single amino acid at the C-terminus. The sequence of events leading to the degradation of hGH in this system was found to be initiated by a cleavage between Tyr143–Ser144 to produce a two-chain form of the protein. This was followed by the cleavage of the initial two-chain form at Tyr42–Ser43, liberating the N-terminal peptide Phe1–Phe42. Subsequent events over a 4.5 h incubation led to the degradation of hGH to a set of more than 20 peptides with masses at or below 2300 [66].

The metabolism of des(64,65)-human proinsulin was examined in rats after subcutaneous administration. Circulating insulin-like immunoreactivity in the plasma 25 min after injection was evaluated by anion-exchange and reversed-phase HPLC. Both techniques indicated the presence of a metabolite comprising 5–10% of the circulating immunoreactivity and having the retention characteristics of human insulin. The remainder had retention characteristics of des(64,65)-human proinsulin. The peaks of immunoreactive material were isolated and their structures determined using reversed-phase HPLC and ESI-MS. The major circulating component co-eluted with des(64,65)-human proinsulin and had an identical mass spectrum. Two circulating metabolites were identified. These metabolites co-eluted by reversed-phase HPLC with human insulin and diarginyl(B31,32)-human insulin and had mass spectra identical with the standard compounds. The data indicate that proteolytic processing of des(64,65)-human proinsulin involves an initial tryptic cleavage at the carboxy side of ArgB32, with the formation of human insulin by the subsequent action of a carboxypeptidase to remove the ArgB31–ArgB32 dipeptide from diarginyl(B31,32)-human insulin. The results suggest that some of the pharmacological activity of des(64,65)-human proinsulin may be mediated in part by circulating insulin-like metabolites [67].

3.2.5. Disulfide bond determination

Many proteins, especially secreted or non-cytoplasmic proteins, contain disulfide bonds between pairs of proximal cysteines. The disulfide bonds play a role in the stabilization of the tertiary structure of these proteins. Proteins that are produced through recombinant DNA technology are often refolded from strong denaturants into biologically active forms and it is critical to determine whether the "correct" cysteines are involved in forming the disulfide bonds. Likewise, for naturally occurring proteins, it is important to characterize the disulfide pairing pattern. Some proteins contain free cysteine residues, not involved in disulfide bonds, and it is important to identify these residues. For example, the protease clostripain contains a cysteine at the active site and this free thiol group is essential for the activity [68]. Although analytical methods have been developed to determine the thiol content in a protein [69], the determination of the locations of disulfide bonds is more of a challenge. Disulfide-containing peptides can be identified by their amino acid composition or sequence only if they are purified to homogeneity. For a protein containing many disulfide bonds, this approach cannot be applied. However, total cysteine content can be determined by complete reduction of

the disulfide bonds with a suitable reagent (DTT or β-mercaptoethanol are commonly used) followed by alkylation with iodoacetic acid or iodoacetamide. The total mass increment of the intact protein after this procedure divided by the mass of the alkylating group (e.g., 58 for acetamide) yields the number of cysteines in the protein. If this procedure is repeated, but without prior reduction, only the free cysteines, not involved in disulfide bonds, are alkylated. It should be cautioned that some proteins may have free cysteines buried inside the tertiary structure, preventing them from being alkylated. Therefore, it might be necessary to perform the alkylation in strong denaturing conditions. Again, the mass increment divided by the mass of the alkylating group yields the number of free cysteines in the protein. The difference between the total number of cysteines and the number of free cysteines divided by 2 (two cysteines per disulfide bond) gives the number of disulfide bonds. One of the greatest problems in localizing the disulfide bonds is disulfide exchange. Application of mass spectrometry to solve this problem originates from the observation by Yazdanparast et al. [70] that disulfide bonds are reduced in situ during peptide analysis by FAB-MS.

The determination of the cysteines involved in a disulfide pair is accomplished by a peptide mapping approach. The key to applying this approach successfully is to isolate peptides that contain a single disulfide bond. For large proteins cleavage with cyanogen bromide can be used to generate large peptide fragments that are more susceptible to proteolytic digestion. The protein or large protein fragments are then digested with a suitable protease. Although trypsin can be used, care must be taken that disulfide interchange not occur. If disulfide interchange is a problem, an acid protease such as pepsin should be used. Following proteolytic cleavage, the peptides are isolated and subjected to both Edman sequencing and mass spectral analysis. The peptides that contain multiple sequences are candidates for peptides that contain disulfide bonds. If more than two sequences are found in an isolated peptide peak, this indicates the existence of more than a single disulfide bond and requires subfragmentation with a second protease. Even the existence of only two sequences does not guarantee a single disulfide bond if cleavage has not occurred between the two cysteines of one disulfide bond. Once a set of peptides, each peptide containing a single disulfide bond, is obtained, the disulfide bonds can be assigned unambiguously. If the protein being analyzed has a free sulfhydryl, it will be necessary to alkylate it before starting the fragmentation procedure.

3.2.6. Ligand binding

In the field of drug discovery, a common theme is the identification of low-molecular mass compounds that bind very tightly to a target protein. An example of this is the covalent attachment of protease inhibitors to proteases, often at the active site of the protease. Covalently bound ligands can be studied using the normal peptide mapping procedure. Sall and Kaiser [71] used peptide mapping and ESI-MS to localize the site of binding of methyl 3-(2-methyl-1-oxopropoxy)[1]benzothieno[3,2-*b*]furan-2-carboxylate, a potent and highly selective inhibitor of thrombin, to the active site of thrombin. They were able to show that this compound formed a stable acyl enzyme complex at the active site Ser-205 which, along with His-43 and Asp-99, make up the active site triad of thrombin.

A similar approach was used with the viral enzyme, rhinovirus 3C protease, to show that an inhibitor of this enzyme bound covalently to the active site Cys-146 (Zimmerman and Becker, personal communication). In additional to localizing the site of attachment, information on the nature of the chemical reaction involved can be obtained from the increment in mass that the protein undergoes.

3.2.7. Enzyme active sites

A related application is the characterization of enzyme active sites. The general approach is to utilize a specific affinity label which forms a covalent bond with one of the active site residues. Peptide mapping is then applied to localize the peptide containing the modification.

Several examples of this approach can be found in the recent literature.

The apparent active site of human leukocyte glycoasparaginase, an enzyme involved in the degradation of the N-glycosidic linkage between asparagine and N-acetylglucosamine in various glycopeptides, has been identified by labeling with an inhibitor, 5-diazo-4-oxo-L-norvaline, an asparagine analog [72]. The labeled protein was digested with trypsin and analysis of the peptides by mass spectrometry revealed that the inhibitor was attached through an α-ketone ether linkage to the hydroxyl group of the N-terminal threonine.

Inactivation of histidine ammonia lyase from *Pseudomonas putida* was accomplished by treatment with L-cysteine at pH 10.5 in the presence of oxygen [73]. Inactivation was accompanied by the formation of a new species with a UV absorbance maximum at 340 nm. Following trypsin and staphylococcal V8 protease digestion, a peptide was isolated that was found to contain a modification of mass 184 by ESI-MS. The modification was localized to Ser-143 of this enzyme and the authors conclude that this represents the site for attachment of an electrophilic cofactor required for histidase activity.

3.2.8. Peptide mass maps

An innovative approach to the identification of protein sequences in a large protein sequence database such as the Protein Information Resource (PIR) or the SWISSPROT database has been created by Yates et al. [74]. The approach involves the proteolytic digestion of an unknown protein and then analysis of the peptides by a peptide mapping/MS approach. They have demonstrated that a set of observed masses which is less than 50% of the total number of predicted masses can be used to identify a protein sequence in the database. Mass maps generated by ESI-MS, MALDI-MS and FAB-MS should all work with this approach. When multiple matches are found, tandem mass spectrometry (discussed in the next section) can be used to establish sequence similarity.

3.3. Sequencing

3.3.1. Tandem mass spectrometry

Tandem mass spectrometry (MS–MS) was originally developed for the investigation of ion chemistry. However, its major application currently is in determining the structure of biomolecules. The tandem mass spectrometer could be a multisector magnetic instrument (tandem double focusing) or a triple quadrupole instrument. In both cases, the "first" mass spectrometer (or the first quadrupole) serves as a mass separator. The selected ion (precursor ion) can be induced for fragmentation by colliding with an inert gas. This process is often referred to as collision-induced dissociation (CID). The fragment ions, commonly referred to as product ions, are analyzed in the "second" mass spectrometer (or the third quadrupole). A detailed description of tandem mass spectrometers has been given elsewhere [75]. This section will only describe the use of tandem mass spectrometry in peptide sequencing.

A diagram of tandem mass spectrometry is shown in Fig. 8. A peptide mixture consisting of R, M and N components is ionized. Peptide M is then selected by the first mass analyzer and subsequently induced by colliding with a neutral gas such as argon or helium to give fragment ions A, B and C. These fragment ions can be scanned by the third mass analyzer to give the mass spectrum of the product ions. The use of a collision gas is essential to produce a significant number of intense product ions.

In peptides, the fragmentation generally occurs at the peptide backbone. However, cleavage at peripheral bonds can occur, particularly when using magnetic sector mass analyzers. For peptides with mass greater than 2500, fragmentation seldom occurs because the vibrational energy

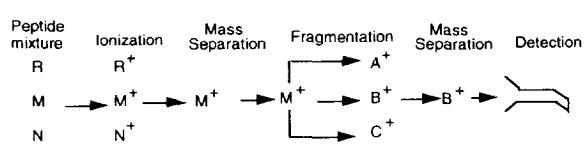


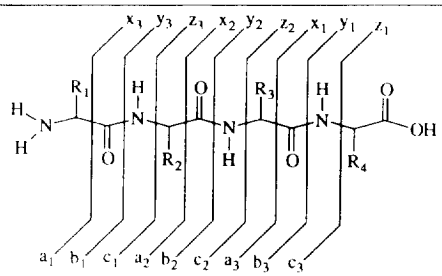
Fig. 8. MS–MS of R, M and N peptide mixture.

induced by collision is distributed through a larger number of bonds (except for post-source decay MALDI-MS where fragmentation of peptide with molecular mass up to 3000 Da can occur). In order to understand peptide fragmentation, a nomenclature system was proposed by Roepstorff and Fohlman [76]. This system was soon modified to use lower case rather than capital letters to avoid confusion, e.g. C_n could be mistaken for a cysteine residue at position n [77]. Table 3 shows the common types of fragment ions.

The mechanism for fragmentation is not fully understood at the present time. The formation of a_n, b_n and y_n is generally accepted to involve initial protonation of the amide nitrogen atom. d_n Ions retain the positive charge at the N-terminus and are formed as a result of the loss of an alkene from the side-chain of the other terminus of a_n ions. d_n Ions provide a means to differentiate the isomeric amino acids leucine and isoleucine. If leucine were the amino acid at the terminus, an isopropyl radical would be eliminated. On the other hand, isoleucine would eliminate a methyl or an ethyl radical [78,79]. Johnson et al. [80] proposed mechanisms for the formation of the fragment ions. It is important to know that these mechanisms are chemically plausible, but there are few experimental data to support them except for the observed ions. In the most common form of ESI-MS, in which quadruple mass analyzers are used, b and y fragment ions predominate.

Tandem mass spectrometry has been used to study a large number of peptides and proteins. Some of the notable works include the studies of bradykinin [81], substance P [82], somatostatin [83] and thioredoxin [84]. So far, most of the proteins studied are well characterized and have known primary sequences. For an unknown protein, the tandem mass spectrum can be very complicated because some proteins exhibit both N- and C-terminal fragments. There are also fragments that do not fit in any of the proposed fragment ions. Despite these complications, mass spectrometry offers protein chemists a means to obtain sequencing information much faster than any other analytical method. There are also

Table 3
Common types of fragment ions



Ion	Type of cleavage	Terminus	Structure
a_n	Backbone	N	$^{+}NH{=}CHR$
b_n	Backbone	N	$-NHCHRCO^{+}$
c_n	Backbone	N	$-[NHCHRCONH_2]H^{+}$
x_n	Backbone	C	$[OCNHCHRCO_2H]H^{+}$
y_n	Backbone	C	$H_3^{+}NCHRCO-$
z_n	Backbone	C	$[CHRCONHCHRCO_2H]H^{+}$
d_n	Side-chain	N	$[-NHCH{=}CHR']H^{+}$
v_n	Side-chain	C	$[HN{=}CHCONHCHRCO_2H]H^{+}$
w_n	Side-chain	C	$[R'CH{=}CHCONHCHRCO_2H]H^{+}$

computer programs which can be used to assist in the analysis of spectral data of peptides. Some of the commonly used programs include MacProMass [85], Mac Mass [86] and COMPOST [87].

3.3.2. Post-source decay MALDI-MS

In addition to tandem mass spectrometry, post-source decay (PSD) is a new technique for peptide sequencing. It has been observed that large peptide and protein ions are unstable on their way through a time-of-flight instrument. The metastable decay of these ions has been detected using two-stage reflectron MALDI-MS. The term post-source decay was introduced by Spengler and co-workers [88–91] to describe the "delayed" fragmentation of the desorbed analyte ions. The fragmentation was due to the multiple collisions of analyte ions with matrix ions during the early plume expansion and ion acceleration followed by collisions with residual or admitted gas molecules in the field-free drift region of a MALDI instrument. Unimolecular decomposition of the precursor ion gives rise to fragmentation patterns which can be used to determine the sequence of the peptide. The fragment ions travel with nearly the same velocities as their precursors, but with lower kinetic energies. Thus, by applying a retarding field at the reflector, one can separate the product ions from the precursor ions. Because of the lower kinetic energies of the product ions, they do not penetrate into the retarding field as deeply as their precursors; therefore, they leave the reflector earlier and arrive sooner at the detector. PSD-MALDI offers many advantages over conventional tandem mass spectrometry. The instrumental sensitivity for product ions is at least two orders of magnitude better than tandem mass spectrometry owing to the high ion transmission and high ion yield (20–80% of precursor ion). Also, MALDI-MS offers a much longer dissociation time range than a conventional collision cell which allows the activation and decay of peptides of M_r up to 3000. The fragmentation patterns of PSD-MALDI are very similar to that of the low-energy CID [92]. Typically, a_n, b_n, C_n, $a_n - 17$, $b_n - 17$, x_n, y_n and z_n ions are observed in a mass spectrum. The loss of 17 u is due to loss of ammonia from arginine residues.

There are drawbacks associating with PSD-MALDI which should be mentioned. First, mass calibration can be difficult because one would have to perform tedious calibration with known precursor ion and product ion masses. Therefore, calibration is normally done using computer software. This software takes into account the instrument geometry, electrical field parameters and the flight times of precursor and product ions to calculate the ion masses. Second, for large peptides (M_r above 1500), it is often necessary to increase the laser irradiance to have sufficient desorption. However, increasing the laser irradiance causes a decrease in resolution which ultimately increases the uncertainty in mass assignment of these peptides. Third, because of the complexity of the fragmentation patterns and the lack of algorithms for cleavage interpretation, analysis of PSD mass spectra requires much more time than it takes to acquire the mass spectrum. Despite these difficulties, PSD-MALDI has gained acceptance in peptide sequencing. For example, Zambias et al. [93] utilized PSD-MALDI for the analysis of a covalently bound peptoid (modified peptide) to a polymeric bead. Hoyes et al. [94] utilized PSD-MALDI for the analysis of ACTH peptide (M_r 2466). Spengler and Kaufmann [95] utilized PSD-MALDI to localize the site of a post-translational modification of a peptide. The analysis of a lipopeptide (M_r 1863) located the fatty acid on the N-terminus of the peptide. One interesting, perhaps very important, application of PSD-MALDI is the possibility of using this technique for the characterization of unknown proteins. Yu et al. [96] performed PSD-MALDI on an unknown tryptic fragment. The resulting primary sequence of this peptide was compared with existing sequences in a database to suggest that this protein resembles CHO MCP-1 protein.

3.3.3. Carboxypeptidases

Carboxypeptidases, exoproteases which cleave only the C-terminal residues in proteins, have been used in protein sequencing with variable success. Four carboxypeptidases, carboxypepti-

dase P (CPP), A (CPA), B (CPB) or Y (CPY), or a combination of these have been most commonly used in sequencing applications. In general, the rates of cleavage by these enzymes are dependent on the polarity of the C-terminal amino acid side-chain, making sequencing progress extremely variable. Because of this variability, carboxypeptidases have traditionally been used to determine only the C-terminal residue of a protein by identifying the amino acid released by its chromatographic retention time [97].

A different approach is to monitor the truncated protein instead of the released amino acid and the techniques of ESI-MS, PD-MS and MALDI-MS now make this approach feasible. A digestion reaction is carried out using carboxypeptidase(s) of choice. At selected intervals, aliquots are removed and the reaction is quenched and stored for later analysis. Using PD-MS, the digestion can be carried out directly on the nitrocellulose support [98,99]. With ESI-MS, the digestion mixture can be monitored by continuous infusion. Smith and Duffin [100] used carboxypeptidase P to sequence interleukin 3 (IL-3) and superoxide dismutase (SOD). IL-3 has a C-terminal sequence of ...TTLSLAIF. The sequencing was performed in 15 m*M* ammonium acetate buffer (pH 4.0). Under these conditions, six amino acid residues were digested by CPP. Additional reaction time did not release any further residues. Superoxide dismutase has a C-terminal sequence of ...CGVIGIAK. Digestion of this protein yielded only one residue. However, after reducing and alkylating the protein with iodoacetic acid, three residues were released by the enzyme. The sequencing did not proceed past the glycine residue even with extended reaction time. Schär et al. [101] reported the use of a mixture of CPA and CPB and also CPY for the digestion of synthetic parathyroid hormone using MALDI-MS. After about 2 h of digestion, 21 amino acid residues were removed from the protein. It was found that CPY digests valine rapidly and very slowly at histidine residues, hence it was not possible to localize all the truncated proteins. Also, truncated peptides of M_r below 1200 were difficult to detect because of interference from the matrix. Rosnack and Stroh [102] described the use of a low-flow reactor to monitor the digestion of glucagon and apomyoglobin using ESI-MS. The protein and enzyme (CPP) were mixed and infused into a fused-silica reactor with continuous monitoring by ESI-MS. Glucagon gave sequence ions for the first 19 amino acids; however, apomyoglobin gave only 50% of the first 30 amino acids. Attempts were made to sequence cytochrome *c* and carbonic anhydrase. Unfortunately, no sequence information was obtained from these digestions.

The application of carboxypeptidases to protein sequencing has been investigated in our laboratory. We decided to use ESI-MS in combination with CPP and CPY for protein sequencing. Glucagon (HSQGTFTSDYSKYLDSRRAQ-DFVQWLMNT) was used as a model compound with a mixture of CPP and CPY as reagents. The detection of the truncated forms of glucagon was carried out by ESI-MS. The combination of CPP and CPY was found to be necessary for successful digestion. For example, digestion with CPY alone was retarded at the tryptophan residue and virtually stopped at the aspartic acid residue. Digestion with CPP alone proceeded past these residues but was retarded at the leucine residue. A combination of CPP and CPY resulted in digestion of the entire glucagon molecule in about 70 min. (Fig. 9). Peaks associated with all of the predicted glucagon fragments could be identified in the mass spectrum taken at 70 min although there was a high background. Obviously, the sequence of glucagon is known and this greatly facilitates the analysis of the mass spectrum. An unknown protein would present a much greater challenge and at present the technique is probably best applied to the confirmation of a known or predicted sequence. The ability to obtain sequence information readily from the C-terminus of a protein is, however, an important advance in protein characterization.

3.4. Non-covalent interactions

Non-covalently bound ligands pose a problem in the analysis of these complexes by most ionization techniques. Henion and co-workers

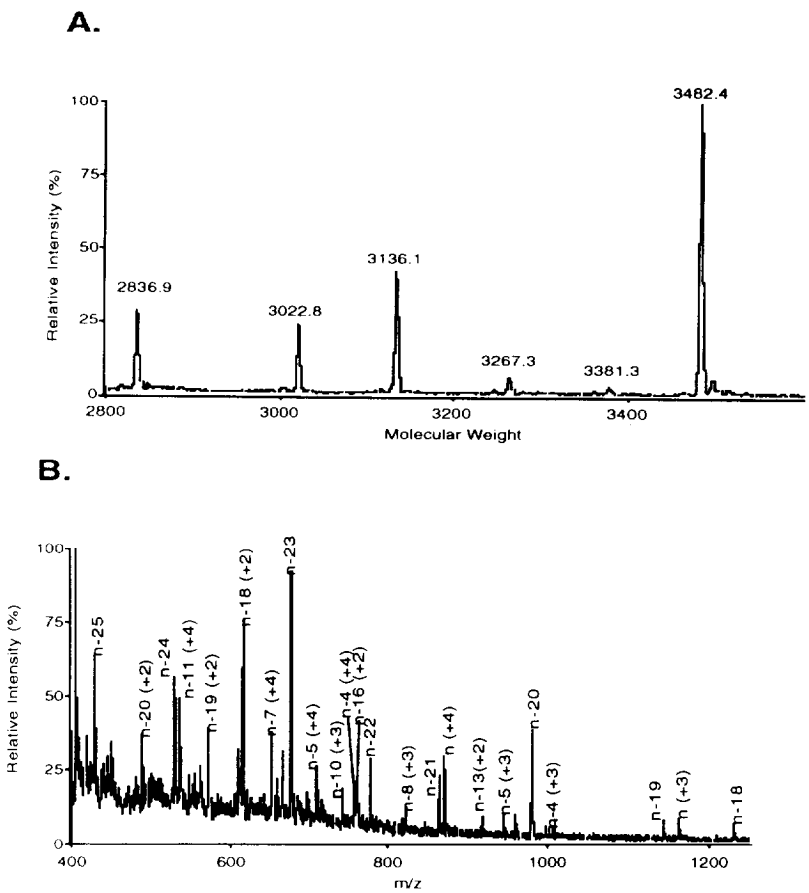


Fig. 9. Sequencing of glucagon using a mixture of CPP and CPY. (A) Reconstructed mass spectrum of digestion using CPP and CPY after 15 min. Data were acquired in the chromatographic mode. (B) The actual ion m/z of digestion after 70 min; n refers to the glucagon and $n-18$ refers to glucagon minus 18 amino acid residues from the C-terminus. The charges of the ions are shown in parentheses. The charge is not specified when the ion is in a +1 charge state. The +1 charge states were assigned by examination of the isotopic pattern of the peak. A total of 21 nmol of glucagon was consumed. See to Table 1 for peak assignments.

suggested that non-covalent molecular association complexes might be detectable under the "soft ionization" conditions offered by ESI and that the reaction might be monitored in a real-time mode [103]. The first successful application of ESI-MS to the detection of non-covalent receptor–ligand complexes was reported by the same laboratory [104]. The macrolide FK506, an immunosuppressive agent, inhibits T-cell activation when in a complex with the cytoplasmic receptor FKBP. FKBP was mixed with a slight excess of FK506 at pH 7.5, and a new signal appeared corresponding to the complex. Two important controls were performed to ensure that covalent adducts did not form. First, when FK506 was combined with denatured FKBP, the signal for the complex was not observed. Second, when FK506 was replaced with a more potent reagent (higher affinity toward the receptor), the peak corresponding to the new complex was much stronger in intensity.

ESI-MS has also been used to study the hydrolysis of a hexasaccharide of N-acetylglucosamine (NAG_6) by lysozyme. A mass spectrum immediately after mixing the enzyme and substrate reveals a new protonated peak which corresponds to the lysozyme–NAG_6 complex. During the time-course reaction, the spectrum

shows the disappearance of lysozyme–NAG_6 and appearance of lysozyme–NAG_4 and lysozyme–NAG_3 [105].

Kata and Chait [106] demonstrated the utility of ESI-MS in analyzing protein–ligand complexes by studying the heme–globin interaction. The oxygen-carrying protein myoglobin contains a non-covalently bound heme group in the hydrophobic pocket of the native globin chain. At low pH values, unfolding of the protein results in complete protonation of buried histidine residues, resulting in a higher charge-state and a loss of the heme group. A mass spectrum obtained from an aqueous myoglobin solution at pH 3.35 (completely denatured protein) shows a single distribution of peaks and the total absence of ions correspond to the heme–globin complex. However, at pH 3.9, where both native and denatured forms of myoglobin coexist, the mass spectrum exhibits two distinct distributions of peaks. One set of peaks corresponds to the denatured myoglobin (higher charge state) and the other set corresponds to the intact heme–globin complex. These results suggest that native, non-covalent complexes of proteins and cofactors in solution can be preserved in the gas phase and observed by mass spectrometry.

Many proteins self-associate to form non-covalent dimers or higher oligomers. For example, human growth hormone (hGH), under certain conditions, will form a tightly complexed, non-covalent dimer [107]. This dimer is stable under aqueous conditions and can be resolved from the monomer by high-resolution size-exclusion chromatography. Re-chromatography of the isolated dimer on the same system reveals that this fraction is still dimeric with the retention time expected for the dimer and has not reverted to monomer. However, addition of 30% acetonitrile to the sample before re-chromatography caused the conversion of the dimer into monomer. Likewise, treatment with SDS sample buffer followed by analysis by SDS-PAGE shows that the dimer has been converted into monomer. This hGH dimer has been studied by ESI-MS (Fig. 10). Under the conditions normally used to obtained protein spectra, 0.1–1.0% acetic acid–50% acetonitrile, only the spectrum

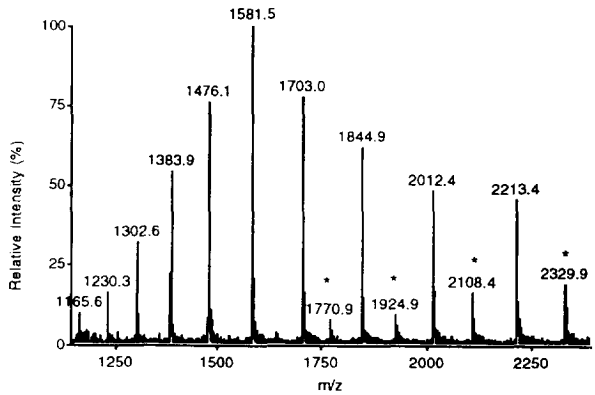


Fig. 10. Spectrum of a non-covalent dimer of human growth hormone. The spectrum was obtained on a Sciex API-III instrument by infusing at 2.5 μl/min a sample of hGH dimer at a concentration of 1 mg/ml in 0.1% acetic acid. The *m/z* values corresponding exclusively to the dimer are marked with asterisks.

of monomeric hGH was observed. This is not surprising, since 30% acetonitrile had been shown to dissociate the dimer. However, by omitting the acetonitrile and infusing a solution at a concentration of 1 mg/ml in 0.1% acetic acid, evidence for the dimer was obtained. The spectrum appears to be an overlay of the spectrum of the dimer and the spectrum of the monomer. To rule out the possibility that the observed dimer spectrum was the result of dimer formation in the mass spectrometer, a solution of monomeric hGH was studied under identical conditions and no evidence of dimer formation was observed.

3.5. Protein folding/higher order structural characterization

3.5.1. Deuterium exchange

While mass spectrometry is most commonly used to assess the primary structure of proteins or peptides, there are various experimental approaches available which can yield information concerning protein folding and higher order (secondary, tertiary or quaternary) structure. One such approach is the determination of hydrogen–deuterium exchange rates under controlled solution conditions. In one application

the exchange rates of the various amide hydrogens in cytochrome *c* were determined using HPLC–FAB-MS [108]. Cytochrome *c* was incubated in ^2H_2O at various temperatures. After incubation for preset time intervals, the protein was transferred into a solution in which deuterium exchange is very slow (pH 2–3, 23°C) and digested with pepsin. The number of deuterium atoms incorporated into each of the proteolytic peptides was determined by HPLC–FAB-MS. This approach can provide information regarding the secondary or tertiary structure of the protein, since amides which exchange readily have a greater solvent accessibility than do amides which exchange slowly. In cases where a biosynthetic form of a protein is being structurally characterized, this approach can be especially useful if a native (natural source) form of the protein is available as a comparator.

3.5.2. Proteolysis

Another useful strategy for protein structure evaluation is limited proteolytic digestion. This approach involves digestion of the protein with a proteolytic enzyme, under conditions such that the rate of proteolysis is sufficiently slow to allow determination of the kinetics of formation of the various peptide fragments. This approach has been used to compare the structures of wild-type and recombinant yeast (mutant) calmodulins, using ESI-MS [109]. In this study, the calmodulins were digested with trypsin using varying protein/enzyme ratios. After digestion for selected time intervals, the reaction was terminated by the addition of soybean trypsin inhibitor. The peptide solution was desalted and the characteristic fragments were determined by ESI-MS (without chromatographic separation). The use of ESI-MS directly, without prior HPLC separation, makes this a particularly rapid assay method. However, the lack of an HPLC separation makes the technique more susceptible to matrix interference (e.g., salts) and the quantitative capability of such a method is therefore relatively limited. Nonetheless, this approach provides significant information regarding the structure of the recombinant protein.

3.5.3. Charge state

The charge-state distribution observed for a protein in ESI-MS can provide significant information regarding protein folding. For example, ESI-MS has been used to study the mechanism of refolding of acid-denatured myoglobin [110]. In this study, the ESI-MS analysis was conducted at neutral pH, since the heme group was demonstrated to remain attached to the protein under these conditions. Myoglobin was initially denatured in 10% acetic acid and then refolded by adjusting to various pH levels (5–8) by addition of ammonia solution. The molecular mass and charge-state distribution of the partially, or fully, refolded protein was then determined by ESI-MS analysis. Based on the pH dependence of the observed molecular mass (affected by the presence or absence of the heme group) and charge state distribution, the ESI-MS data indicated that myoglobin refolds in two major steps. The unfolded polypeptide chain first refolds to form a "native-like" structure without the heme group, and the binding cavity of this structure then binds the heme group by noncovalent interaction. This example clearly demonstrates that ESI-MS can provide useful information regarding protein folding pathways, and this information can be used in conjunction with other more traditional techniques (e.g., circular dichroism) to obtain a more complete understanding of these complex phenomena.

3.5.4. Cross-linking reagents

Addition of cross-linking reagent to proteins, followed by subsequent structural characterization of the linked domains, has been a traditional approach used to probe protein secondary and tertiary structure. Mass spectrometry clearly can provide an additional analytical tool to be used in this approach. The interfacing domains in recombinant human erythropoietin (EPO) have been characterized in this manner [111]. In this case, amino groups were selectively cross-linked by specific cross-linkers such as disuccinimidyl suberate or dithiobis(succinimidyl propionate). The linked regions were then characterized by trypsin digestion followed by HPLC separation. The isolated tryptic peptides were characterized

by sequencing (Edman degradation) and mass spectrometry. These data were used to determine which lysine groups in EPO were adjacent in the three-dimensional structure. The study also demonstrated that non-glycosylated and glycosylated forms of EPO have a high degree of similarity with regard to the protein conformation. A similar approach has been described for studying the subunit association of proteins [112]. In this case the proteins are cross-linked with glutaraldehyde and the molecular mass of the adduct is determined by MALDI-MS. The validity of this approach was confirmed by analyzing a variety of proteins having various known states of association.

References

[1] A.L. Burlingame, T.A. Baillie and D.H. Russell, *Anal. Chem.*, 64 (1992) 467R.
[2] S.A. Carr, *Adv. Drug. Deliv. Rev.*, 4 (1990) 113.
[3] S.A. Carr, M.E. Hemling, M.F. Bean and G.D. Roberts, *Anal. Chem.*, 63 (1991) 2802.
[4] J.A. McCloskey (Editor), *Methods Enzymol.*, 193 (1990) 960.
[5] K. Biemann, *Mass Spectrometry, Applications to Organic Chemistry*, McGraw-Hill, New York, 1962.
[6] M.S.B. Munson and F.H. Field, *J. Am. Chem. Soc.*, 88 (1966) 2621.
[7] M. Barber, R.S. Bordoli, R.D. Sedgwick and A.N. Taylor, *Nature*, 293 (1981) 270.
[8] M. Barber, R.S. Bordoli, R.D. Sedgwick, G.J. Elliot and A.N. Taylor, *Anal. Chem.*, 54 (1982) 645A.
[9] T. Keough, *Anal. Chem.*, 58 (1985) 2027.
[10] J.A. Sunner, R. Kulatunga and P. Kebarle, *Anal. Chem.*, 58 (1986) 1312.
[11] S. Naylor, A.F. Findeis, B.W. Gibson and D.H. Williams, *J. Am. Chem. Soc.*, 108 (1986) 6359.
[12] L.D. Detter, O.W. Hand, R.G. Cooks and R.A. Walton, *Mass Spectrom. Rev.*, 7 (1988) 465.
[13] R.M. Caprioli, W.T. Moore and T. Fan, *Rapid Commun. Mass Spectrom.*, 1 (1987) 15.
[14] H.D. Beckey, *Principles of Field Ionization and Field Desorption Mass Spectrometry (International Series in Analytical Chemistry*, Vol. 61), Pergamon Press, Oxford, 1977.
[15] R.J. Bueler, E. Flanigan, L.J. Greene and L. Friedman, *J. Am. Chem. Soc.*, 96 (1974) 3990.
[16] D.F. Torgerson, R.P. Skowronski and R.D. Macfarlane, *Biophys. Res. Commun.*, 60 (1974) 616.
[17] G. Jonsson, A. Hedin, P. Hakansson, B.U.M. Sunqvist, H. Bennich and P. Roepstorff, *Rapid Commun. Mass Spectrom.*, 3 (1989) 190.
[18] A. Tsabopoulos, G.W. Becker, J.L. Occolowitz and I. Jardine, *Anal. Chem.*, 171 (1988) 113.
[19] N. Furstenau, W. Knippelberg, F.R. Krueger, G. Weisz and K. Wien, *Z. Naturforsch., Teil A*, 32 (1977) 711.
[20] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida and T. Yoshida, *Rapid Commun. Mass Spectrom.*, 2 (1988) 151.
[21] M. Karas and F. Hillenkamp, *Anal. Chem.*, 60 (1988) 2299.
[22] R.C. Beavis and B.T. Chait, *Rapid Commun. Mass Spectrom.*, 3 (1989) 233.
[23] D.S. Nagra and L. Li, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 671a.
[24] R.M. Whittal, D.S. Nagra and L. Li, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 672a.
[25] D.S. Nagra and L. Li, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 987.
[26] R.M. Whittal, D.S. Nagra and L. Li, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 985.
[27] M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson and M.B. Alice, *J. Chem. Phys.*, 49 (1968) 2240.
[28] M. Yamashita and J.B. Fenn, *J. Phys. Chem.*, 88 (1984) 4451 and 4671.
[29] M.L. Alexandrov, L.N. Gall, N.V. Krasnov, V.I. Nikolaev, V.A. Pavlenko and V.A. Shkurov, *J. Anal. Chem. USSR*, 40 (1985; transl. 1986) 1227.
[30] S.F. Wong, C.K. Meng and J.B. Fenn, *J. Phys. Chem.*, 92 (1988) 546.
[31] T.R. Covey, R.F. Bonner, B.I. Shushan and J. Henion, *Rapid Commun. Mass Spectrom.*, 2 (1988) 249.
[32] M. Mann, C.K. Meng and J.B. Fenn, *Anal. Chem.*, 61 (1989) 1702.
[33] J.V. Iribane and B.A. Thomson, *J. Chem. Phys.*, 64 (1976) 2287; 71 (1979) 4451.
[34] C. Schmelzeisen-Redeker, L. Bueftering, and F.W. Röllgen, *Int. J. Mass Spectrom. Ion. Processes*, 30 (1967) 663.
[35] M.A. Abbas and J. Latham, *J. Fluid Mech.*, 30 (1967) 663.
[36] S.K. Chowdhury, V. Katta and B.T. Chait, *Rapid Commun. Mass Spectrom.*, 4 (1990) 81.
[37] J.A. Loo, H.R. Udseth and R.D. Smith, *Anal. Biochem.*, 179 (1989) 404.
[38] O.A. Mirgorodskaya, A.A. Shevchenko, I.V. Chernushevich, A.F. Dodonov and A.I. Miroshnikov, *Anal. Chem.*, 66 (1994) 99.
[39] K.D. Henry, E.R. Williams, B.H. Wang, F.W. MacLafferty, J. Shabanowitz and D.F. Hunt, *Proc. Natl. Acad. Sci. U.S.A.*, 86 (1989) 9075.
[40] T.R. Covey, R.F. Bonner, B.I. Shushan and J. Henion, *Rapid Commun. Mass Spectrom.*, 2 (1988) 249.

[41] H.R. Morris, M. Panico, M. Barber, R.S. Bordoli, R.D. Sedgwick and A.N. Taylor, *Biochem. Biophys. Res. Commun.*, 101 (1981) 623.
[42] K. Kobayashi and J.A. Smith, *J. Biol. Chem.*, 262 (1987) 11435.
[43] G. Rhadhakrishna and F. Wold, *J. Protein Chem.*, 9 (1990) 309.
[44] S. Tsunasawa, H. Takakura and F. Sakiyama, *J. Protein Chem.*, 9 (1990) 265.
[45] B.W. Gibson and K. Biemann, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 1956.
[46] K. Biemann, *Biomed. Environ. Mass Spectrom.*, 16 (1988) 99.
[47] S.A. Carr, M.E. Hemling, M.F. Bean and G.D. Roberts, *Anal. Chem.*, 63 (1991) 2802.
[48] V. Ling, A.W. Guzzetta, E. Canova-Davis, J.T. Stults, W.S. Hancock, T.R. Covey and B.I. Shushan, *Anal. Chem.*, 63 (1991) 2909.
[49] K. Biemann, *Annu. Rev. Biochem.*, 61 (1992) 977.
[50] J.T. Stults, T.M. Billeci, W.J. Henzel and S.C. Wong, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 421a.
[51] S.R. Weinberger, K.O. Boernsen, J.W. Finchy, V. Robertson and B.D. Musselman, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 775a.
[52] G.B. Hurst, T.J. Czartoski and M.V. Buchanan, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 975.
[53] F. Xiang and R.C. Beavis, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, 1994, p. 974.
[54] K.G. Krishna and F. Wold, in K. Imahori (Editor), *Methods in Protein Sequence Analysis*, Plenum Press, New York, 1993, pp. 167–172.
[55] T. Covey, B. Shushan, R. Bonner and W. Schröder, in H. Jörnvall, J. Höög and A. Gustavson (Editors), *Methods in Protein Sequence Analysis*, 1991, p. 249–256.
[56] R.M. Caprioli, B. Whaley, K.K. Mock and J.S. Cottrell, in J.J. Villafranca (Editor), *Techniques in Protein Chemistry II*, Academic Press, San Diego, 1991, pp. 497–510.
[57] K. Rose, M.G. Simona, R.E. Offord, C.P. Prior, B. Otto and D.R. Thatcher, *Biochem. J.*, 215 (1983) 273.
[58] C.W. Sutton, J.A. O'Neill and J.S. Cottrell, *Anal. Biochem.*, 218 (1994) 34.
[59] K.L. Hsi, L. Chen, D.H. Hawke, L.R. Zieske and P.M. Yuan, *Anal. Chem.*, 198 (1991) 238.
[60] D.H. Hawke, K.L. Hsi, L.R. Zieske, L. Chen and P.M. Yuan, in J.J. Villafranca (Editor), *Techniques in Protein Chemistry III*, Academic Press, San Diego, 1992, pp. 315–326.
[61] J. Gonzales, T. Takao, H. Hori, V. Besada, R. Rodriguez, G. Padron and Y. Shimonishi, *Anal. Biochem.*, 205 (1992) 151.
[62] S.A. Carr, M.J. Huddleston and M.F. Bean, *Protein Sci.*, 2 (1993) 183.
[63] G.W. Becker, P.M. Tackitt, W.W. Bromer, D.S. Lefeber and R.M. Riggin, *Biotechnol. Appl. Biochem.*, 10 (1988) 326.
[64] E. Canova-Davis, M. Eng, V. Mukku, D.H. Reifsnyder, C.V. Olson and V.T. Ling, *Biochem. J.*, 285 (1992) 207.
[65] S.A. Carr, M.F. Bean, M.E. Hemling and G.D. Roberts, in A.L. Burlingame and J.A. McCloskey (Editors), *Biological Mass Spectrometry*, Elsevier, Amsterdam, 1990, pp. 621–652.
[66] V.J. Wroblewski, R.E. Kaiser and G.W. Becker, *Pharm. Res.*, 10 (1993) 1106.
[67] V.J. Wroblewski, M. Masnyk and R.E. Kaiser, *Diabetes*, 42 (1993) 1407.
[68] A.M. Gilles, A. Lecroisey and B. Keil, *Eur. J. Biochem.*, 145 (1984) 469.
[69] T.W. Thannhauser, Y. Yonishi and H.A. Scheraga, *Methods Enzymol.*, 143 (1987) 115.
[70] Y. Yazdanparast, P. Andrews, D.L. Smith and J.E. Dixon, *Anal. Biochem.*, 153 (1986) 348.
[71] D.J. Sall and R.E. Kaiser, Jr., *J. Med. Chem.*, 36 (1993) 2350.
[72] V. Kaartinen, J.C. Williams, J. Tomich, J.R. Yates, III, L.E. Hood and I. Mononen, *J. Biol. Chem.*, 266 (1991) 5860.
[73] D. Hernandez, J.G. Stroh and A.T. Phillips, *Arch. Biochem. Biophys.*, 307 (1993) 126.
[74] J.R. Yates, III, S. Speicher, P.R. Griffin and T. Hunkapiller, *Anal. Biochem.*, 214 (1993) 397.
[75] R.A. Yost and R.K. Boyd, *Methods Enzymol.*, 193 (1990) 154.
[76] P. Roepstorff and J. Fohlman, *Biomed. Mass Spectrom.*, 11 (1984) 601.
[77] A.A. Tuinman and G.R. Pettit, *Int. J. Pept. Protein. Res.*, 36 (1990) 331.
[78] R.S. Johnson, S.A. Martin and K. Biemann, *Int. J. Mass Spectrom. Ion Processes*, 86 (1988) 137.
[79] J.T. Stults and J.T. Watson, *Biomed. Environ. Mass Spectrom.*, 14 (1987) 583.
[80] R.S. Johnson, S.A. Martin, K. Biemann, J.T. Stults and J.T. Watson, *Anal. Chem.*, 59 (1987) 2621.
[81] L.C.E. Taylor and L. Poulter, in *Proceedings of the American Society for Mass Spectrometry*, American Society for Mass Spectrometry, San Francisco, 1988.
[82] H.A. Scoble, S.A. Martin and K. Biemann, *Biomed. J.*, 245 (1987) 621.
[83] S.A. Martin and K. Biemann, *Int. J. Mass Spectrom. Ion Processes*, 78 (1987) 213.
[84] R.S. Johnson and K. Biemann, *Biochemistry*, 26 (1987) 1209.
[85] T.D. Lee, *MacProMass*, Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, CA.

[86] J.-C. Gesquiere, *Mac Mass*, Faculté de Pharmacie, URA 1309 CNRS, Institut Pasteur, Lille.
[87] K. Biemann, *COMPOST*, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA.
[88] B. Spengler, D. Kirsch and R. Kaufmann, *J. Phys. Chem.*, 96 (1992) 9678.
[89] R. Kaufmann, B. Spengler and F. Lützenkirchen, *Rapid Commun. Mass Spectrom.*, 7 (1993) 902.
[90] B. Spengler, D. Kirsch, R. Kaufmann and E. Jaeger, *Rapid Commun. Mass Spectrom.*, 6 (1992) 105.
[91] B. Spengler, D. Kirsch and R. Kaufmann, *Rapid Commun. Mass Spectrom.*, 5 (1991) 198.
[92] J.C. Rouse, W. Yu and S.A. Martin, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 676.
[93] R.B. Zambias, D.A. Boulton and P.R. Griffin, *Tetrahedron Lett.*, 35 (1994) 4283.
[94] J.B. Hoyes, S.G. Curbishley, P. Doorbar, P. Tatterton, R.H. Bateman, B. Beer and J. Lockett, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 682.
[95] B. Spengler and R. Kaufmann, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 774a.
[96] W. Yu, M. Flemming, I. Mazsarroff, H. Scoble and S. Martin, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 442.
[97] K. Breddam, *Carlsberg Res. Commun.*, 51 (1986) 83.
[98] K. Klarskov, K. Breddam and P. Roepstorff, *Anal. Biochem.*, 180 (1989) 38.
[99] R. Wang, R.J. Cotter, J.F. Meschia and S.S. Sisodia, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry III*, Academic Press, San Diego, 1992, pp. 505–513.
[100] C.E. Smith and K.L. Duffin, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry IV*, Academic Press, San Diego, 1993, pp. 463–470.
[101] M. Schär, K.O. Börnsen and E. Gassmann, *Rapid Commun. Mass Spectrom.*, 5 (1991) 319.
[102] K.J. Rosnack and J.G. Stroh, *Rapid Commun. Mass Spectrom.*, 6 (1992) 637.
[103] E.D. Lee, W. Mück, T.R. Covey and J.D. Henion, *J. Am. Chem. Soc.*, 111 (1989) 4600.
[104] B. Ganem, Y.T. Li and J.D. Henion, *J. Am. Chem. Soc.*, 113 (1991) 6294.
[105] B. Ganem, Y.T. Li and J.D. Henion, *J. Am. Chem. Soc.*, 113 (1991) 7818.
[106] V. Kata and B.T. Chait, *J. Am. Chem. Soc.*, 113 (1991) 8534.
[107] G.W. Becker, R.R. Bowsher, W.C. MacKellar, M.L. Poor, P.M. Tackitt and R.M. Riggin. *Biotechnol. Appl. Biochem.*, 9 (1987) 478.
[108] Z. Zhang and D.L. Smith, *Protein Sci.*, 2 (1993) 522.
[109] S.E. Brockerhoff, C.G. Edmonds and T.N. Davis, *Protein Sci.*, 1 (1992) 504.
[110] R. Feng and Y. Konishi, *J. Am. Soc. Mass Spectrom.*, 4 (1993) 638.
[111] M. Haniu, L.O. Narhi, T. Arakawa, S. Elliott and M.F. Rohde, *Protein Sci.*, 2 (1993) 1441.
[112] T.B. Farmer and R.M. Caprioli, *Biol. Mass Spectrom.*, 20 (1991) 796.