9 Mass spectrometric techniques

A. AITKEN

9.1 INTRODUCTION

9.1.1 General

Mass spectrometry (MS) is an extremely valuable analytical technique in which the molecules in a test sample are converted to gaseous ions that are subsequently separated in a mass spectrometer according to their **mass-to-charge** (m/z) **ratio** and detected. The **mass spectrum** is a plot of the (relative) abundance of the ions at each m/z ratio. Note that it is the mass to charge ratios of ions (m/z) and not the actual mass that is measured. If for example, a biomolecule is ionised by the addition of one or more protons (H^+ ions) the instrument measures the m/z after addition of 1 Da for each proton if the instrument is measuring positive ions or m/z minus 1 Da for each proton lost if measuring negative ions.

The development of two ionisation techniques, **electrospray** (ESI) and **matrix-assisted laser desorption/ionisation** (MALDI), has enabled the accurate mass determination of high-molecular-mass compounds as well as low-molecular-mass molecules and has revolutionised the applicability of mass spectrometry to almost any biological molecule. Applications include the new science of proteomics as well as in drug discovery. The latter includes combinatorial chemistry where a large number of similar molecules (combinatorial libraries) are produced and analysed to find the most effective compounds from a group of related organic chemicals.

M_r is sometimes used to designate relative molar mass. Molecular weight (which is a force not a mass) is also frequently and incorrectly used. M_r is a relative measure and

has no units. However, M_r is numerically equivalent to the mass, M, which does have units and the Dalton is frequently used (see Section 1.2.2).
The essential features of all mass spectrometers are therefore:

- production of ions in the gas phase;
- acceleration of the ions to a specific velocity in an electric field;
- separation of the ions in a mass analyser; and
- detection of each species of a particular m/z ratio.

The instruments are calibrated with standard compounds of accurately known M_r values. In mass spectrometry the carbon scale is used with $^{12}C = 12.000000$. This level of accuracy is achievable in high-resolution magnetic sector double-focussing, accelerator mass spectrometers and Fourier transform mass spectrometers (Sections 9.3.5, 9.3.6 and 9.3.13).

The mass analyser may separate ions either by use of a magnetic or an electrical field. Alternatively the time taken for ions of different masses to travel a given distance in space is measured accurately in the **time-of-flight** (TOF) **mass spectrometer** (Section 9.3.8).

Any material that can be ionised and whose ions can exist in the gas phase can be investigated by MS, remembering that very low pressures, i.e. high vacuum, in the region of 10^{-6} Torr are required (Torr is measure of pressure which equals 1 mm of mercury (133.3 Pa; atmospheric pressure is 760 Torr)). The majority of biological MS investigations on proteins, oligosaccharides and nucleic acids is carried out with **quadrupole, quadrupole–ion trap** and TOF mass spectrometers. In the organic chemistry/biochemistry area of analysis, the well-established magnetic sector mass spectrometers still find wide application and their main principles will also be described.

The treatment of mass spectrometry in this chapter will be strictly non-mathematical and non-technical. However, the intention is to give an overview of the types of instrumentation that will be employed, the main uses of each, complementary techniques and advantages/disadvantages of the different instruments and particular applications most suited to each type. Data analysis and sample preparation to obtain the best sensitivity for a particular type of compound will also be covered.

9.1.2 Components of a mass spectrometer

All mass spectrometers are basically similar (Fig. 9.1). They consist of the following:

- *A high vacuum system (10^{-6} torr or 1 μtorr)*: These include **turbomolecular pumps**, **diffusion pumps** and **rotary vane pumps**.
- *A sample inlet*: This comprises a sample or target plate; a high-performance liquid chromatography (HPLC), gas chromatography (GC) or capillary electrophoresis system; solids probe; **electron impact** or **direct chemical ionisation** chamber.
- *An ion source* (to convert molecules into gas-phase ions): This can be MALDI; ESI; **fast atom bombardment** (FAB); electron impact or direct chemical ionisation.
- *A mass filter/analyser*: This can be: TOF; quadrupole; ion trap; magnetic sector or **ion cyclotron Fourier transform** (the last is also actually a detector).
- *A detector*: This can be a **conversion dynode**, **electron multiplier**, **microchannel plate** or **array detector**.

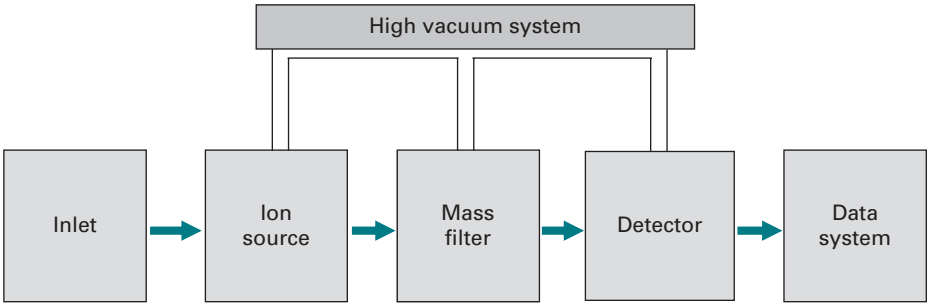


Fig. 9.1 Basic components of mass spectrometers.

9.1.3 Vacuum system

All mass analysers operate under vacuum in order to minimise collisions between ions and air molecules. Without a high vacuum, the ions produced in the source will not reach the detector. At atmospheric pressure, the mean free path of a typical ion is around 52 nm; at 1 mtorr, it is 40 mm; and at 1 µtorr, it is 40 m.

In most instruments, two vacuum pump types are used, e.g. a rotary vane pump (to produce the main reduction in pressure) followed by a turbomolecular pump or diffusion pump to produce the high vacuum.

The rotary vane pump can be an oil pump to provide initial vacuum (approximately 1 torr), while the turbomolecular pump provides working high vacuum (1 mtorr to 1 ntorr). This is a high-speed gas turbine with interspersed rotors (moving blades) and stators (i.e. fixed or stationary blades) whose rotation forces molecules through the blade system.

9.2 IONISATION

Ions may be produced from a neutral molecule by removing an electron to produce a positively charged **cation**, or by adding an electron to form an **anion**. Both positive- and negative-ion mass spectrometry may be carried out but the methods of analysis in the following sections will be described mainly for positive-ion MS, since this is more common and the principles of separation and detection are essentially the same for both types of ion.

9.2.1 Electron impact ionisation (EI)

Electron impact ionisation (EI) is widely used for the analysis of metabolites, pollutants and pharmaceutical compounds, for example in drug testing programmes. Electron impact (EI) has major applications as a mass detector for gas chromatography (GC/MS, Section 11.9.3). A stream of electrons from a heated metal filament is accelerated to 70 eV potential (the electron volt, eV, is a measure of energy). Sample ionisation occurs when the electrons stream across a high vacuum chamber into which molecules of the substance to be analysed (**analyte**) are allowed to diffuse (Fig. 9.2). Interaction with the analyte results in either loss of an electron from the

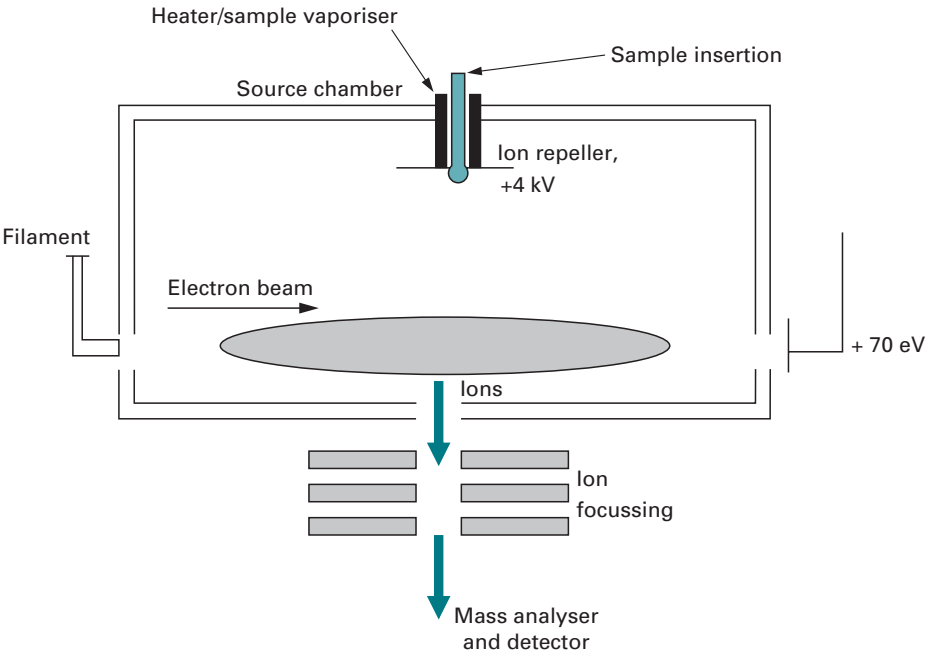


Fig. 9.2 Electron impact source. Electrons are produced by thermionic emission from a filament of tungsten or rhenium. The filament current is typically 0.1 mA. Electrons are accelerated toward the ion source chamber (held at a positive potential equal to the accelerating voltage) and acquire an energy equal to the voltage between the filament and the source chamber, typically 70 eV. The electron trap is held at a fixed positive potential with respect to the source chamber. Gaseous analyte molecules are introduced into the path of the electron beam where they are ionised. Owing to the positive ion repeller voltage and the negative excitation voltage that produce an electric field in the source chamber, the ions leave the source through the ion exit slit and are analysed.

substance (to produce a cation) or electron capture (to produce an anion). The analyte must be in the vapour state in the electron impact source, which limits the applicability to biological materials below ca. 400 Da. Before the advent of electrospray and MALDI, the method did have some applicability to peptides, for example, whose volatility could be increased by chemical modification. A large amount of fragmentation of the sample is common, which may or may not be desirable depending on the information required.

Chemical bonds in organic molecules are formed by the pairing of electrons. Ionisation resulting in a cation requires loss of an electron from one of these bonds (effectively knocked out by the bombarding electrons), but it leaves a bond with a single unpaired electron. This is a radical as well as being a cation and hence the representation as $M^{\cdot+}$, the ($^+$) sign indicating the ionic state and the ($^\cdot$) a radical. Conversely, electron capture results both in an anion but also the addition of an unpaired electron and therefore a negatively charged radical, hence the symbol $M^{\cdot-}$. Such radical ions are termed **molecular ions, parent ions** or **precursor ions** and under the conditions of electron bombardment are relatively unstable. Their energy in excess of that required for ionisation has to be dissipated. This latter process results in the

precursor ion disintegrating into a number of smaller **fragment ions** that may be relatively unstable and further fragmentation may occur. This gives rise to a series of **daughter ions** or **product ions**, which are recorded as the mass spectrum.

For the production of a radical cation, as it is not known where either the positive charge or the unpaired electron actually reside in the molecule, it has been the practice to place the dot signs outside the abbreviated bracket sign, $^{\cdot}]^{\cdot}$. The recent recommendation by IUPAC for mass spectrometry notation is to write the sign first followed by the superscripted dot, i.e. $M^{+\cdot}$ or $M^{-\cdot}$.

When the precursor ion fragments, one of the products carries the charge and the other the unpaired electron, i.e. it splits into a radical and an ion. The product ions are therefore true ions and not radical ions. The radicals produced in the fragmentation process are neutral species and therefore do not take any further part in the mass spectrometry but are pumped away by the vacuum system. Only the charged species are accelerated out of the source and into the mass analyser. It is also important to recognise that almost all possible bond breakages can occur and any given fragment will arise both as an ion and a radical. The distribution of charge and unpaired electron, however, is by no means equal. The distribution depends entirely on the thermodynamic stability of the products of fragmentation. Furthermore, any fragment ion may break down further (until single atoms are obtained) and hence not many ions of a particular type may survive, resulting in a low signal being recorded.

A simple example is given by *n*-butane ($CH_3CH_2CH_2CH_3$) and some of the major fragmentations are shown Fig. 9.3.a. The resultant EI spectrum is shown in Fig. 9.3b.

9.2.2 Chemical ionisation

Chemical ionisation (CI) is used for a range of samples similar to those for EI. It is particularly useful for the determination of molecular masses, as high intensity molecular ions are produced due to less fragmentation. CI therefore gives rise to much cleaner spectra. The source is essentially the same as the EI source but it contains a suitable reagent gas such as methane (CH_4) or ammonia (NH_3) that is initially ionised by EI. The high gas pressure in the source results in ion–molecule reactions between reagent gas ions (such as NH_3^+ and CH_4^+) some of which react with the analyte to produce analyte ions. The mass differences from the neutral parent compounds therefore correspond to these adducts.

9.2.3 Fast atom bombardment (FAB)

At the time of its development in the early 1980s, fast atom bombardment (FAB) revolutionised MS for the biologist. The important advance was that this **soft ionisation** technique, which leads to the formation of ions with low internal energies and little consequent fragmentation, permitted analysis of biomolecules in solution without prior derivatisation. The sample is mixed with a relatively involatile, viscous matrix such as glycerol, thioglycerol or *m*-nitrobenzyl alcohol. The mixture, placed on a probe, is introduced into the source housing and bombarded with an **ionising beam** of neutral

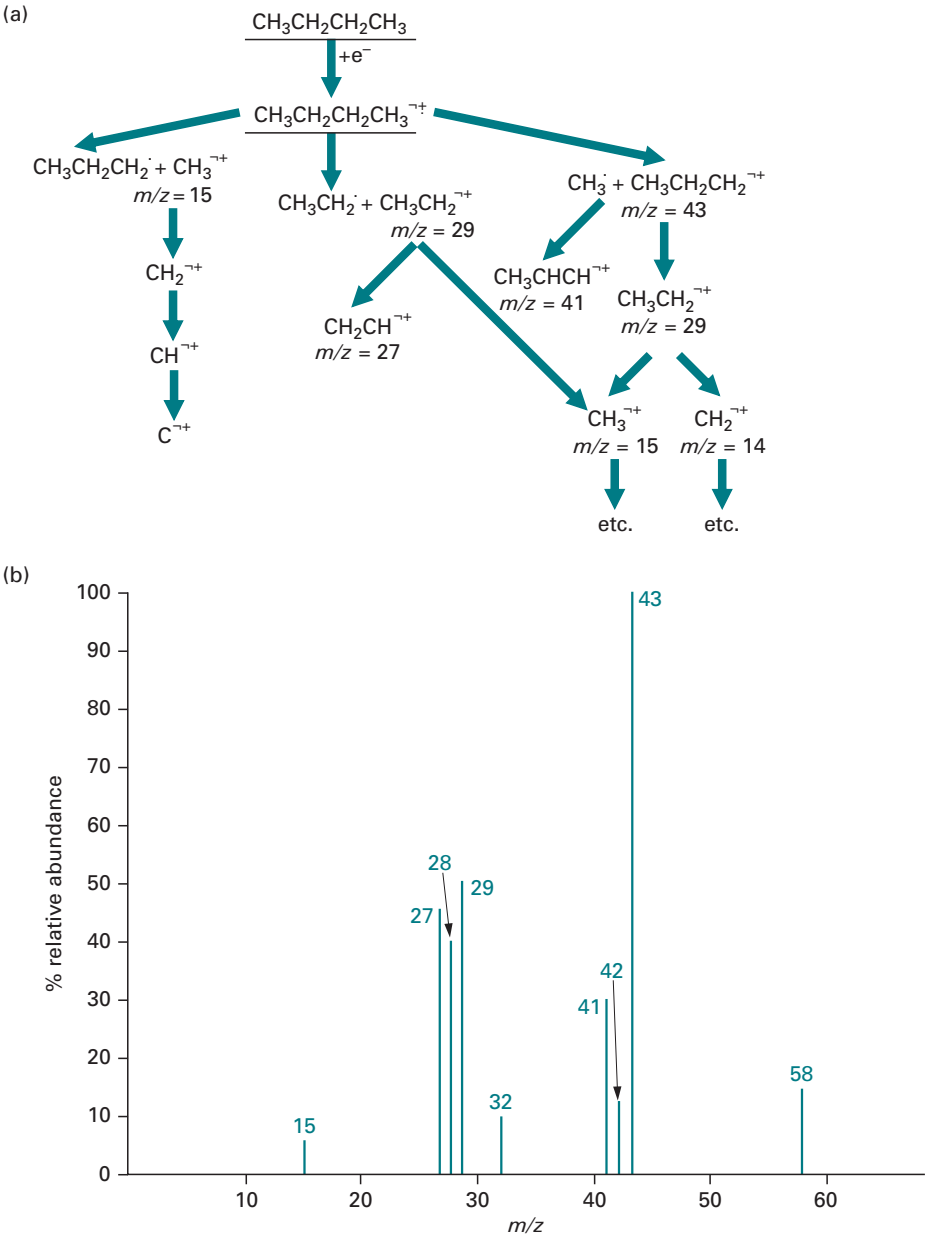


Fig. 9.3 Fragmentation pathways in *n*-butane and the EI spectrum. The pathway for fragmentation of *n*-butane is shown in (a) and the EI spectrum in (b). In the spectrum, the relative abundance is plotted from 0 to 100% where the largest peak is set at 100% (base peak). Spectra represented in this way are said to be **normalised**.

atoms (such as Ar, He, Xe) of high velocity. A later development was the use of a beam of caesium (Cs^+) ions and the term **liquid secondary ion mass spectrometry** (LSIMS) was introduced to distinguish this from FAB–MS. **Pseudomolecular ion species** arise as either protonated or deprotonated entities $(M + H)^+$ and $(M - H)^-$ respectively, which allows

positive and negative ion mass spectra to be determined. The term pseudomolecular implies the mass of the ion formed from a substance of a given mass by the gain or loss of one or more protons. Other charged adducts can also be formed such as $(M + Na)^+$ and $(M + K)^+$.

9.2.4 Electrospray ionisation (ESI)

This involves the production of ions by spraying a solution of the analyte into an electrical field. This is a soft ionisation technique and enables the analysis of large intact (underivatised) biomolecules, such as proteins and DNA. The electrospray (ES) creates very small droplets of solvent-containing analyte. The essential principle in ES is that a spray of charged liquid droplets is produced by atomisation or nebulisation. Solvent (typically 50 : 50 water and organic solvent) is removed as the droplets enter the mass spectrometer. ESI is the result of the strong electric field (around 4 keV at the end of the capillary and 1 keV at the counter electrode) acting on the surface of the sample solution. As the solvent evaporates in the high-vacuum region, the droplet size decreases and eventually charged analyte (free of solvent) remains. Ionisation can occur at atmospheric pressure and this method is also sometimes referred to as atmospheric pressure ionisation (API).

The concentration of sample is usually around 1–10 pmol mm^{-3}. Typical solvents are 50/50 acetonitrile (or methanol)/H_2O with 1% acetic acid or 0.1% formic acid. Ammonium hydroxide or trifluoroacetic acid (TFA, 0.02%) in 50/50 acetonitrile (or methanol)/H_2O can also be used. The organic acid (or the NH_4OH) aids ionisation of the analyte. At low pH, basic groups will be ionised. In the example of peptides these are the side groups of Lys, His, Arg and the N-terminal amino group. At alkaline pH the carboxylic acid side chains as well as stronger anions such as phosphate and sulphate groups will be ionised. The presence of organic solvent assists in formation of small droplets and facilitates evaporation.

The flow rate into the source is normally around a few mm^3 min^{-1} although higher flow rates can be tolerated (up to 1 cm^3) if the solution is an eluant from on-line HPLC for example.

Smaller molecules usually produce singly charged ions but multiply charged ions are frequently formed from larger biomolecules, in contrast to MALDI, resulting in m/z ratios that are sufficiently small to be observed in the quadrupole analyser. Thus masses of large intact proteins, DNA and organic polymers can also be accurately measured in electrospray MS although the m/z limit of measurement is normally 2000 or 3000 Da. For example, proteins are normally analysed in the positive ion mode where charges are introduced by addition of protons. The number of basic amino acids in the protein (mainly lysine and arginine) determines the maximum number of charges carried by the molecule. The distribution of basic residues in most proteins is such that the multiple peaks (one for each $M + nH)^{n+}$ ion, are centred on m/z about 1000. In Fig. 9.6 a large protein with a mass of over 100 000 Da behaves as if it were multiple mass species around 1020 Da. For the species with 100 protons (H^+) i.e. with 100 charges, $z = 100$, $m/z = 1027.6$ therefore $(M + 100H)^{100+} = 1027.6$. When the computer processes the data for the multiple peaks, the average for each set of peaks gives a mass determination

Example 1 **PROTEIN MASS DETERMINATION BY ESI**

Question A protein was isolated from human tissue and subjected to a variety of investigations. Relative molecular mass determinations gave values of approximately 12 000 by size exclusion chromatography and 13 000 by gel electrophoresis. After purification, a sample was subjected to electrospray ionisation mass spectrometry and the following data obtained.

m/z	773.9	825.5	884.3	952.3	1031.3
Abundance (%)	59	88	100	66	37

Given that $n_2 = (m_1 - 1)/(m_2 - m_1)$ and $M = n_2\,(m_2 - 1)$ and assuming that the only ions in the mixture arise by protonation, deduce an average molecular mass for the protein by this method.

Answer M_r by exclusion chromatography = 12 000

M_r by gel electrophoresis = 13 000

Taking ESI peaks in pairs:

$m_1 - 1$	$m_2 - m_1$	n_2	$m_2 - 1$	M (Da)	z
951.3	79.0	12.041	1030.3	12406.6	12
883.3	68.0	12.989	951.3	12357.1	13
824.5	58.8	14.022	883.3	12385.7	14
772.9	51.6	14.978	824.5	12349.9	15

ΣM = 49 499.3 Da

Mean M = 12 374.8 Da

Note: Relative abundance values are not required for the determination of the mass.

to high accuracy. The peaks can be **deconvoluted** and presented as a single peak representing the M_r (in this example M = 102 658).

A diagrammatic representation of the ESI source is shown in Fig. 9.4. A **curtain** or **sheath gas** (usually nitrogen) around the spray needle at a slow flow rate may be used to assist evaporation of the solvent at or below room temperature. This may be an advantage for thermally labile compounds.

9.3 MASS ANALYSERS

9.3.1 Introduction

Once ions are created and leave the ion source, they pass into a mass analyser, the function of which is to separate the ions and to measure their masses. (Remember, what

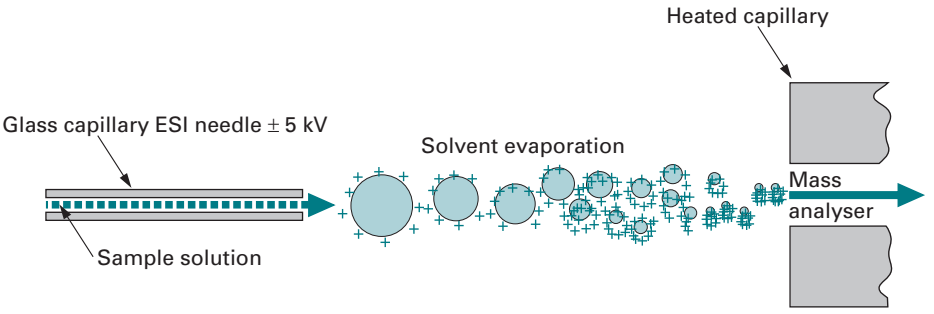


Fig. 9.4 Electrospray ionisation source. The ESI creates very small droplets of solvent-containing analyte by atomisation or **nebulisation** as the sample is introduced into the source through the fine glass (or other material) hollow needle capillary. The solvent evaporates in the high-vacuum region as the spray of droplets enters the source. As the result of the strong electric field acting on the surface of the sample droplets, and electrostatic repulsion, their size decreases and eventually single species of charged analyte (free of solvent) remain. These may have multiple charges depending on the availability of ionisable groups.

is really measured is the mass-to-charge ratio (m/z) for each ion.) At any given moment, ions of a particular mass are allowed to pass through the analyser where they are counted by the detector. Subsequently, ions of a different mass are allowed to pass through the analyser and again the detector counts the number of ions. In this way, the analyser scans through a large range of masses.

In the majority of instruments, a particular type of ionisation is coupled to a particular mass analyser that operates by a particular principle. That is, EI, CI and FAB are combined with magnetic sector instruments; ESI and its derivatives with quadrupole (or its variant ion-trap) and MALDI is coupled to TOF detection.

9.3.2 Quadrupole mass spectrometry

The quadrupole analyser consists of four parallel cylindrical rods (Fig. 9.5). A direct current (DC) voltage and a superimposed radio frequency (RF) voltage are applied to each rod, creating a continuously varying electric field along the length of the analyser. Once in this field, ions are accelerated down the analyser towards the detector. The varying electric field is precisely controlled so that during each stage of a scan, ions of one particular mass-to-charge ratio pass down the length of the analyser. Ions with any other mass-to-charge value impact on the quadrupole rods and are not detected. By changing the electric field (**scanning**), the ions of different m/z successively arrive at the detector.

Quadrupoles can routinely analyse up to m/z 3000, which is extremely useful for biological MS since, as we have seen, proteins and other biomolecules normally give a charge distribution of m/z that is centred below this value (see Fig. 9.6).

Note that **hexapole** and **octapole** devices are also used, to direct a beam into the next section of a **triple quadrupole** or into the ion trap for example, but the principle is the same.

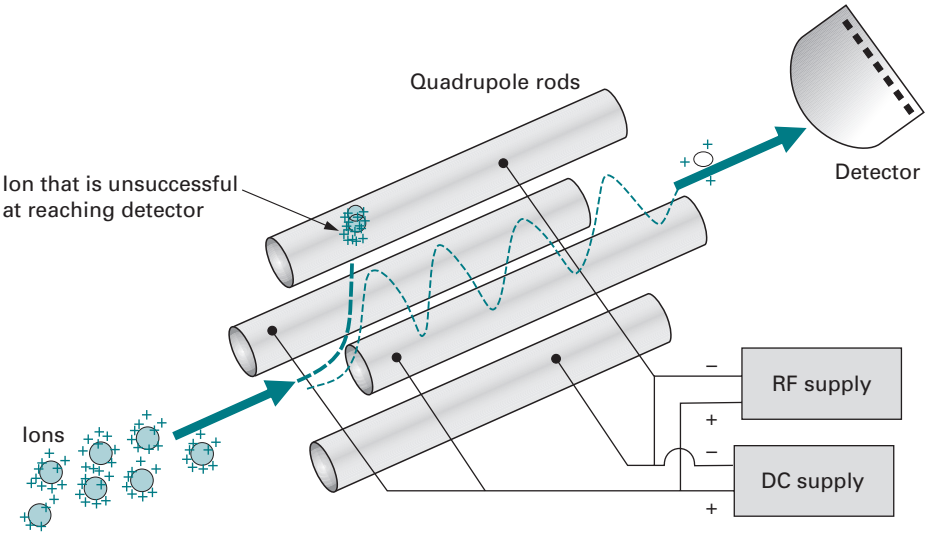


Fig. 9.5 Quadrupole analyser. The fixed (DC) and oscillating (RF) fields cause the ions to undergo complicated trajectories through the quadrupole filter. For a given set of fields, only certain trajectories are stable, which only allows ions of specific *m/z* to travel through to the detector. The efficiency of the quadrupole is impaired after a build-up of ions that do not reach the detector. Therefore a set of pre-filters is added to the quadrupole to remove the ions that would otherwise affect the main quadrupole.

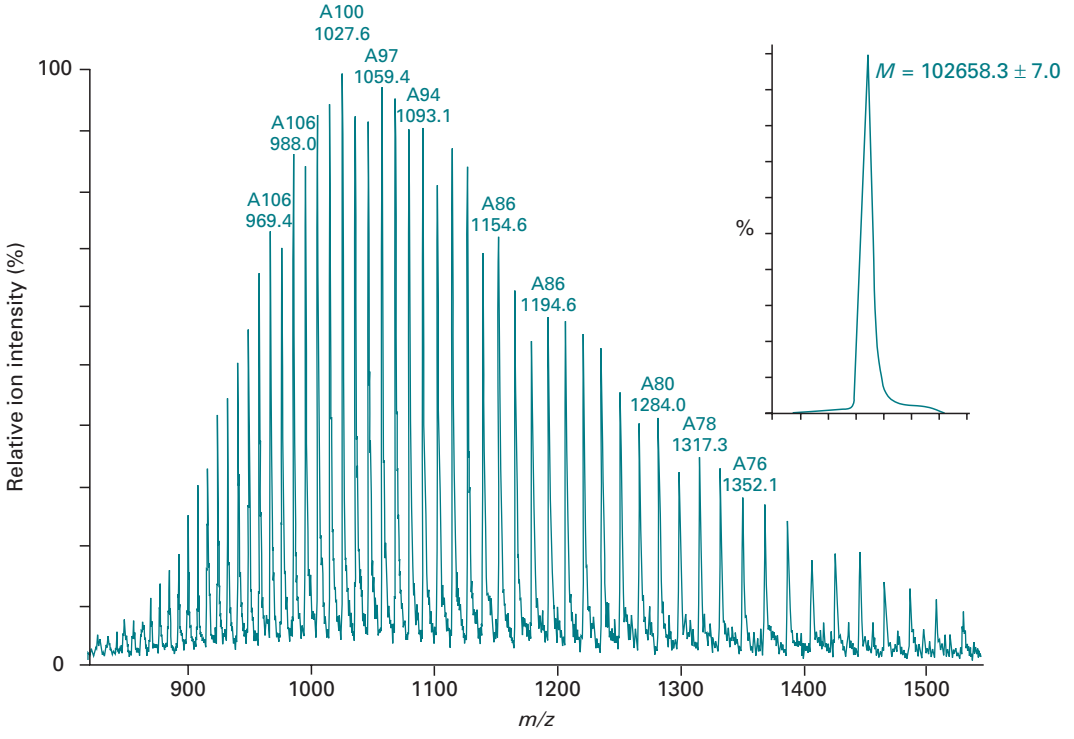


Fig. 9.6 Large intact protein mass accurately measured in electrospray MS. The species of ions are annotated by the charge state, e.g. with 99, 100, 101 charges, etc., and the associated *m/z* value. The inset shows the 'deconvoluted spectrum'.

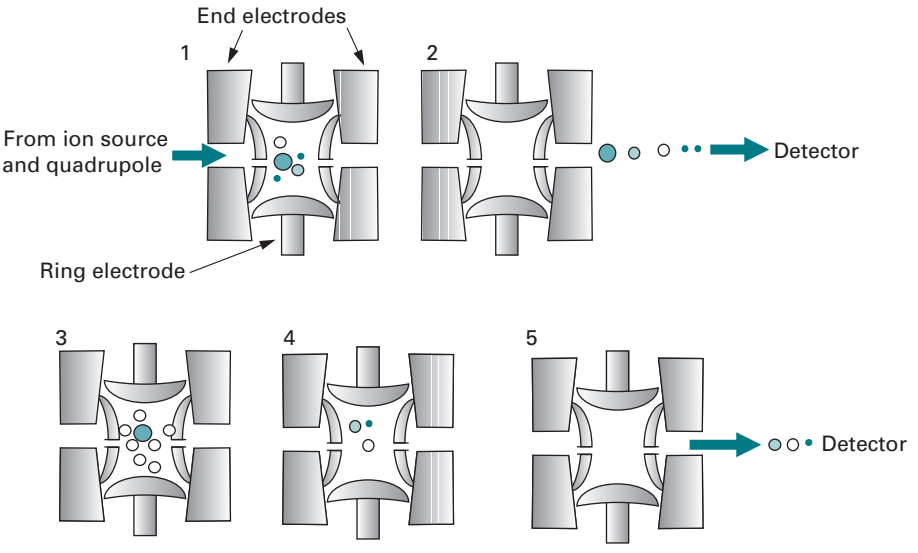


Fig. 9.7 Diagram of an ion trap. The ion trap contains three hyperbolic electrodes which form a cavity in a cylindrical device of around 5 cm diameter in which the ions are trapped (stored) and subsequently analysed. Each end-cap electrode has a small hole in the centre. Ions produced from the source enter the trap through the quadrupole and the entrance end-cap electrode. Potentials are applied to the electrodes to trap the ions (diagrams 1 and 2). The ring electrode has an alternating potential of constant radio frequency but variable amplitude. This results in a three-dimensional electrical field within the cavity. The ions are trapped in stable oscillating trajectories that depend on the potentials and the *m*/*z* of the ions. To detect these ions, the potentials are varied, resulting in the ion trajectories becoming unstable and the ions are ejected in the axial direction out of the trap in order of increasing *m*/*z* into the detector. A very low pressure of helium is maintained in the trap, which 'cools' the ions into the centre of the trap by low-speed collisions that normally do not result in fragmentation. These collisions merely slow the ions down so that during scanning, the ions leave quickly in a compact packet, producing narrower peaks with better resolution. In sequencing, all the ions are ejected except those of a particular *m*/*z* ratio that has been selected for fragmentation (see diagrams 3, 4 and 5). The steps are: (3) selection of precursor ion, (4) collision-induced dissociation of this ion, and (5) ejection and detection of the fragment ions.

9.3.3 Ion trap mass spectrometry

Ion trap mass spectrometers use ESI to produce ions, all of which are transferred into and subsequently measured almost simultaneously (within milliseconds) in a device called an **ion trap** (Fig. 9.7). The trap must then be refilled with the ions that are arriving from the source. Therefore, although the trap does not measure 100% of all ions produced (it depends on the **cycle time** to refill the trap then analyse the ions) this results nevertheless in a great improvement in sensitivity relative to quadrupole mass spectrometers where at any given moment only ions of one particular *m*/*z* are detected. ESI–ion trap mass spectrometers have found wide application for analysis of peptides and small biomolecules such as in protein identification by tandem MS; liquid chromatography/mass spectrometry (LC/MS); combinatorial libraries and rapid analysis in drug discovery and drug development.

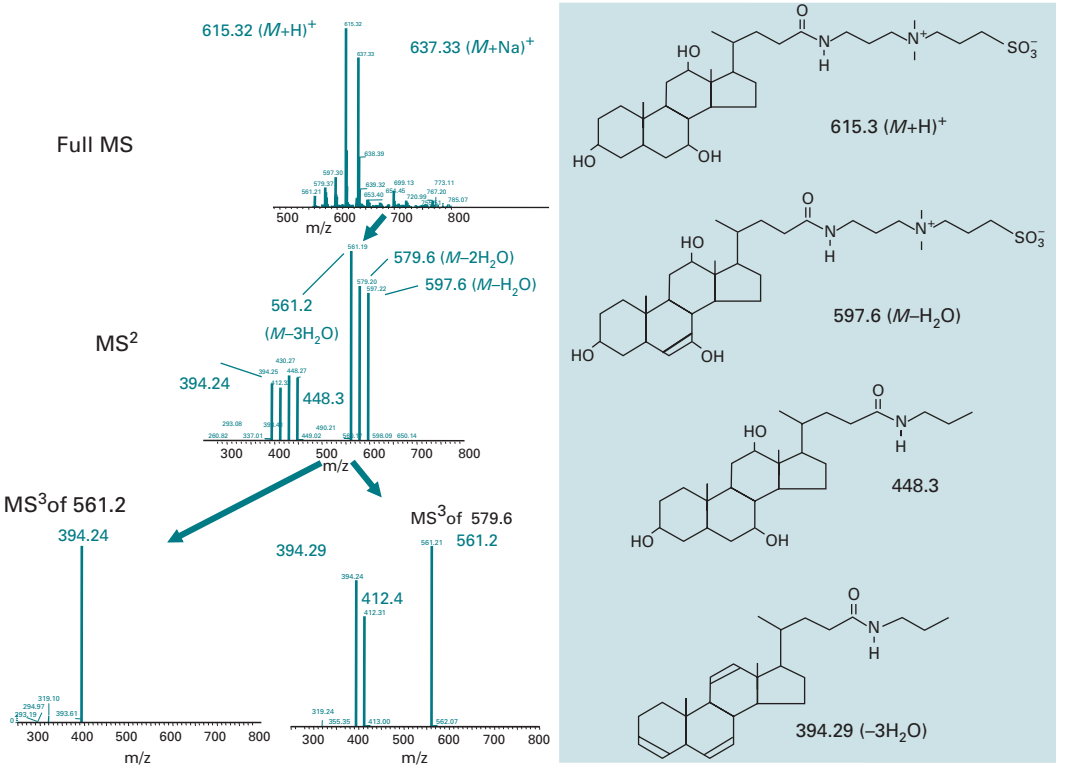


Fig. 9.8 Structural analysis, MS^n in an ion trap. In this example, of a steroid-related compound, the structure can be analysed when the $(M + H)^+$ ions at 615.3 are selected to be retained in the ion trap. These ions are subjected to collision-induced dissociation (CID) resulting in loss of the aliphatic sulphonate from the quaternary ammonium group and partial loss of some hydroxyl groups in the tandem MS (MS^2) experiment. The major fragment ions (561.2 and 579.6) are further selected for CID (MS^3) resulting in subsequent losses of more hydroxyl groups from specific parts of the steroid ring.

Ion trap MS permits structural information to be readily obtained (and sequence information in the case of polypeptides). Not only can MS–MS analysis be carried out but also due to the high efficiency of each stage, further fragmentation of selected ions may be carried out to MS to the power n (MS^n) (Fig. 9.8). The instrument still allows accurate molecular mass determination to over 100 000 Da at greater than 0.01% mass accuracy.

The **MS^n procedure** in an ion trap involves ejecting all ions that are stored in the trap, except those corresponding to the selected m/z value. To perform **tandem MS** (MS^2) a collision gas is introduced (a low pressure of helium) and **collision-induced dissociation** (CID) occurs (Fig. 9.7). The **fragment ions** are then ejected in turn and the **fragment spectrum** determined. The process can be repeated successively where all the fragment ions stored in the trap except those fragment ions corresponding to another selected m/z value are ejected. This fragment ion can then be further fragmented to obtain more structural information, as illustrated for the example shown in Fig. 9.8. This technique has a big advantage since no additional mass spectrometers

or collision cells are required. The limitation is sensitivity, which decreases with each MS experiment, although the claimed record in an ion trap is currently MS^{14}.

9.3.4 Nanospray and on-line tandem mass spectrometry

The sensitivity with ESI can be greatly improved with a reduction in flow rate. Nanospray is therefore the technique of choice for ultimate sensitivity when sample amounts are limited. There are two ways of achieving this. Both static and dynamic nanospray techniques are widely used. Flow rates in both nanospray techniques are in the order of tens of $nm^3\ min^{-1}$, which leads to low sample consumption and low signal-to-noise ratio.

Firstly, in static nanospray, glass needles are used with a very finely drawn out capillary tip (coated with gold to allow the needle to be held at the correct kV potential; see Fig. 9.4). The needles are filled with 1–2 mm^3 of sample and accurately positioned at the entrance to the source. Closed-circuit television (CCTV) is used to determine accurately the position of the capillary. The solution is drawn into the source by electrostatic pressure, although a low pressure may be applied with an air-filled syringe behind the other (open) end of the needle if necessary.

In dynamic nanospray experiments, small-diameter microbore HPLC or capillary columns are also used to achieve separation at low flow rates. This can be combined with a stream splitter device that can further reduce flow rate (Section 11.9.3). The stream splitter can be used to divert a percentage of the solvent flow from the pump, say 99% to 99.9% to waste and allow the remainder to pass through the column. This allows for much more accurate flow rates since it is extremely difficult to directly and accurately pump at 0.5 mm^3 or even 50 $nm^3\ min^{-1}$ with a high-pressure pump. Therefore one can use a pump that functions more efficiently at flow rates of 50 to 500 $mm^3\ min^{-1}$ to pass 0.5 $nm^3\ min^{-1}$ or less into the micro column.

Nanospray sources are used in triple quadrupole, ion trap and hybrid MALDI instruments.

Computer programs can be set up to perform tandem MS during the chromatographic separation on each component as it elutes from the column, if it gives a signal above a threshold that is set by the operator.

9.3.5 Magnetic sector analyser

A magnetic sector analyser is shown diagrammatically in Fig. 9.9. The ions are accelerated by an electric field. The electric sector acts as a kinetic energy filter and allows only ions of a particular kinetic energy to pass, irrespective of the m/z. This greatly increases the resolution since the ions emerge from the electrostatic analyser (ESA) with the whole range of masses but the same velocity. A given ion with the appropriate velocity then enters the magnetic sector analyser. It will travel in a curved trajectory in the magnetic field with a radius depending on the m/z and the velocity of the ion (the latter has already been selected). Thus only ions of a particular m/z will be detected at a particular magnetic field strength. The trajectory

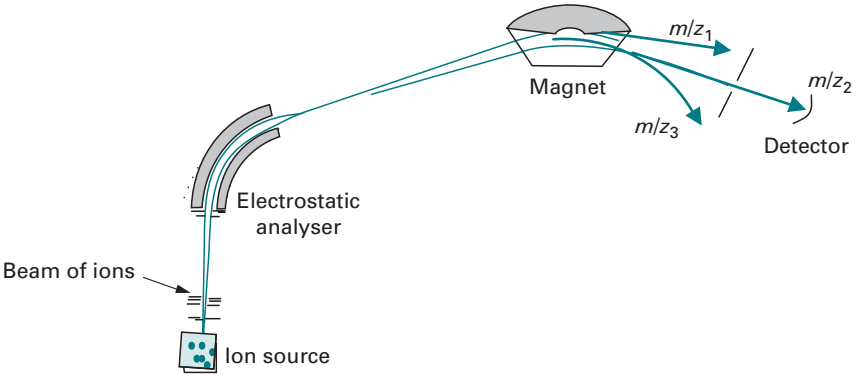


Fig. 9.9 Double-focussing magnetic sector mass spectrometer. The figure shows the 'forward geometry' arrangement where the electrostatic analyser is before the magnetic sector (known as EB; E for electric, B for magnetic). Similar results may be obtained if the reverse geometry (BE) type is used. The radial path followed by each ion is shown by scanning the magnetic field, B, and each ion of a particular *m/z* can be brought into the detector slit in turn.

of the ions is through a sector of the circular poles of the magnet, hence the term **magnetic sector.**

Figure 9.9 shows several possible trajectories for a given ion in the magnetic field. Only one set of ions will be focussed on the detector. If the field is changed, these ions will be defocussed because they will not be deflected to the correct extent. A new set of ions will be deflected and collected at the detector. By starting at either end of the magnet range the ions can be scanned from high to low mass or from low to high mass. This **magnetic scanning** is the most commonly used type of analysis in this instrument. Alternatively, the mass spectrum can be scanned electrically by varying the voltage, V, while holding the magnetic field, *B*, constant. This type of instrument is called a **two-sector** or **double-focussing** mass spectrometer and resolving power to parts per million may be obtained.

9.3.6 Accelerator mass spectrometry

Accelerator mass spectrometry (AMS) has proved to be extremely useful for quantifying rare isotopes and has had a major impact in archaeology (to measure ^{14}C) and geochronology. AMS can also measure radioisotopes such as ^{3}H, ^{10}Be, ^{26}Al, ^{36}Cl and ^{41}Ca with attomole (10^{-18}) to zeptomole (10^{-21}) levels of sensitivity and very high precision. AMS has found increasing application in human microdosing studies in drug development. This enables metabolites to be measured in human plasma or urine after administration of low, pharmacologically relevant doses of labelled drugs. Among the many applications of AMS are long-term pharmacokinetic studies to determine low-dose and chronic effects and the analysis of molecular targets of neurotoxins (see Section 18.3.1).

9.3.7 Plasma desorption ionisation

Plasma desorption ionisation mass spectrometry (PDMS) was the first mass spectrometer to be able to analyse proteins and other large biomolecules (although only those of relatively low M_r, less than 35 k). The technique and instruments developed are now obsolete and clearly overtaken by the much more powerful, sensitive and accurate instruments described elsewhere in this chapter. PDMS instruments are however still in use in some laboratories and research publications still appear with mass spectra obtained on this instrument. A basic understanding of the principle is therefore worth including. The source of the plasma (atomic nuclei stripped of electrons) was radioactive californium, ^{252}Cf, and two typical emission nuclei were the 100 MeV Ba^{20+} and Tc^{18+}, formed by the decay of the Cf, which are ejected in opposite directions, almost collinearly and with equal velocity. This is a pulsed technique, i.e. particles are emitted at discrete time intervals and require a TOF mass detector. The plasma particle emitted in the opposite direction to that passing through the sample triggers a time counter and the desorbed sample ions are accelerated electrically and detected as for other TOF analysers (Section 9.3.8).

9.3.8 MALDI, TOF mass spectrometry, MALDI-TOF

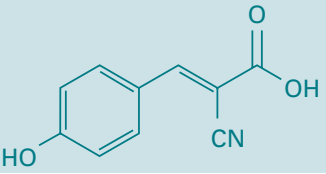
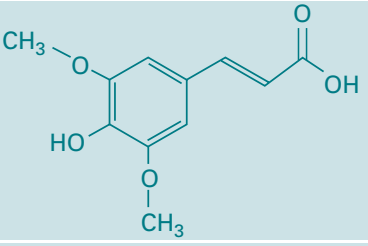
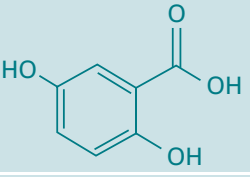
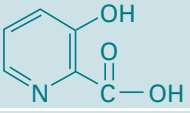
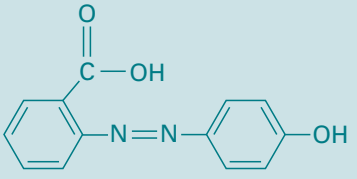
Matrix-assisted laser desorption ionisation (MALDI) produces gas phase protonated ions by excitation of the sample molecules from the energy of a laser transferred via a UV light-absorbing matrix. The matrix is a conjugated organic compound (normally a weak organic acid such as a derivative of cinnamic acid and dihydroxybenzoic acid) that is intimately mixed with the sample. Examples of MALDI matrix compounds and their application for particular biomolecules are shown in Table 9.1. These are designed to maximally absorb light at the wavelength of the laser, typically a nitrogen laser of 337 nm or a neodymium/yttrium-aluminium-garnet (Nd-YAG) at 355 nm.

The sample (1–10 pmol mm^{-3}) is mixed with an excess of the matrix and dried on to the target plate, where they co-crystallise on drying. Pulses of laser light of a few nanoseconds duration cause rapid excitation and vaporisation of the crystalline matrix and the subsequent ejection of matrix and analyte ions into the gas phase (Fig. 9.10). This generates a plume of matrix and analyte ions that are analysed in a TOF mass analyser.

The particular advantage of MALDI is the ability to produce large mass ions, with high sensitivity. MALDI is a very soft ionisation method that does not produce abundant amounts of fragmentation compared with some other ionisation methods. Since the molecular ions are produced with little fragmentation, it is a valuable technique for examining mixtures (see Fig. 9.14 and compare this to the more complex spectrum in Fig. 9.6).

TOF is the best type of mass analyser to couple to MALDI, as this technique has a virtually unlimited mass range. Proteins and other macromolecules of M_r greater than 400 000 have been accurately measured. The principle of TOF is illustrated in Fig. 9.11 and the main components of the instrument are shown in Fig. 9.12.

Table 9.1 **Examples of MALDI matrix compounds**

Compound	Structure	Application
α-Cyano-4-hydroxycinnamic acid (CHCA)		Peptides <10 kDa (glycopeptides)
Sinapinic acid (3,5-dimethoxy-4-hydoxycinnamic acid) (SA)		Proteins >10 kDa
'Super DHB', mixture of 10% 5-methoxysalycilic acid (2-hydroxy-5-methoxybenzoic acid) with DHB		Proteins, glycosylated proteins
2,5-Dihydroxybenzoic acid (DHB) (gentisic acid)		Neutral carbohydrates, synthetic polymers (oligos)
3-Hydroxypicolinic acid		Oligonucleotides
2,-(4-hydroxy-phenlyazo)-Benzoic acid (HABA)		Oligosaccharides, proteins

Sample concentration for MALDI

Maximum sensitivity is achieved in MALDI–TOF if samples are diluted to a particular concentration range. If the sample concentration is unknown a dilution series may be needed to produce a satisfactory sample/matrix spot of suitable concentration on the MALDI plate.

Peptides and proteins seem to give best spectra at around 0.1 to 10 pmol mm^{-3} (Figs. 9.13, 9.14). Some proteins, particularly glycoproteins, may yield better results at concentrations up to 10 pmol mm^{-3}. Oligonucleotides give better spectra at around 10 to 100 pmol mm^{-3} while polymers require a concentration around 100 pmol mm^3. (Note: 1 pmol $nm^{-3} = 10^{-6}$ mol dm^{-3}.)

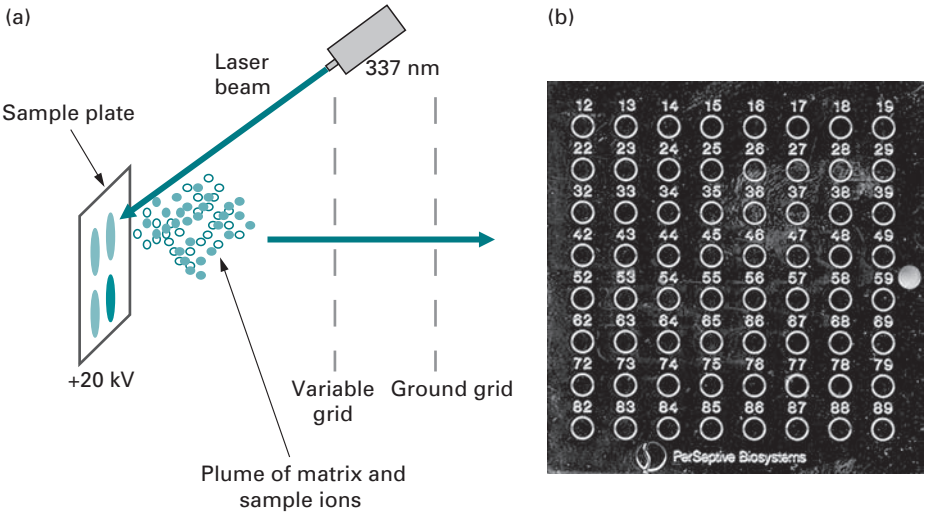


Fig. 9.10 MALDI ionisation mechanism and MALDI–TOF sample plate. (a) The sample is mixed, in solution, with a 'matrix' – the organic acid in excess of the analyte (in a ratio between 1000 : 1 to 10 000 : 1) and transferred to the MALDI plate. An ultraviolet laser is directed to the sample (with a beam diameter of a few micrometres) for desorption. The laser radiation of a few nanoseconds' duration is absorbed by the matrix molecules, causing rapid heating of the region around the area of laser impact and electronic excitation of the matrix. The immediate region of the sample explodes into the high vacuum of the mass spectrometer, creating gas phase protonated molecules of both the acid and the analyte. The laser flash ionises matrix molecules: neutrals (M) and matrix ions $(MH)^+$, $(M-H)^-$ and sample neutral fragments (A). Sample molecules are ionised by gas phase proton transfer from the matrix:

$$MH^+ + A- > M + AH^+.$$
$$(M-H)^- + A- > (A-H)^- + M.$$

The matrix serves as an absorbing medium for the ultraviolet light converting the incident laser energy into molecular electronic energy, both for desorption and ionisation and as a source of H^+ ions to transfer to, and ionise, the analyte molecule. (b) A MALDI sample plate.

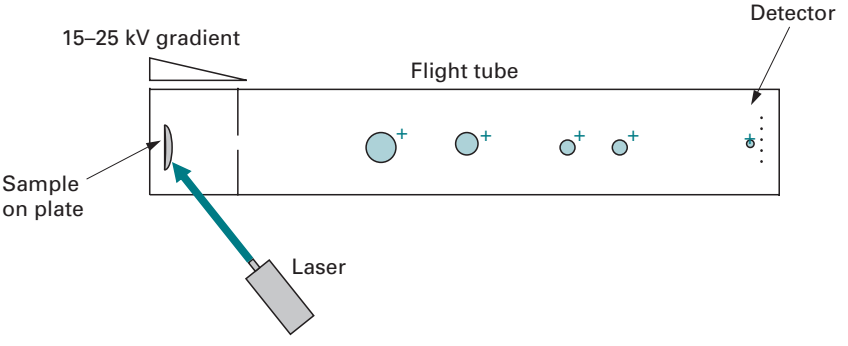


Fig. 9.11 Principle of time-of-flight (TOF). The ions enter the flight tube, where the lighter ions travel faster than the heavier ions to the detector. If the ions are accelerated with the same potential at a fixed point and a fixed initial time, the ions will separate according to their mass to charge ratios. This time of flight can be converted to mass. Typically a few 100 pulses of laser light are used, each of around a few nanoseconds' duration and the information is accumulated to build up a good spectrum. With the benefit of a camera that is used to follow the laser flashes one can move or 'track' the laser beam around the MALDI plate to find so called **sweet spots** where the composition of co-crystallised matrix and sample is optimal for good sensitivity.

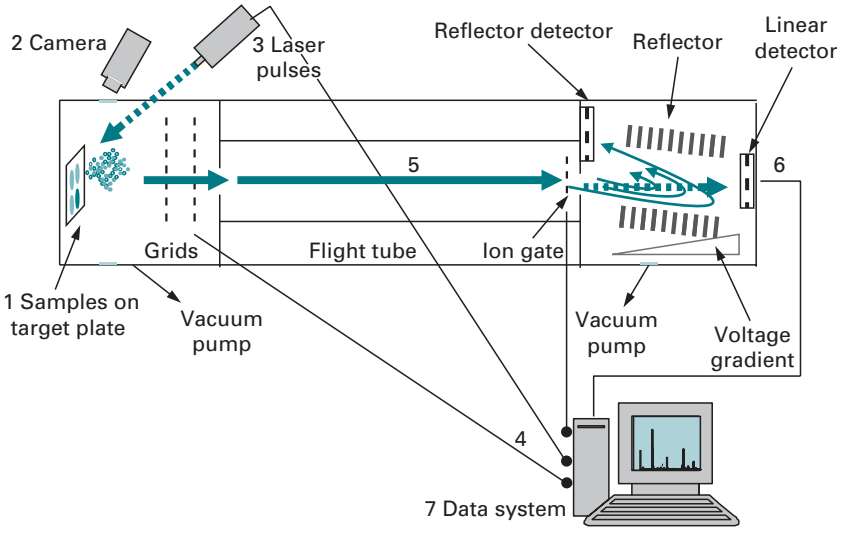


Fig. 9.12 MALDI–TOF instrument components. (1) Sample mixed with matrix is dried on the target plate which is introduced into high-vacuum chamber. (2) The camera allows viewing of the position of the laser beam which can be tracked to optimise the signal. (3) The sample/matrix is irradiated with laser pulses. (4) The clock is started to measure time-of-flight. (5) Ions are accelerated by the electric field to the same kinetic energy and are separated according to mass as they fly through the flight tube. (6) Ions strike the detector either in linear (dashed arrow) or reflectron (full arrows) mode at different times, depending on their *m/z* ratio. (7) A data system controls instrument parameters, acquires signal versus time and processes the data.

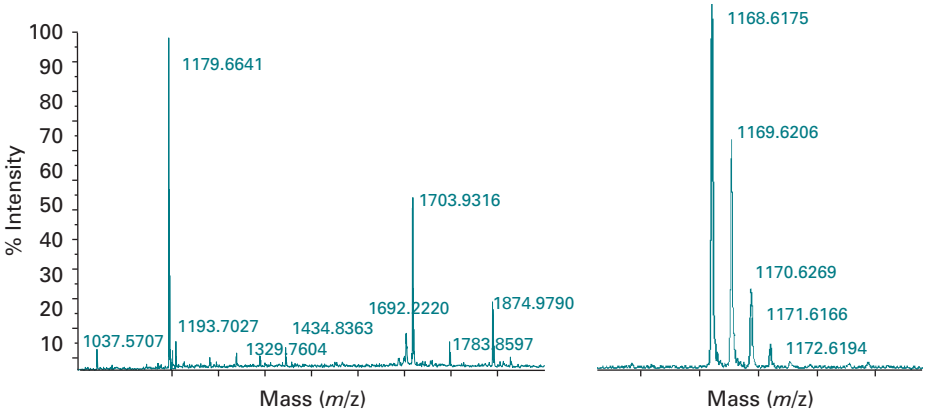


Fig. 9.13 Two examples of MALDI–TOF peptide spectra. The left-hand spectrum is from a protein digest mixture and the right-hand image is an expanded one of a small part of a spectrum showing ^{13}C-containing forms (see Section 9.5.4).

9.3.9 Delayed extraction

In the first MALDI–TOF instruments, the ions in the plume of material generated by the laser pulse were continuously extracted by a high electrostatic field. Since this plume of material occupies a small but finite volume of space, ions arising at different

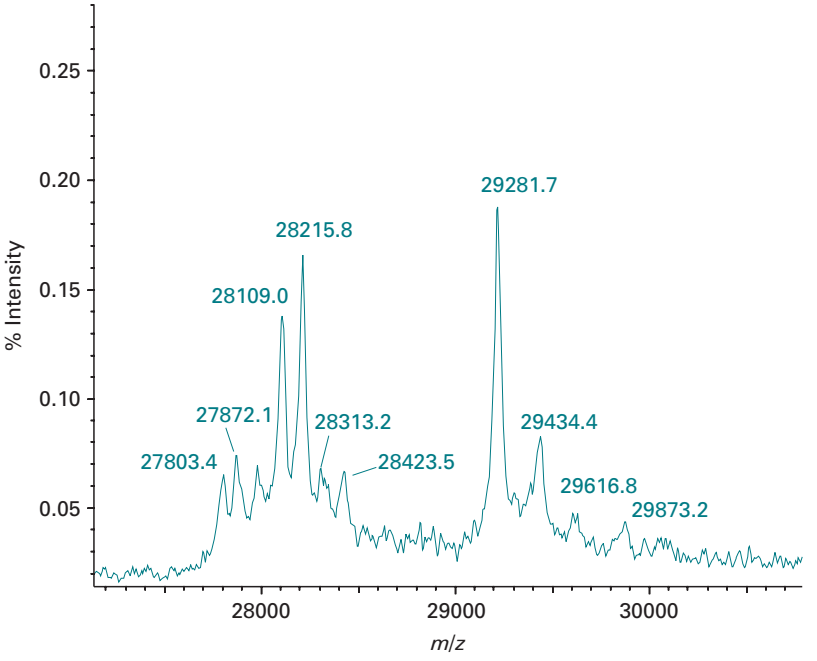


Fig. 9.14 MALDI–TOF spectrum of protein isoforms. The spectrum is almost exclusively singly charged ions representing the molecular ion species of the constituent proteins. Compare this spectrum with the electrospray spectrum of another protein (Fig 9.6) where the multiply charged ions result in multiple peaks which would make it harder to interpret masses of mixtures. (I acknowledge the assistance of Bruker Daltonics who carried out the analysis.)

places could have different energies. This energy spread (and fragmentation occurring during this initial extraction period) usually broadens the peak corresponding to any particular ion which leads to lower mass accuracy. However, if extraction is delayed until all ions have formed, this spread is minimised. The procedure is known as **delayed extraction** (DE), whereby the ions are formed in either a weak field or no field during a predetermined time delay, and then extracted by the application of a high-voltage pulse. The degree of fragmentation of ions (Section 9.3.10) can also be controlled, to some extent, by the length of the time delay. Delayed extraction is illustrated in Fig. 9.15.

9.3.10 Post-source decay

Post-source decay (PSD) is the process of fragmentation that may occur after an ion (the **precursor** ion) has been extracted from the source. Many biological molecules, particularly peptides, give rise to ions that dissociate over a timespan of microseconds and most precursor ions will have been extracted before this dissociation is complete. The fragment ions generated will have the same velocity as the precursor and cause peak broadening and loss of resolution in a linear TOF analyser (Fig. 9.16). The problem is overcome by the use of a reflector.

Example 2 PEPTIDE MASS DETERMINATION (I)

Question A peptide metabolite and an enzyme digest of it were analysed by a combination of mass spectrometric techniques giving the data listed below:

(i) The peptide showed two signals at 3841.5 and 1741 in the MALDI–TOF.

(ii) Five signals could be discerned when the peptide was introduced into a mass spectrometer via an electrospray ionisation source:

m/z	498.2	581.1	697.1	871.2	1161.2

(iii) HPLC-MS of the digest indicated *four* components, the $(M + H)^+$ data for the components being $m/z = 176$, 625, 1229 and 1508. The ions corresponding to the MS of the '625' component appeared at $m/z = 521$, 406, 293, 130 and 113.

(iv) HPLC-MS–MS of the $m/z = 406$ ion of the '625' component identified two ions at $m/z = 378$ and 336, and that of the $m/z = 113$ ion gave $m/z = 85$ and 57, in the product ion spectra.

Use the above data to compare and contrast the different ionisation methods, deduce a molecular mass for the peptide and determine a sequence for the '625' component. Use the amino acid residue mass values in Table 9.2.

Answer The data in (i) are $m/z = 3481.5$ and $m/z = 1741$. These data could represent either of the following possibilities:

(a) $m/z = 3481.5 \equiv (M + H)^+$
when $m/z = 1741 \equiv (M + H)^{2+}$, giving $M = 3480.5$

(b) $m/z = 3481.5 \equiv (2M + H)^+$
when $m/z = 1741 \equiv (M + H)^+$, giving $M = 1740$

Consideration of the data in (ii) allows a choice to be made between these two alternatives, using $n_2 = (m_1 - 1)/(m_2 - m_1)$ and $M = n_2(m_2 - 1)$.

$m_1 - 1$	$m_2 - m_1$	n_2	$m_2 - 1$	M (Da)	z
870.2	290	3.0006	1160.2	3481.2	3
696.1	174.1	3.9982	870.2	3479.3	4
580.1	116	5.0000	696.1	3481.1	5
497.2	82.9	5.9975	580.1	3479.2	6

$\Sigma M = 13920.8$ Da

Mean $M = 3480.2$ Da

The mean M result confirms set (a) of the conclusions above concerning the data obtained from the MALDI experiments.

Example 2 (*cont.*)

The data in (iii) indicate that four products arise from the enzymatic digest of the original peptide. As these products arise directly from the original, the sum of these masses will be related to the *M* of the pepide.

Therefore

$176 + 625 + 1229 + 1508 = 3538\,\text{Da}$

The difference between this mass and the M determined above is

$3538 - 3480.2 = 57.8 \approx 58\,\text{Da}$

The difference of 58 mass units is explained as follows.

Each of the enzyme digest products is protonated (to be 'seen' in the mass spectrometer). Hence this accounts for 4 units. The remaining 54 unit increase arises from the enzymic hydrolysis. From a linear peptide, four products arise from three cleavage points (three cuts in a piece of string give four pieces). Each cleavage point requires the input of one water molecule (hydrolysis, $H_2O = 18$). Three cleavage points require $3 \times 18 = 54$.

The $m/z = 625$, $(M + H)^+$, peak was subjected to further mass spectrometry and sequence ions were observed.

m/z	624	521	406	293	130	113
Δ	103	115	113	163	17	
aa	Cys	Asp	Ile/Leu	Tyr	Ile/Leu	

The loss of 113 from the $m/z = 406$ ion indicates either Ile or Leu. MS2 shows consecutive losses of 28 (CO) and 42 ($CH_2 = CH = CH_3$) which is indicative of Leu. The loss of 17 (not a sequence ion) from 130 confirms this as the C-terminal amino acid.

The predicted sequence from the N-terminal end is
Cys - Asp - Leu - Tyr - Ile

The reflector

A **reflector** (or **reflectron**) is a type of **ion mirror** that provides higher resolution in MALDI–TOF. The reflector increases the overall path length for an ion and it corrects for minor variation in the energy spread of ions of the same mass. Both effects improve resolution. The device has a gradient electric field and the depth to which ions will penetrate this field, before reversal of direction of travel, depends upon their energy. Higher-energy ions will travel further and lower-energy ions a shorter distance. The flight times thus become focussed, while neutral fragments are unaffected by the deflection. Figure 9.16 shows a diagrammatic representation of a MALDI–TOF

Example 3 PEPTIDE MASS DETERMINATION (II)

Question Consider the following mass spectrometric data obtained for a peptide metabolite.

(i) The MALDI spectrum showed two signals at $m/z = 1609$ and 805.

(ii) There were two significant signals in positive ion trap MS mass spectrum at $m/z = 805$ and 827, the latter signal being enhanced on addition of sodium chloride.

(iii) Signals at $m/z = 161.8$, 202.0, 269.0 and 403.0 were observed when the sample was introduced into the mass spectrometer via an electrospray ionisation source.

Use these data to give an account of the ionisation methods used. Discuss the significance of the data and deduce a relative molecular mass for the metabolite.

Use the amino acid residue mass values in Table 9.2.

Answer (i) Signals in the MALDI spectrum were observed at $m/z = 1609$ and 805. These data could represent the following possibilities:

(a) $m/z = 1609 \equiv (M + H)^+$
when $m/z = 805 \equiv (M + 2H)^{2+}$
and $m/z = 403 \equiv (M + 4H)^{4+}$, giving $M = 1608$Da

(b) $m/z = 1609 \equiv (2M + H)^+$
when $m/z = 805 \equiv (M + H)^+$
and $m/z = 403 \equiv (M + 2H)^{2+}$, giving $M = 804$Da

(ii) The distinction between the above options can be made by considering the ion trap data. This mode of ionisation gave peaks at $m/z = 805$ and 827, the latter being enhanced on the addition of sodium chloride. This evidence suggests:
$m/z = 805 \equiv (M + H)^+$
$m/z = 827 \equiv (M + Na)^+$
giving $M = 804$ Da and supports option (b) from the MALDI data.

(iii) The multiply charged ions observed in the electrospray ionisation method allow an average M to be calculated. Using the standard formula:

$m_1 - 1$	$m_2 - m_1$	n_2	$m_2 - 1$	M (Da)	z
268.0	134	2.0	402.0	804	2
201.0	67	3.0	268.0	804	3
160.8	40.2	4.0	201.0	804	4

The molecular mass is clearly 804 Da, confirming the above conclusions.

instrument that includes the facility for both linear and reflectron modes of ion collection. The reflectron improves resolution and mass accuracy and also allows structure and sequence information (in the case of peptides) to be obtained by PSD analysis.

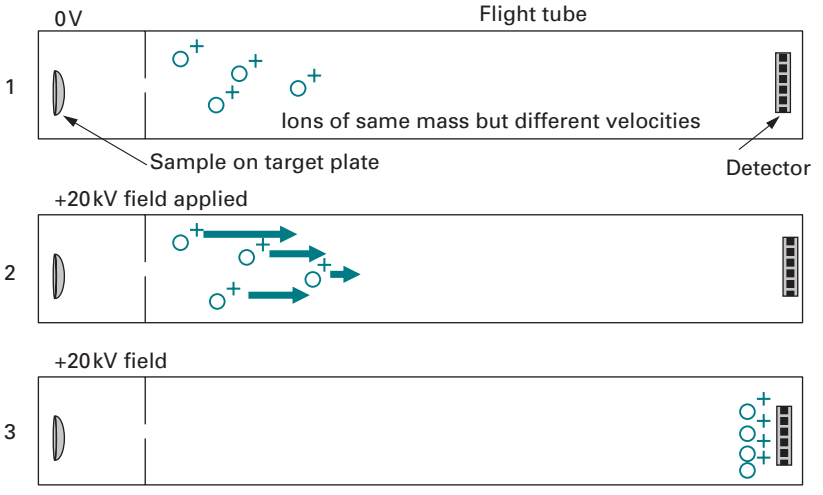


Fig. 9.15 Delayed extraction (DE). (1) No applied electric field. The ions spread out. (2) Field applied. The potential gradient accelerates slow ions more than fast ones. (3) Slow ions catch up with faster ones at the detector.

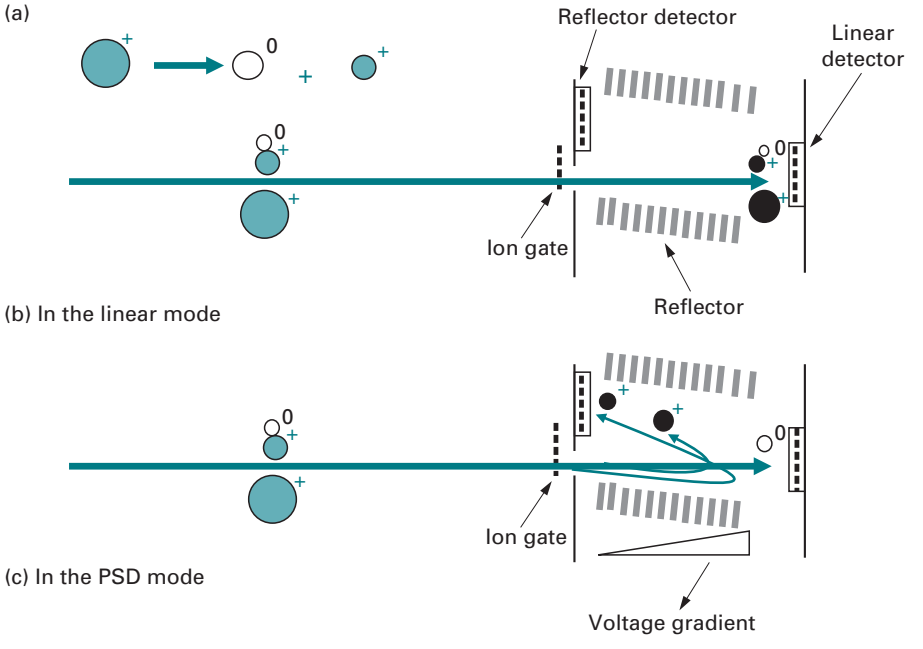


Fig. 9.16 The MALDI–TOF reflector. Post-source decay (PSD) theory. (a) Fragment ions arising by PSD as well as the neutral fragments and the precursor ions have the same velocity and reach the detector simultaneously. This prevents a distinction between precursor and PSD fragment. (b) In the linear mode the charged fragments are not separated. (c) In the reflector mode, the fragment that does not retain the charge (neutral, denoted by 0°) is not deflected in the reflector but the charged fragments ($\bullet^{+}$) are deflected according to their *m/z* and a spectrum of the fragment (daughter) ions is recorded, albeit of a limited *m/z* range for each setting of the reflector voltages.

Sequencing peptides by PSD analysis in MALDI–TOF is less straightforward (and in a large percentage of experiments is unsuccessful) than tandem MS on a quadrupole ESI or ion trap instrument. At any given setting of the reflector/ion mirror, charged fragments of a particular range of m/z are focussed in the reflector (Fig. 9.16). Fragment ions of m/z above and below this narrow range are poorly focussed. Therefore, since only fragment ions of a limited mass range are focussed for a given **mirror ratio** in the reflector, a number of spectra are run at different settings and **stitched** together to generate a composite spectrum.

Types of MALDI sample plates

MALDI sample plate types that are available include 100-well stainless steel flat plates. These are good for multiple sample analysis where close **external calibration** is used, that is the use of a compound or compounds of known molecular mass placed on an adjacent spot to calibrate the instrument. It is also easier to see crystallisation of the matrix on this type of surface.

Four-hundred-spot Teflon-coated plates have particular application for concentrating sample for increased sensitivity. Due to the very small diameter of the spots, it is difficult to spot accurately manually but these plates are good for automated sample spotting. Only in the centre of each spot is the surface of the plate exposed therefore the sample does not 'wet' over the whole surface but concentrates itself into the centre of each spot as it dries. Gold-coated plates with wells (2 mm diameter, see Fig. 9.10b) are good surfaces on which to contain the spread of sample and matrix when used with highly organic solvents, e.g. tetrahydrofuran (THF) preparations for polymers. They also allow on-plate reactions within the well with thiol-containing reagents that bind to the gold surface.

9.3.11 Novel hybrid instruments

There are a number of commercial developments of hybrid MS instruments that involve coupling an electrospray, ion trap or a MALDI ion source with a hybrid quadrupole orthogonal acceleration time-of-flight mass spectrometer (Fig. 9.17). This potentially leads to improved tandem MS performance from MALDI phase samples.

The intention of the development of these instruments is to combine the best features of both types of ion source with the best features of all types of analyser in order to improve tandem MS capability and increase sensitivity. Hybrid magnetic sector instruments are also manufactured where the first mass spectrometer is a two-sector device and the second mass spectrometer is a quadrupole.

9.3.12 Fourier-transform ion cyclotron resonance MS

The recent development of **Fourier-transform ion cyclotron resonance** (FT-ICR) mass spectrometry has great potential in analysis of a wide range of biomolecules. It is potentially the most sensitive mass spectrometric technique and has very high

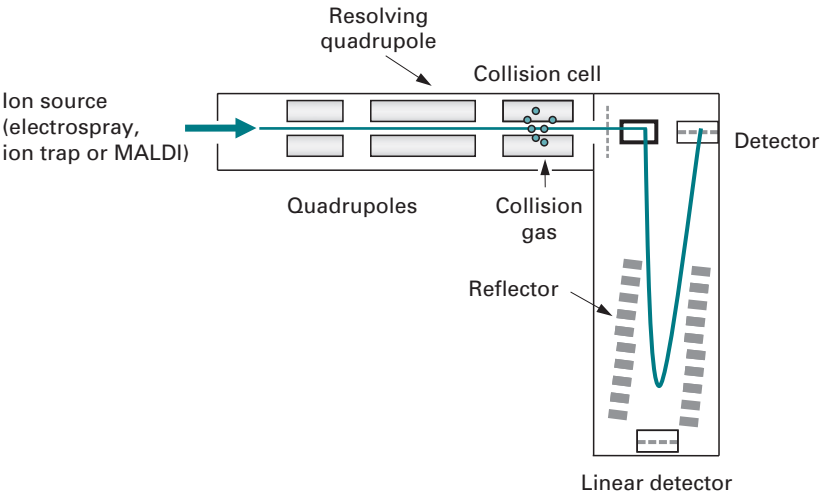


Fig. 9.17 Diagram of a hybrid quadrupole TOF MS. The diagram shown here does not represent any specific instrument from a particular manufacturer. The source may be an ion trap device, an electrospray or even a MALDI source (such as in the 'MALDI Q-TOF' from Micromass). Other hybrid instruments include the Bruker Daltonics 'BioTOF III, ESI-Q-q-TOF System' and the 'QSTAR' Hybrid LC/MS/MS from Applied Biosystems with an electrospray, nanospray or an optional MALDI source. The Shimadzu Biotech 'AXIMA MALDI QIT TOF' combines a MALDI source with an ion trap and reflectron TOF mass analyser.

mass resolution; $>10^6$ is observable with most instruments. The instrument also allows tandem MS to be carried out. The ions can be generated by a variety of techniques, such as an ESI or a MALDI source. FT-ICR MS is based on the principle of ions, which while orbiting in a magnetic field, are excited by radio frequency (RF) signals. As a result, the ions produce a detectable **image current** on the **cell** in which they are trapped. The time-dependent image current is Fourier transformed to obtain the component frequencies of the different ions, which correspond to their m/z (Fig. 9.18).

9.3.13 Orbitrap mass spectrometer

The resolving power of FTICR-MS is proportional to the strength of the magnetic field therefore superconducting magnets (3–12 tesla) are required, which makes for high maintenance cost. The Orbitrap mass analyser is a lower-cost alternative to the high magnetic field FTICR mass spectrometer. Ions are trapped in the **orbitrap**, where they undergo harmonic ion oscillations, along the axis of an electric field (see Fig. 9.19). Their m/z values are measured from the frequency of the ions, measured non-destructively, using Fourier transforms to obtain the mass spectrum. The instrument has high mass resolution (up to 150 000), high mass accuracy (2–5 p.p.m.), an m/z range of >6000 and a dynamic range greater than 10^3 with sub-femtomol sensitivity.

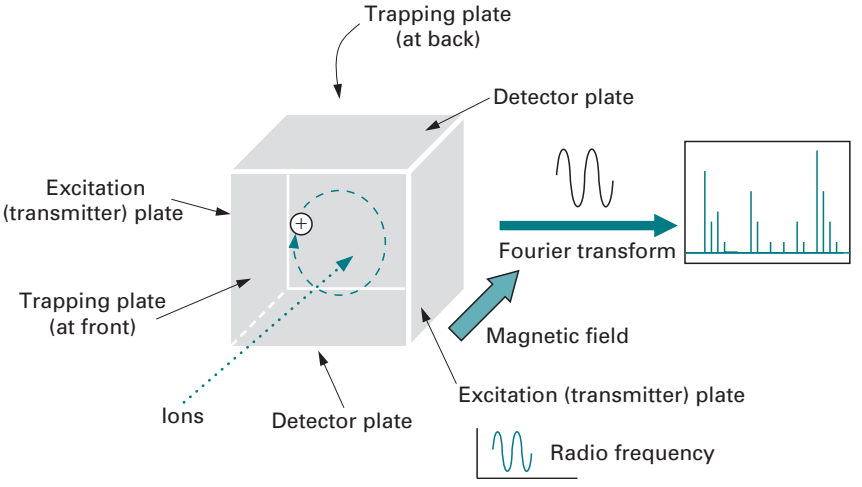


Fig. 9.18 Schematic diagram of the Fourier-transform ion cyclotron resonance (FT-ICR) instrument. The technique involves trapping, excitation and detection of ions to produce a mass spectrum. The **trapping plates** to maintain the ions in orbit are at the front and back in the schematic. The **excitation** or **transmitter plates** where the radio frequency (RF) pulse is given to the ions are shown at each side and the **detector plates** that detect the **image current** which is Fourier transformed are shown at the top and bottom. The sample source is normally electrospray (described in Section 9.2.4) or MALDI (see Section 9.3.8 and Fig. 9.10). The ions are focussed and transferred into the **analyser cell** under high vacuum. The analyser cell is a type of ion trap in a spatially uniform strong magnetic field which constrains the ions in a circular orbit, the frequency of which is determined by the mass, charge and velocity of the ion. While the ions are in these stable orbits between the detector electrodes they will not give a measurable signal. In order to achieve this, ions of a given *m/z* are excited to a wider orbit by applying a RF signal of a few milliseconds' duration. One frequency excites ions of one particular *m/z* which results in the ions producing a detectable **image current**. This time-dependent image current is Fourier transformed to obtain the component frequencies which correspond to the *m/z* of the different ions. The angular frequency measurements produce values for *m/z*. Therefore the mass spectrum is determined to a very high mass resolution since frequency can be measured more accurately than any other physical property. After excitation, the ions relax back to their previous orbits and high sensitivity can be achieved by repeating this process many times.

9.4 DETECTORS

9.4.1 Introduction

The ions from the mass analyser impinge on a surface of a detector where the charge is neutralised, either by collection or donation of electrons. An electric current flows that is amplified and ultimately converted into a signal that is processed by a computer. The **total ion current** (TIC) is the sum of the current carried by all the ions being detected at any given moment and is a very useful parameter to measure during **on-line** MS. A plot of ion current versus time complements the ultraviolet trace that is also normally recorded during the chromatography run. Unlike the ultraviolet trace which depends on the absorbance of each component at the particular wavelength(s) set on the ultraviolet detector, the TIC is of course independent of the light-absorbing properties of a substance and depends only on its ionisability in the instrument.

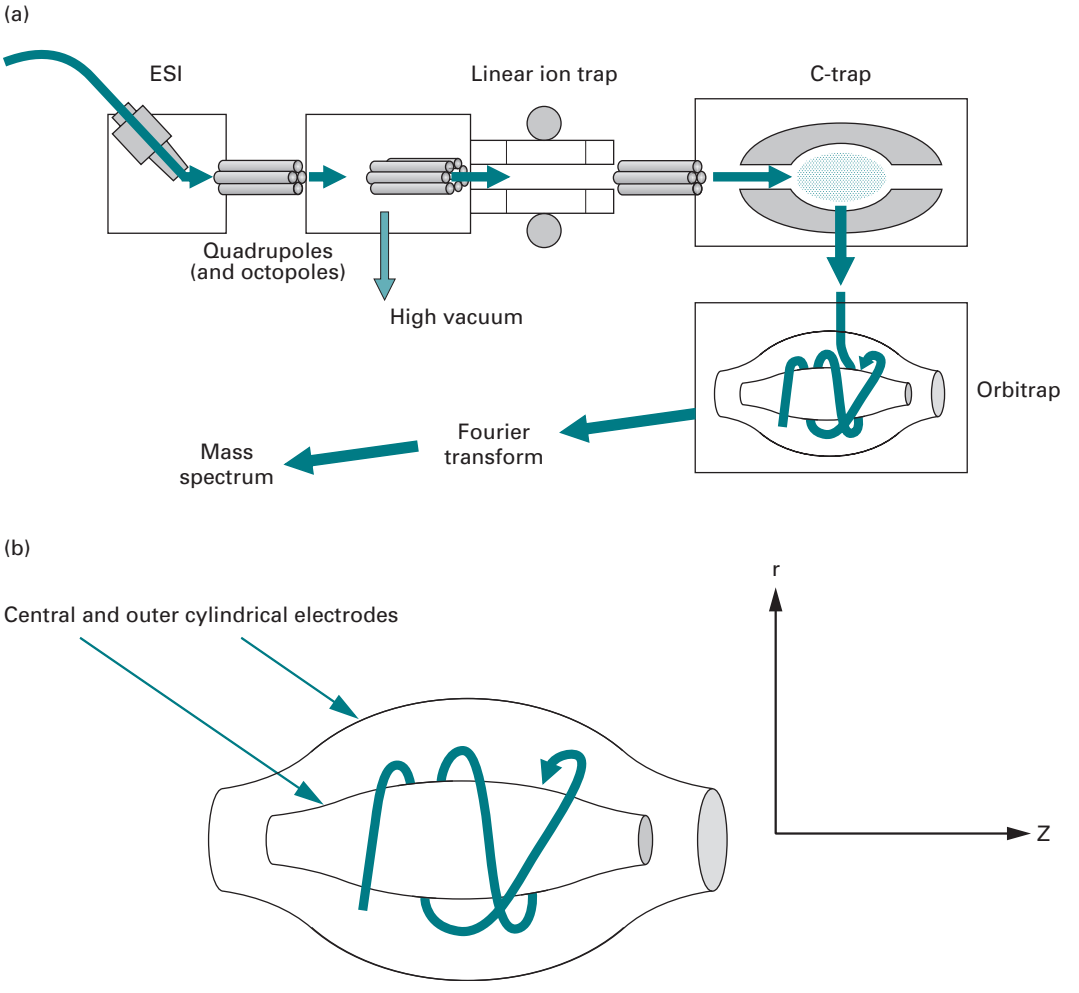


Fig. 9.19 (a) Simplified schematic of an Orbitrap mass spectrometer with an electrospray ionisation (ESI) source. Ions are transferred from an ESI source (described in Section 9.2.4) through three stages of differential vacuum pumping using RF guide quadrupoles and octopoles which focus and guide the ions through the various parts of the instrument. The ions are stored in the linear ion trap then axially ejected to the C-trap where they are squeezed into a small cloud. 'Bunches' of ions are then injected into the Orbitrap analyser. The third quadrupole, which is pressurised to less than 10^{-3} torr with collision gas, acts as an ion accumulator where ion/neutral collisions slow the ions which pool in an axial potential well at the end of the quadrupole (the C-trap). The linear ion trap operates on a similar principle to the ion trap described in Section 9.3.3. and Fig. 9.7. MALDI (see Section 9.3.8. and Fig. 9.10) is an alternative ion source. (b) Detail of the Orbitrap analyser. In the Orbitrap, the ions are trapped in a radial electric field between a central and an outer cylindrical electrode. They orbit around the central electrode with axial oscillations. The superimposed harmonic oscillations in the *Z* direction are detected by measuring the image current at the outer electrode. The frequency of oscillation is proportional to the *m/z* ratio and is detected and processed by fast Fourier transform, as in FT-ICR MS.

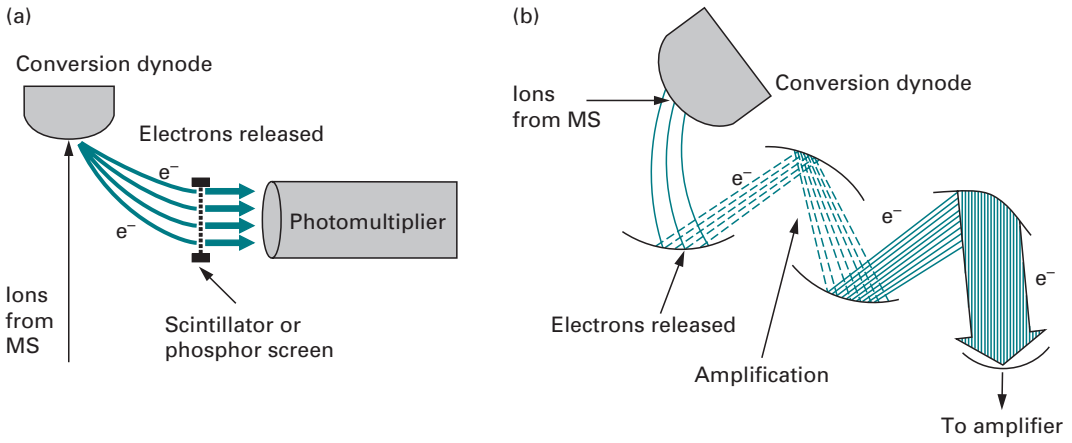


Fig. 9.20 Conversion dynode and electron multiplier. (a and b) Each ion strikes the conversion dynode (which converts ions to electrons) which emits a number of electrons that travel to the next, higher-voltage dynode. The secondary electrons from the conversion dynode are accelerated and focussed onto a second dynode, which itself emits secondary electrons. Each electron then produces several more electrons. Amplification is achieved through the 'cascading effect' of secondary electrons from dynode to dynode that finally results in a measurable current at the end of the electron multiplier. The cascade of electrons continues until a sufficiently large current for normal amplification is obtained. A series of up to 10–20 dynodes (set at different potentials) provides an amplification gain of 10^6 or 10^7.

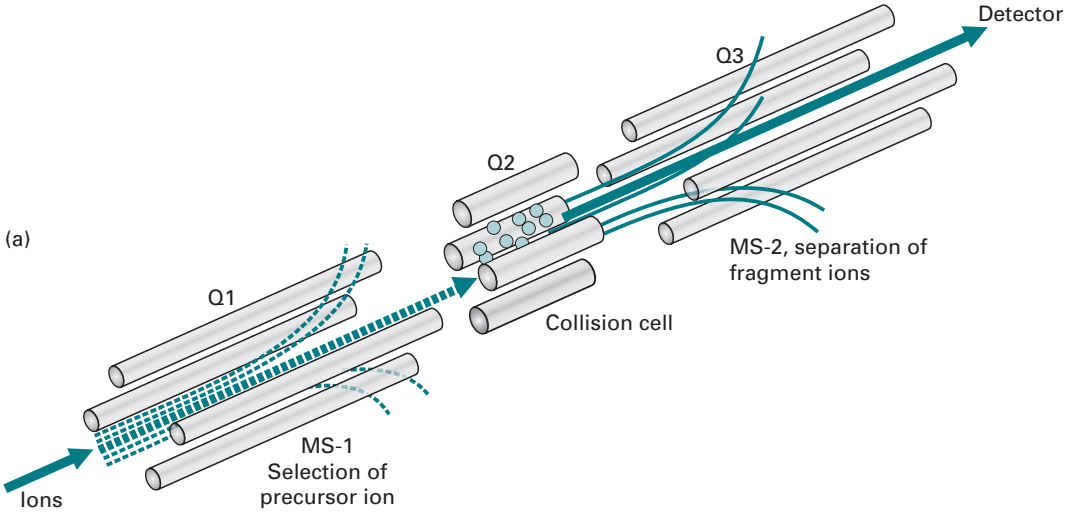
9.4.2 Electron multiplier and conversion dynode

Electron multipliers are used as detectors for many types of mass spectrometers. These are frequently combined with a conversion dynode which is a device to increase sensitivity. The ion beam from the mass analyser is focussed onto the conversion dynode, which emits electrons in direct proportion to the number of bombarding ions. A positive ion or a negative ion hits the conversion dynode, causing the emission of secondary particles containing secondary ions, electrons and neutral particles (see Fig. 9.20). These secondary particles are accelerated into the dynodes of the electron multiplier. They strike the dynodes with sufficient energy to dislodge electrons, which pass further into the electron multiplier, colliding with the dynodes, producing more and more electrons.

9.5 STRUCTURAL INFORMATION BY TANDEM MASS SPECTROMETRY

9.5.1 Introduction

As mentioned above, the newer ionisation techniques ESI and MALDI are soft ionisation techniques (as is FAB and its derivative techniques). In contrast to EI, they do not produce significant amounts of fragment ions. Therefore in order to obtain structural information on biomolecules and sequence information (in the case of proteins and peptides), tandem MS has been developed. The technique can also be applied to obtain sequence information on oligosaccharides (see Sections 9.5.5 and 9.5.6) and oligonucleotides. Although it is unlikely that this method will ever replace DNA sequencing



(b)

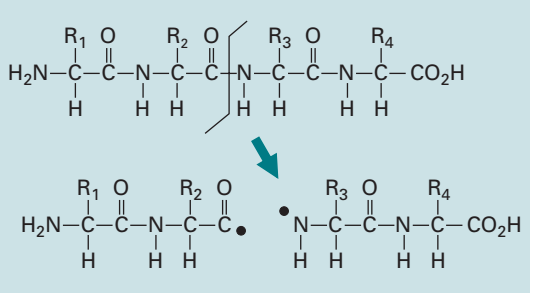


Fig. 9.21 Quadrupole MS sequencing. An ion of a particular *m/z* value is selected in the first quadrupole, Q1, as in Fig. 9.5, but instead of being detected, it passes through the second quadrupole, Q2, where it is subjected to collision with the collision gas. The Q3 acts like a second quadrupole mass spectrometer, MS-2, to scan *m/z* to obtain a spectrum of the fragment ions. The collision cell, Q2, is frequently a radio frequency (RF)-only quadrupole containing the appropriate collision gas. No mass filtering occurs here, the RF merely constrains the ions to allow a greater number of collisions to occur. The fragmentation depicted here is at the peptide bond and one of the fragments will retain the charge, resulting in either a y-series or a b-series ion (see Fig. 9.22).

gels, it can be used to identify positions of modified or labelled bases that might not be picked up by the Sanger dideoxy sequencing method.

Structural information can be obtained on almost any type of organic molecule, on an instrument that is suitable for that type of sample. This includes investigation of organic compounds on a magnet sector MS where two double-focussing magnetic sector machines can be combined into a four-sector device coupled through a **collision cell.**

The general procedure is that a mixture of ions is generated in the ion source of the mass spectrometer as normal and the ions are allowed to pass through the first mass

analyser where an ion of a particular m/z is selected (but not detected). This ion then enters the **collision cell** and collides with an inert **collision gas** such as helium or argon. The kinetic energy of this ion is converted to vibrational energy and the ion fragments. This is known as collision-induced dissociation (CID) or **collision-activated dissociation** (CAD). The m/z values of the fragment ions are then determined in a second mass spectrometer (see Fig. 9.21 for an illustration of the principle in a quadrupole mass spectrometer). Collision cells may be placed in any of the field-free regions, leading to a wide variety of experimental methodologies for many different applications.

For example, as well as in the triple quadrupole MS this can be done in a hybrid instrument such as the Q-TOF (described in Section 9.3.11).

Since the principles of tandem MS are similar for most instrument configurations, further discussion will focus on electrospray tandem MS.

The procedure for obtaining structural and sequence information on polypeptides in ion trap MS has been described above (Section 9.3.3).

9.5.2 Sequencing of proteins and peptides

The identification of proteins involves protease cleavage, mostly by trypsin. Owing to the specificity of this protease, tryptic peptides usually have basic groups at the N- and C-terminis. Trypsin cleaves after lysine and arginine residues, both of which have basic side chains (an amino and a guanidino group respectively). This results in a large proportion of high-energy doubly charged positive ions that are more easily fragmented.

The digestion of the protein into peptides is followed by identification of the peptides by mass charge ratio (m/z) either as very accurate masses alone or by using a second fragmentation that gives ladders of fragments cleaved at the peptide bonds.

Although a wide variety of fragmentations may occur, there is a predominance of peptide bond cleavage which gives rise to peaks in the spectrum that differ sequentially by the residue mass. The mass differences are thus used to reconstruct the amino acid sequence (**primary structure**) of the peptide (Table 9.2).

Different series of ions, a, b, c and x, y, z, may be recognised, depending on which fragment carries the charge. Ions x, y and z arise by retention of charge on the C-terminal fragment of the peptide. For example, the z_1 ion is the first C-terminal residue; y_1 also contains the NH group (15 atomic mass units greater) and x_1 includes the carbonyl group; y_2 comprises the first two C-terminal residues, and so on. The a, b, and c ion series arise from the N-terminal end of the peptide, when the fragmentation results in retention of charge on these fragments.

Figure 9.22a shows an idealised peptide subjected to fragmentation. Particular series will generally predominate so that the peptide may be sequenced from both ends by obtaining complementary data (Fig. 9.22b). In addition, ions can arise from side chain fragmentation, which enables a distinction to be made between isomeric amino acids such as leucine and isoleucine.

The protein is identified by searching databases of expected masses from all known peptides from every protein (or translations from DNA) and theoretical masses from fragmented peptides. Sensitivity of tandem MS has been claimed down to zeptomole level.

Table 9.2 **Symbols and residue masses of the protein amino acids**

Name	Symbol	Residue mass[a]	Side chain
Alanine	A, Ala	71.079	CH_3-
Arginine	R, Arg	156.188	$HN{=}C(NH_2){-}NH{-}(CH_2)_3$-
Asparagine	N, Asn	114.104	$H_2N{-}CO{-}CH_2$-
Aspartic acid	D, Asp	115.089	$HOOC{-}CH_2$-
Cysteine	C, Cys	103.145	$HS{-}CH_2$-
Glutamine	Q, Gln	**128.131**	$H_2N{-}CO{-}(CH_2)_2$-
Glutamic acid	E, Glu	129.116	$HOOC{-}(CH_2)_2$-
Glycine	G, Gly	57.052	H-
Histidine	H, His	137.141	Imidazole-CH_2-
Isoleucine	I, Ile	**113.160**	$CH_3{-}CH_2{-}CH(CH_3)$-
Leucine	L, Leu	**113.160**	$(CH_3)_2{-}CH{-}CH_2$-
Lysine	K, Lys	**128.17**	$H_2N{-}(CH_2)_4$-
Methionine	M, Met	131.199	$CH_3{-}S{-}(CH_2)_2$-
Metsulphoxide	Met.SO	**147.199**	$CH_3{-}S(O){-}(CD_2)_2$-
Phenylalanine	F, Phe	**147.177**	Phenyl-CH_2-
Proline	P, Pro	97.117	Pyrrolidone-CH-
Serine	S, Ser	87.078	$HO{-}CH_2$-
Threonine	T, Thr	101.105	$CH_3{-}CH(OH)$-
Tryptophan	W, Trp	186.213	Indole-$NH{-}CH{=}C{-}CH_2$-
Tyrosine	Y, Tyr	163.176	4-OH-Phenyl-CH_2-
Valine	V, Val	99.133	$CH_3{-}CH(CH_2)$-

Note: [a]Residue mass is the mass in a peptide bond, i.e. after loss of H_2O when the peptide bond is formed. The numbers in bold in the residue mass column indicate amino acids that may be ambiguous in a sequence determined by tandem MS due to close similarity or identity in mass.

9.5.3 Comparison of MS and Edman sequencing

Edman degradation (Section 8.4.3) to obtain the complete sequence of a protein is uncommon nowadays since genomes are available to search with fragmentary sequences. Most intact proteins, if they are not processed from a secretory or pro-peptide form, are blocked at the N-terminus, most commonly with an acetyl group. Other amino terminal blocking includes fatty acylation, most commonly with a myristoyl, C_{12} fatty acid, attached through a glycine residue but the presence of many shorter-chain fatty acids is known to occur. Cyclisation of glutamine to a pyroglutamyl

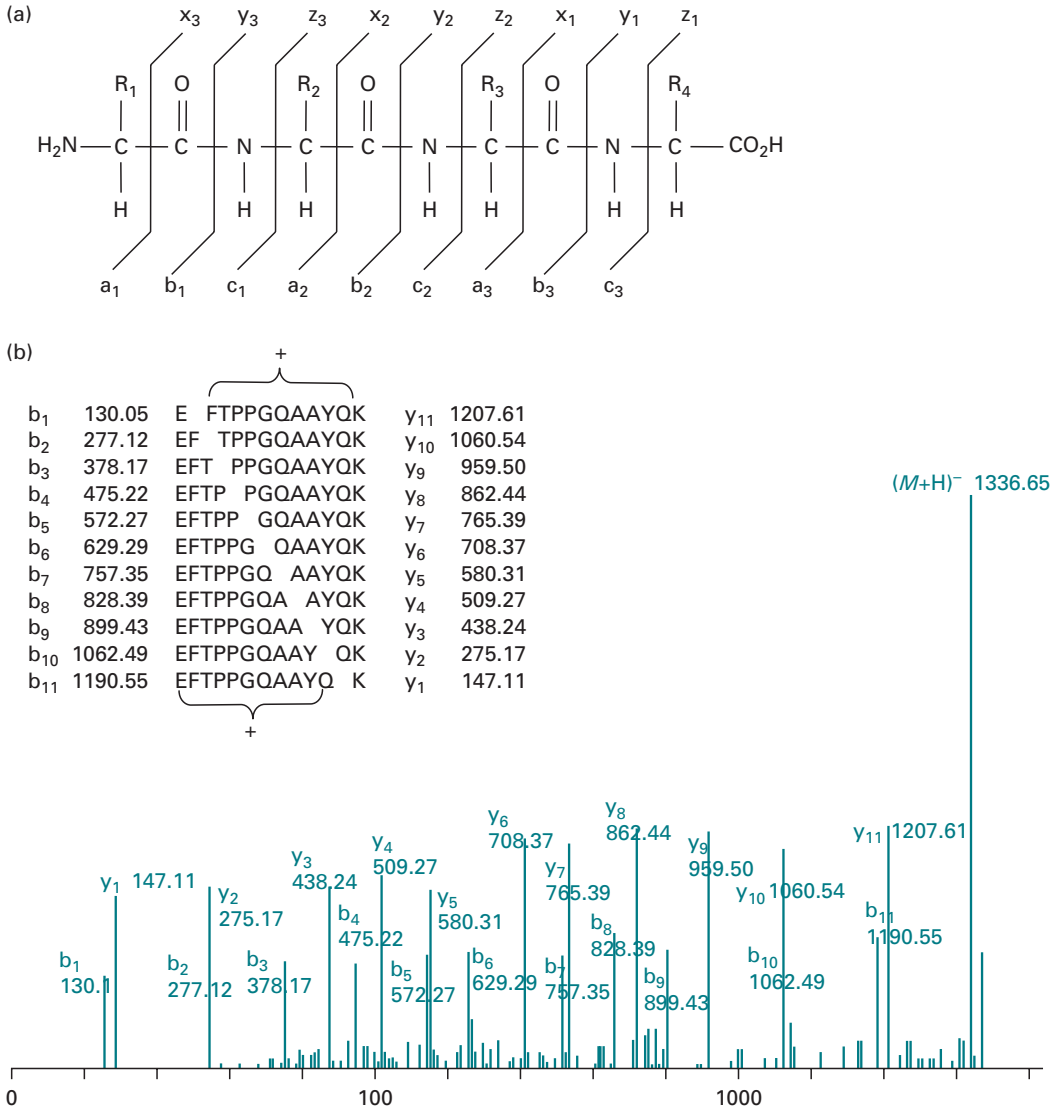


Fig. 9.22 Peptide fragment ion nomenclature and tandem MS spectrum of a peptide. (a) Charge may be retained by either the N- or C-terminal fragment, resulting in the a, b and c series of ions or x, y and z series respectively. Ions in the b and y series frequently predominate. Corresponding neutral fragments are of course not detected. (b) The sequence of the peptide from a mutant haemoglobin is: EFTPPGQAAYQK. The figure shows the tandem mass spectrum from collision-induced dissociation of the doubly charged $(M + 2H)^{2+}$ precursor, $m/z = 668.3$. Cleavage at each peptide bond results in the b or y ions when the positive charge is retained by the fragment containing the N- or C-terminus of the peptide respectively (see inset).

residue and post-translational modification to N-terminal trimethylalanine and dimethylproline also occur. In the case of recombinant proteins over-expressed in *E. coli*, the initiator residue *N*-formyl methionine is often incompletely removed. All these modifications leave the N-terminal residue without a free proton on the alpha nitrogen and Edman chemistry cannot proceed. Mass spectrometry has therefore been essential for their correct structural identification. The protein sequencing

instruments are still important for solid phase sequencing to identify post-translational modifications; in particular, sites of phosphorylation and a combination of micro-sequencing and mass spectrometry techniques are now commonly employed for complete covalent structure determination of proteins.

Example 4 **PEPTIDE SEQUENCING (I)**

Question An oligopeptide obtained by tryptic digestion was investigated by ESI–MS and ion trap MS–MS both in positive mode, and gave the following *m/z* data:

ESI	223.2	297.3								
Ion trap	146	203	260	357	444	591	648	705	802	890

(i) Predict the sequence of the oligopeptide. Use the amino acid residual mass values in Table 9.2.
(ii) Determine the average molecular mass.
(iii) Identify the peaks in the ESI spectrum.

Note: Trypsin cleaves on the C-terminal side of arginine and lysine.

Answer (i) The highest mass peak in the ion trap MS spectrum is $m/z = 890$, which represents $(M + H)^+$.

Hence $M = 889$ Da.

m/z	146	203	260	357	444	591	648	705	802	889
Δ		57	57	97	87	147	57	57	97	87
aa		Gly	Gly	Pro	Ser	Phe	Gly	Gly	Pro	Ser

The mass differences (Δ), between sequence ions, represent the amino acid (aa) residue masses. The lowest mass sequence ion, $m/z = 146$, is too low for arginine and must therefore represent Lys + OH. The sequence in conventional order from the N-terminal end would be:

Ser-Pro-Gly-Gly-Phe-Ser-Pro-Gly-Gly-Lys

(ii) The summation of the residues = 889 Da, which is a check on the mass spectrometry value for *M*.
(iii) The *m/z* values in the ESI spectrum represent multiply charged species and may be identified as follows:
$m/z = 223.2 \equiv (M + 4H)^{4+}$ from $889/223.2 = 3.98$
$m/z = 297.3 \equiv (M + 3H)^{3+}$ from $889/297.3 = 2.99$

Remember that *z* must be an integer and hence values need to be rounded to the nearest whole number.

Table 9.3 **Mass differences due to isotopes in multiply charged peptides**

Charge on peptide	Apparent mass	Mass difference between isotope peaks
Single charge	$[(M + H)/1]$	1 Da
Double charge	$[(M + 2H)/2]$	0.5 Da
Triple charge	$[(M + 3H)/3]$	0.33 Da
n charges	$[(M + nH)/n]$	$1/n$ Da

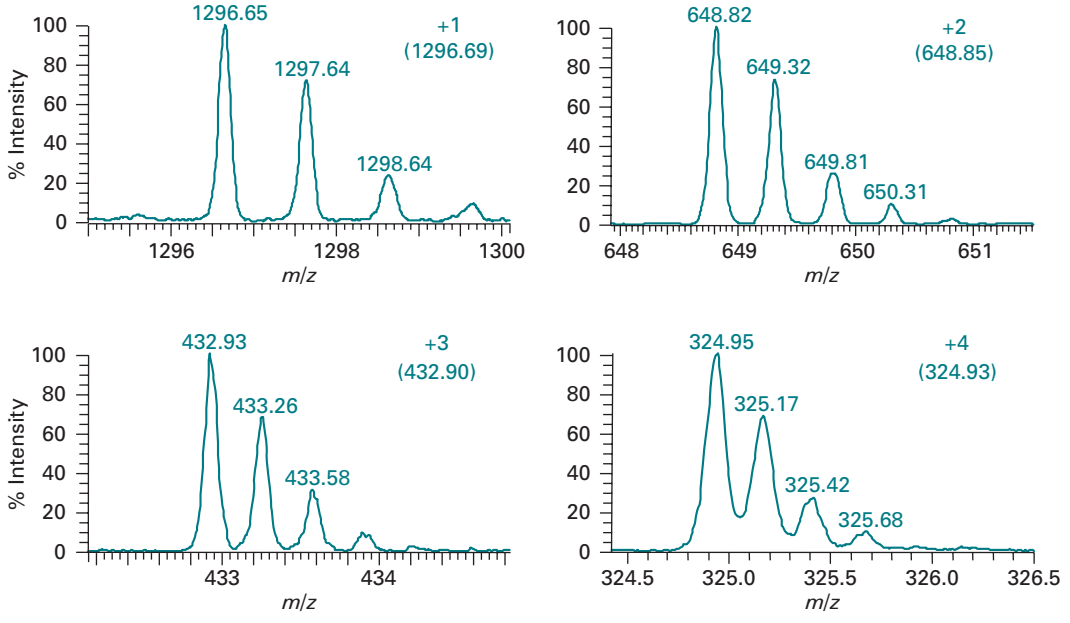


Fig. 9.23 Spectra of a multiply charged peptide. Finding the charge state of a peptide involves **zooming in** on a particular part of the mass spectrum to obtain a detailed image of the mass differences between different peaks that arise from the same biomolecule, due to isotopic abundance. This is mainly due to ^{12}C and its ^{13}C isotope, as described in the text.

9.5.4 Carbon isotopes and finding the charge state of a peptide

Since the mass detector operates on the basis of mass-to-charge ratio (m/z), mass assignment is normally made assuming a single charge per ion (i.e. $m/z = m + 1$ in positive ion mode). However, since there is around 1.1% ^{13}C natural abundance, with increasing size, peptides will have a greater chance of containing at least one ^{13}C and two ^{13}C, etc. A peptide of 20 residues has approximately equal peak heights of the 'all ^{12}C peptide' and of the peptide with one ^{13}C.

A singly charged peptide will show adjacent peaks differing in one mass unit; a doubly charged peptide will show adjacent peaks differing in half a mass unit and so on (Fig. 9.23 and Table 9.3). In the example illustrated, the peptide has a mass

calculated from its sequence as 1295.69. The experimentally derived values are, for the singly charged ion, $[(M + H)/1] = 1296.65$ and for the doubly charged ion, $[(M + 2H)/2] = 648.82$.

For elements such as chlorine, the isotopic abundance is approximately 3 : 1, ^{35}Cl : ^{37}Cl. If a compound contains a single chlorine atom, two ion species will be observed, with peak intensities in an approximate ratio of 3 : 1. If a compound contains two chlorine atoms then three peaks will be seen.

The technique is particularly useful for determining which are the high-energy doubly charged tryptic peptides, for tandem MS.

Example 5 PEPTIDE SEQUENCING (II)

Question Determine the primary structure of the oligopeptide that gave the following, positive mode, MS–MS data:

m/z	149	305	442	529	617

Use the amino acid residual mass values in Table 9.2.

Answer $m/z = 617 \equiv (M + H)^+$

m/z	149	305	442	529	616
Δ	156	137	87	87	
aa	Arg	His	Ser	Ser	

Conventional order for the sequence would be:

Ser-Ser-His-Arg-?

It is important to note that no assignment has been given for the remaining $m/z = 149$. It may not in fact be a sequence ion and more information would be required, such as an accurate molecular mass of the oligopeptide, in order to proceed further. It is, however, possible to speculate as to the nature of this ion. If the $m/z = 149$ ion is the C-terminal amino acid then it would end in -OH and be 17 mass units greater than the corresponding residue mass. The difference between 149 and 17 is 132, which is extremely close to methionine, so this amino acid remains a possibility to end the chain.

9.5.5 Post-translational modification of proteins

Many chemically distinct types of post-translational modification of proteins are known to occur. These include the wide variety of acylations at the N-terminus of proteins (mentioned above) as well as acylations at the C-terminus and at internal sites. In this section, examples of the application of MS techniques employed for analysis of glycosylation, phosphorylation and disulphide bonds are given.

An up-to-date list of the broad chemical diversity of known modifications and the side chains of the amino acids to which they are attached is on the website

Example 6 PEPTIDE MASS DETERMINATION (III)

Question An unknown peptide and an enzymatic digest of it were analysed by mass spectrometric and chromatographic methods as follows:

(i) MALDI–TOF mass spectrometry of the peptide gave two signals at $m/z = 3569$ and 1785;
(ii) MALDI–TOF of the hydrolysate showed signals at $m/z = 766$, 891, 953 and 1016;
(iii) the data obtained from analysis of the peptide using coupled HPLC–MS operating through an electrospray ionisation source were $m/z = 510.7$, 595.7, 714.6, 893.0 and 1190.3;
(iv) when the hydrolysate was analysed by HPLC, four distinct components could be discerned.

Explain what information is available from these observations and determine a molecular mass, using the amino acid residue mass values in Table 9.2, for the unknown peptide.

Answer (i) Signals from MALDI–TOF were observed at $m/z = 3569$ and 1785. These data could represent either of the following possibilities:
(a) $m/z = 3569 \equiv (M + H)^+$
when $m/z = 1785 \equiv (M + 2H)^{2+}$, giving $M = 3568$
(b) $m/z = 3569 \equiv (2M + H)^+$,
when $m/z = 1785 \equiv (M + H)^+$, giving $M = 1784$

(ii) It is possible to distinguish between these two options by considering the MALDI–TOF of the products of hydrolysis. Four m/z values were obtained: 766, 891, 953 and 1016.
Each is a protonated species and the sum of these masses, 3626, will be of the order of the M of the original peptide. The value of this sum supports option (a) in (i) above.

(iii) Electrospray ionisation data represent multiply charged ions. Using the standard formula the mean M may be obtained.

$m_1 - 1$	$m_2 - m_1$	n_2	$m_2 - 1$	M (Da)	z
892.0	297.3	3.0003	1189.3	3568.3	3
713.6	178.4	4.0000	892.0	3568.0	4
594.7	118.9	5.0016	713.6	3569.2	5
509.7	85.0	5.9964	594.7	3566.1	6

$\Sigma M = 14271.6$ Da
Mean $M = 3567.9$ Da
This more precise value confirms the conclusions found above. For an explanation of the mass difference between M_r and the sum of the hydrolysate products, refer to the answer to Example 2.

The data in (iv) are confirmatory chromatographic evidence that only four hydrolysis products were obtained.

'Delta Mass', which is a database of protein post-translational modifications that can be found at http://www.abrf.org/index.cfm/dm.home. There are hyperlinks to references to the modifications.

Protein phosphorylation and identification of phosphopeptides

Phosphate is reversibly covalently attached to eukaryotic proteins in order to regulate activity (Section 15.5.4). The modified residues are *O*-phosphoserine, *O*-phosphothreonine and O-phosphotyrosine but many other amino acids in proteins can be phosphorylated: *O*-phospho-Asp; *S*-phospho-Cys; *N*-phospho-Arg; *N*-phospho-His and *N*-phospho-Lys. Analysis of modified peptides by mass spectrometry is essential to confirm the exact location and number of phosphorylated residues, especially if no ^{32}P or other radiolabel is present. Identification of either positive or negative ions may yield more information, depending on the mode of ionisation and fragmentation of an individual peptide. Phosphopeptides may give better spectra in the negative ion mode since they have a strong negative charge due to the phosphate group. Phosphopeptides may not run well on MALDI–TOF and methods have been successfully developed for this type of instrument that employ examination of spectra before and after dephosphorylation of the peptide mixture with phosphatases.

Mass spectrometry of glycosylation sites and structures of the sugars

The attachment points of *N*-linked (through asparagine) and *O*-linked (through serine) glycosylation sites and the structures of the complex carbohydrates can be determined by MS. The loss of each monosaccharide unit of distinct mass can be interpreted to reconstruct the glycosylation pattern (see example in Fig. 9.24).

The 'GlycoMod' website, part of the ExPASy suite, provides valuable assistance in interpretation of the spectra. GlycoMod is a tool that can predict the possible oligosaccharide structures that occur on proteins from their experimentally determined masses. The program can be used for free or derivatised oligosaccharides and for glycopeptides. Another algorithm, GlycanMass, also part of the ExPASy suite, can be used to calculate the mass of an oligosaccharide structure from its oligosaccharide composition. GlyocoMod and GlycanMass are found at http://us.expasy.org/tools/glycomod/ and http://us.expasy.org/tools/glycomod/glycanmass.html respectively.

Identification of disulphide linkages by mass spectrometry

Mass spectrometry is also used in the location of disulphide bonds in a protein. Identification of the position of the disulphide linkages involves the fragmentation of proteins into peptides under low pH conditions to minimise disulphide exchange. Proteases with active site thiols should be avoided (e.g. papain, bromelain). Pepsin and cyanogen bromide are particularly useful. The disulphide-linked peptide fragments are separated and identified under mild oxidising conditions by HPLC–MS. The separation is repeated after reduction with reagents such as mercaptoethanol and dithiothreitol (DTT) to cleave –S–S– bonds and the products reanalysed as before. Peptides that were disulphide linked disappear from the spectrum and reappear at the appropriate positions for the individual components.

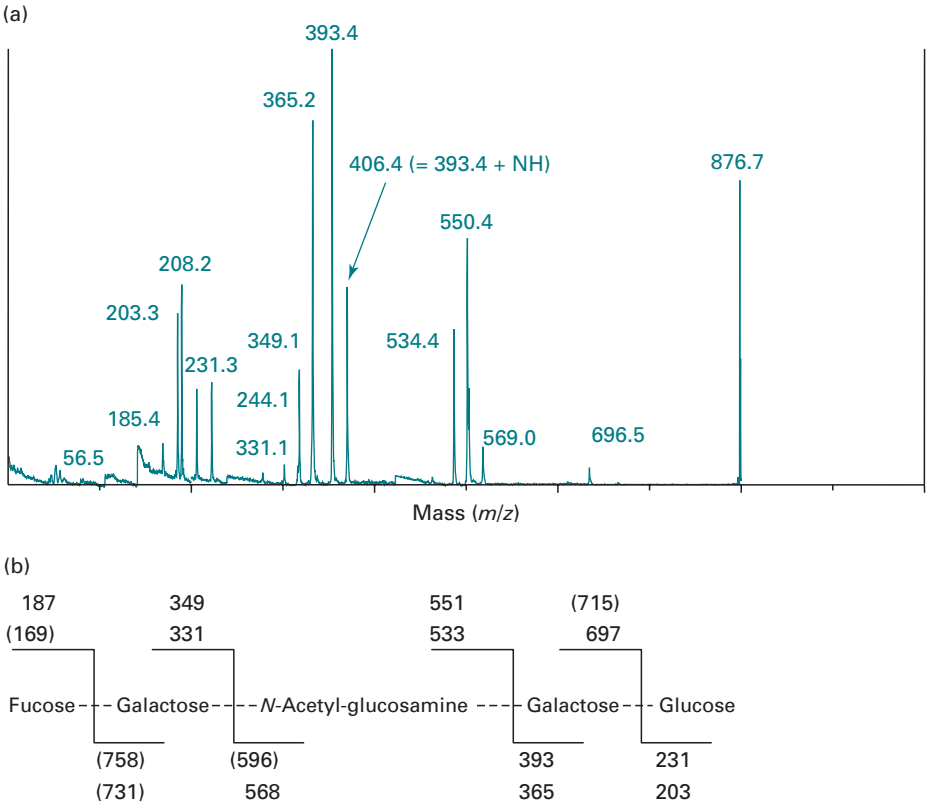


Fig. 9.24 MALDI–TOF PSD MS of carbohydrates. (a) PSD MS spectrum of the carbohydrate Fuc1–2Gal1–3Gluc*N*Ac1–3Gal1–4Glc using 2,5-dihydroxybenzoic acid (DHB) as matrix. On careful inspection of the spectrum one can observe a number of abrupt changes in baseline corresponding to where the PSD spectra have been stitched together. The peak at 876.7 Da is due to the mass of the intact molecule as a sodium adduct, i.e. the parent ion at 876.7 = $[M + Na]^+$ ion. (Spectrum courtesy of Dr Andrew Cronshaw.) (b) Interpretation of the spectrum. Experimentally derived fragment masses are mainly within 1 Da of the theoretical. The masses in parentheses were not seen in this experiment.

9.5.6 Selected ion monitoring

Selected ion monitoring (SIM) is typically used to look for ions that are characteristic of a target compound or family of compounds. This technique has particular application for on-line chromatography/MS where the instruments can be set up to monitor selected ion masses as the components elute successively from the capillary LC or reverse-phase HPLC column for example (Sections 11.3.3 and 11.9.3). Detection programmes or algorithms that are set up to carry out tandem MS on each component as it elutes from a chromatography column can be adapted to enable selective detection of many types of post-translationally modified peptides. This technique can selectively detect low-mass fragment ions that are characteristic markers that identify the presence of post-translational modifications such as phosphorylation, glycosylation, sulphation and acylation in any particular peptide. For example, phosphopeptides can be identified by production of phosphate-specific fragment ions of 63 Da (PO_2^-) and 79 Da (PO_3^-) by collision-induced dissociation during negative ion

HPLC–ES MS. Glycopeptides can be identified by characteristic fragment ions including hexose$^+$ (163 Da) and *N*-acetyl hexosamine$^+$ (204 Da).

Phosphoserine- and phosphothreonine-containing peptides can also be identified by a process known as **neutral loss scanning** where these peptides show loss of 98 Da by β-elimination of H_3PO_4 (Fig. 9.25).

9.6 ANALYSING PROTEIN COMPLEXES

Mass spectrometry is frequently used to identify partner proteins that interact with a particular protein of interest. Interacting proteins can be isolated by a number of methods including immunoprecipitation of tagged proteins from cell transfection; affinity chromatography and surface plasmon resonance. Surface plasmon resonance (SPR) (Section 13.3) technology has widespread application for biomolecular interaction analysis and during characterisation of protein–ligand and protein–protein interactions, direct analysis by MALDI–TOF MS of samples bound to the Biacore chips is now possible (where interaction kinetic data is also obtained; see Sections 13.3 and 17.3.2). Direct analysis of protein complexes by mass spectrometry is also possible. As well as accurate molecular weight of large biopolymers such as proteins of mass greater than 400 kDa, intact virus particles of M_r 40×10^6 (40 MDa) have been analysed using ESI–TOF. An icosahedral virus consisting of a single-stranded RNA surrounded by a homogeneous protein shell with a total mass of 6.5×10^6 Da and a rod-shaped RNA virus with a total mass of 40.5×10^6 Da were studied on a ESI–TOF hybrid mass spectrometer.

9.6.1 Sample preparation and handling

Mass analysis by ES–MS and MALDI–TOF is affected, seriously in some cases, by the presence of particular salts, buffers and detergents. Keratin contamination from flakes of skin and hair can be a major problem particularly when handling gels and slices; therefore gloves and laboratory coats must be worn. Work on a clean surface in a hood with air filter if possible and use a dedicated box of clean polypropylene microcentrifuge tubes tested to confirm that they do not leach out polymers, mould release agents, plasticisers, etc. Sample clean-up to remove or reduce levels of buffer salts, EDTA, DMSO, non-ionic and ionic detergents (e.g. SDS) etc. can be achieved by dilution, washing, drop dialysis and ion exchange resins. If one is analysing samples by MALDI–TOF, on-plate washing can remove buffers and salts. Sample clean-up can also be achieved by pipette tip chromatography (Section 11.2.5). This consists of a miniature C_{18} reverse-phase chromatography column, packed in a 10 nm^3 pipette tip. The sample, in low or zero organic solvent-containing buffer, is loaded into the tip with a few up- and-down movements of the pipette piston to ensure complete binding of the sample. Since most contaminants described above will not bind, the sample is trapped on the reverse phase material and eluted with a solvent containing high organic solvent (typically 50–75% acetonitrile). This is particularly applicable for clean-up of samples after in-gel digestion of protein bands separated on SDS-PAGE. Coomassie Brilliant Blue dye is also removed by this procedure. The technique can be used to concentrate samples and fractionate a mixture. Purification

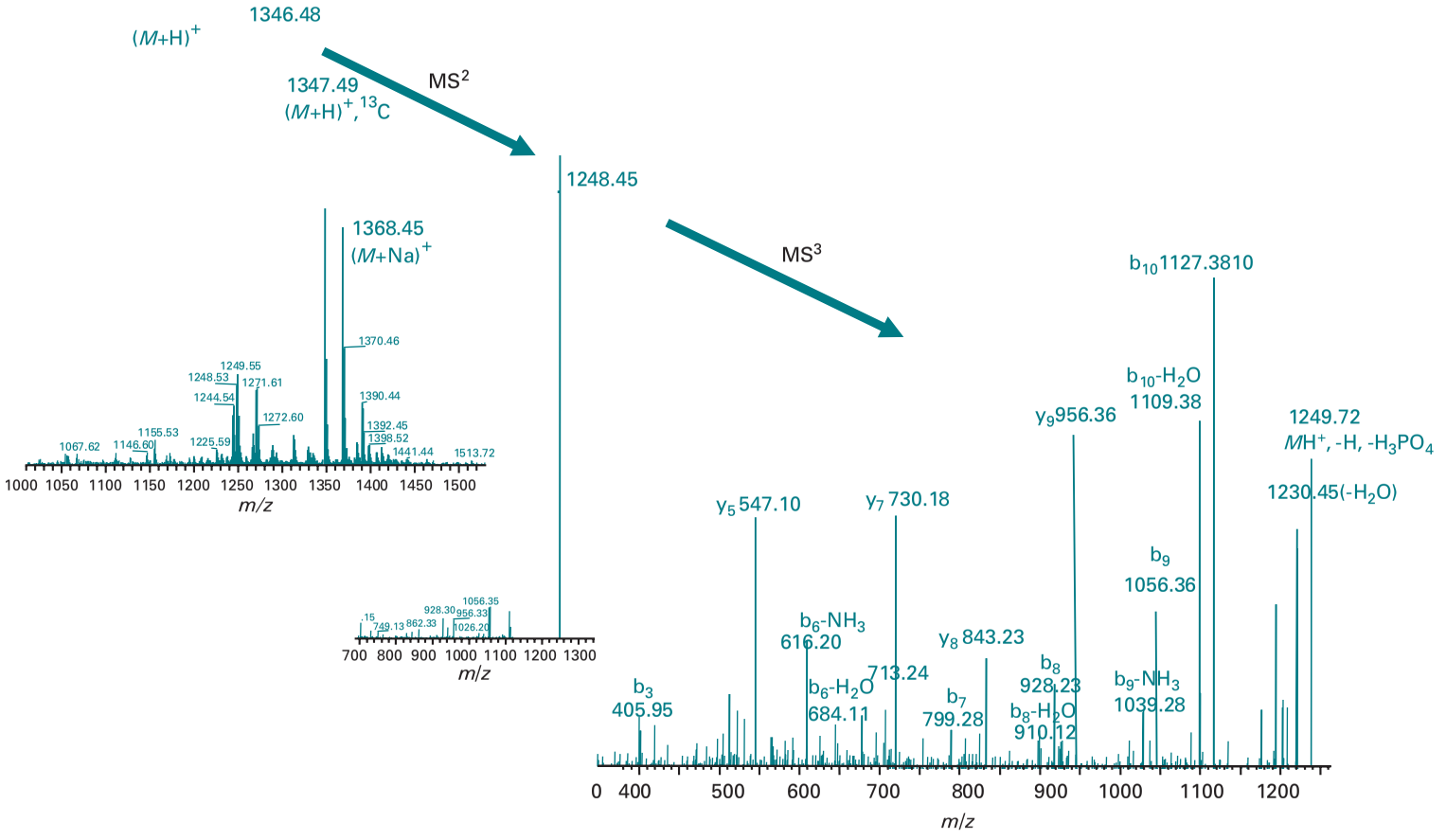


Fig. 9.25 MS identification of phosphopeptides. Sequence is YEILNS^{P}EKAC where S^{P} is phosphoserine. The MS^{2} and MS^{3} spectra are shown. The first tandem MS experiment mainly results in loss of H_3PO_4, 98 Da. Particular problems may also be associated with electrospray mass spectrometry of phosphopeptides, where a high level of Na+ and K+ adducts is regularly seen.

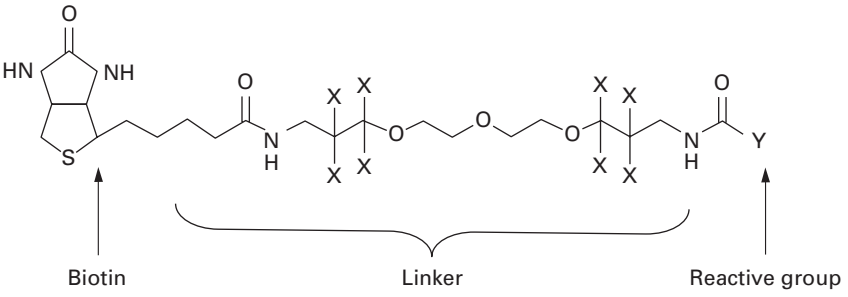


Fig. 9.26 Structure of the ICAT reagent. The ICAT reagent is in two forms, heavy (eight deuterium atoms) and light (no deuterium). The reagent has three elements: an affinity tag (biotin), to isolate ICAT-labelled peptides; a linker in two forms that has stable isotopes incorporated; and a reactive group (Y) with specificity toward thiol groups or other functional groups in proteins (e.g. SH, NH_2, COOH). The heavy reagent is D8-ICAT (where X is deuterium) and light reagent is D0-ICAT (where X is hydrogen). Two protein mixtures representing two different cell states are treated with the isotopically light and heavy ICAT reagents; an ICAT reagent is covalently attached to each cysteine residue in every protein. The protein mixtures are combined; proteolysed and ICAT-labelled peptides are isolated on an avidin column utilising the biotin tag. Peptides are separated by microbore HPLC. Since each pair of ICAT-labelled peptides is chemically identical they are easily visualised because they co-elute, with an 8 Da mass difference. The ratios of the original amounts of proteins from the two cell states are strictly maintained in the peptide fragments. The relative quantification is determined by the ratio of the peptide pairs. The protein is identified by database searching with the sequence information from tandem MS analysis by selecting peptides that show differential expression between samples.

can also be carried out to specifically bind one particular component in a mixture. Immobilised metal ion affinity columns are used to enrich phosphopeptides.

9.6.2 Quantitative analysis of complex protein mixtures by mass spectrometry

Proteome analysis (described in Section 8.5) involves the following basic steps:

- run a gel (one-dimensional (1D) or two dimensional (2D)),
- stain,
- scan to identify spots of interest,
- excise gel spots,
- extract and digest proteins,
- mass analyse the resulting peptides,
- search database.

The initial separation of proteins currently relies on gel electrophoresis which has a number of limitations including the difficulty in analysing all the proteins expressed due to huge differences in expression levels. Although thousands of proteins can be reproducibly separated on one 2D gel from approximately 1 mg of tissue/biopsy or biological fluid, the dynamic range of protein expression can be as high as nine orders of magnitude. One development that has helped to overcome some of the problems is the **isotope-coded affinity tag** (ICAT) strategy for quantifying differential protein expression.

The heavy and light forms of the sulphydryl (thiol-)-specific ICAT reagent (whose structure is illustrated in Fig. 9.26) are used to derivatise proteins in respective samples

isolated from cells or tissues in different states. The two samples are combined and proteolysed, normally with trypsin, for reasons explained above. The labelled peptides are purified by affinity chromatography utilising the biotin group on the ICAT reagent then analysed by MS on either LC–MS MS (including ion trap) or MALDI–TOF instruments. The relative intensities of the ions from the two isotopically tagged forms of each specific peptide indicate their relative abundance. These pairs of peptides are easily detected because they co-elute from reverse-phase microcapillary liquid chromatography (RP–μLC) and contain eight mass units of difference due to the two forms of the ICAT tag. An initial MS scan identifies the peptides from proteins that show differential expression by measuring relative signal intensities of each ICAT-labelled peptide pair. Peptides of interest are then selected for sequencing by tandem MS and the particular protein from which a peptide originated can be identified by database searching the tandem MS spectral data.

9.6.3 iTRAQ

An alternative method for quantitative analysis of protein mixtures is to use the iTRAQ reagents from Applied Biosystems. These are a set of four isobaric (same mass) reagents which are amine-specific and yield-labelled peptides identical in mass and hence also identical in single MS mode, but which produce strong, diagnostic, low-mass tandem MS signature ions allowing quantitation of up to four different samples (or triplicate analyses plus control of the same sample) simultaneously (Fig. 9.27). Information on post-translational modifications is not lost and since all peptides are tagged, proteome coverage is expanded, and since multiple peptides per protein are analysed, this improves confidence and quantitation.

As a consequence of mixing the multiple proteome samples together, the complexity of both MS and tandem MS data is not increased and since there is no signal splitting in either mode, low-level analysis is enhanced as a result of the signal amplification.

The protocol involves reduction, alkylation and digestion with trypsin of the protein samples in parallel, in an amine-free buffer system (Fig. 9.28). The resulting peptides are labelled with the iTRAQ reagents. The samples are then combined and depending on sample complexity, they are may be directly analysed by LC–MS MS after one-step elution from a cation exchange column to remove reagent by-products. Alternatively, to reduce overall peptide complexity of the sample mixture, fractionation can be carried out on the cation-exchange column by stepwise elution of part of the complex mixture. Very recently the company has launched a kit with eight different isobaric tagging reagents (the principle is the same).

9.6.4 Stable isotope labelling with amino acids in cell culture

Alternatives to the above include **stable isotope labelling with amino acids in cell culture** (SILAC) which is useful in investigation of signalling pathways and protein interactions. As the name implies, this technique involves metabolically labelling protein samples in cell cultures that are grown with different stable isotopically labelled amino acids such as ^{13}C and/or ^{15}N lysine and arginine. It is also possible to use ^{12}C and ^{13}C leucine.

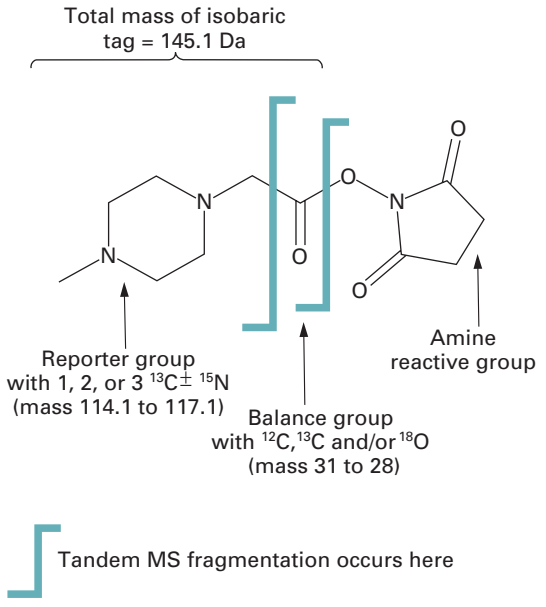


Fig. 9.27 iTRAQ reagent structure. The iTRAQ reagents consist of a charged reporter group, a peptide reactive group (an NHS, *n*-hydroxysuccinimide) which reacts with amino groups and a neutral balance group. The last part maintains an overall mass of 145. The term 'isobaric' is defined as two or more species that have the same mass. The peptide reactive group covalently links an iTRAQ reagent isobaric tag with each lysine side chain and N-terminus of a polypeptide, labelling all peptides in a given sample digest. By combining multiple iTRAQ-reagent-labelled digests into one sample mixture, the MS resembles that of an individual sample (assuming the same peptides are present). The balance group ensures that an iTRAQ-reagent-labelled peptide displays the same mass, whether labelled with iTRAQ reagent 114, 115, 116, or 117. The reporter group retains the charge after fragmentation and the balance group undergoes neutral loss. In the case of '8plex' iTRAQ reagents, additional isotopes are employed to provide reporter groups of 113 to 121 Da (with corresponding masses of balance groups of 192 to 184 (the chemical structure is changed to permit a greater variety of stable isotopes). A reporter group of 120 Da (and of course the balance group of 185) is not used as this may be confused with the ammonium ion for phenylalanine.

9.7 COMPUTING AND DATABASE ANALYSIS

9.7.1 Organic compound databases

Mass spectrometry organic compound databases are available to identify the compound(s) in analyte. The spectra in the databases are obtained by electron impact ionisation. Two such databases are:

- Integrated Spectral Data Base System for Organic Compounds (SDBS) from the National Institute of Advanced Industrial Science and Technology NIMC Spectral Database System. Data on specific compounds can be searched with compound name, molecular formula, number of atoms (CHNO) and molecular mass. The database contains 24 000 electron impact mass spectra as well as over 51 000 Fourier-transform infrared (FT-IR) spectra, 14 700 ^{1}H-NMR spectra and 13 000 ^{13}C-NMR spectra. The URL is http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi?lang=eng

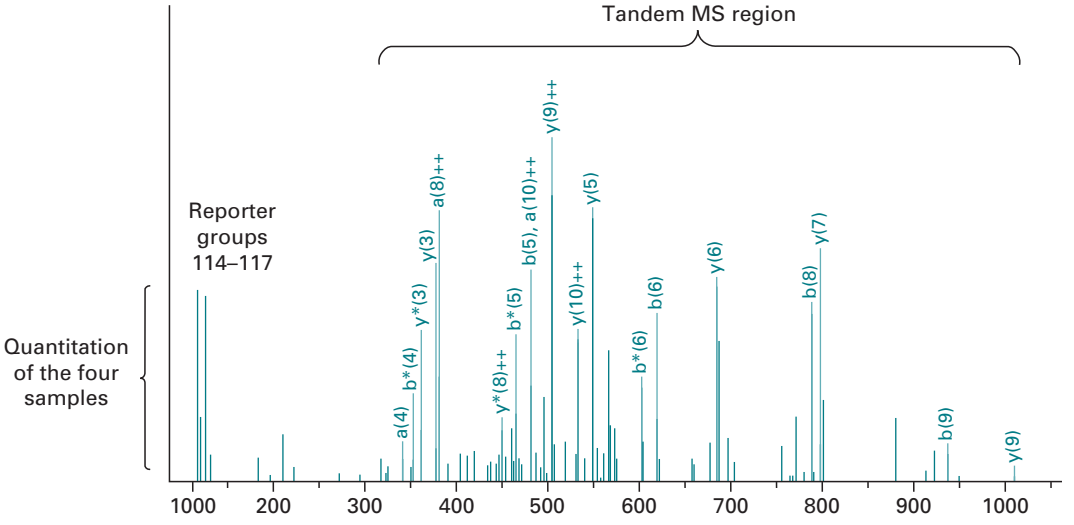


Fig. 9.28 Identification and quantitation of proteins by iTRAQ reagents. During MS/MS (along with the usual peptide fragmentation) the isobaric tag cleaves at the sites indicated. As result of fragmentation, there is neutral loss of the balance group, and the reporter groups are generated, displaying diagnostic ions in the low-mass region at between *m/z* values 114 to 117. Because this region is free of other common ions, quantification of the peak areas of these resultant ions represents the relative amount of a given peptide in the respective sample.

- **NIST Chemistry WebBook.** NIST is the National Institute for Standards and Technology. Data on specific compounds in the Chemistry WebBook can be searched by name, chemical formula, CAS (Chemical Abstracts Service) registry number, molecular mass or selected ion energetics and spectral properties. This site comprises electron impact mass spectra for over 15 000 compounds as well as thermochemical data for over 7000 organic and small inorganic compounds and IR spectra for over 16 000 compounds. The URL is: http://webbook.nist.gov/chemistry/.

9.7.2 Identification of proteins

Database searches to identify a particular protein that has been analysed by mass spectrometry are particularly important. This section gives an overview of websites for proteomic identification.

Identification of proteins can be carried out by using many websites, for example: 'Mascot' from Matrix Science (http://www.matrixscience.com/search_form_select.html) and 'Protein Prospector' (http://prospector.ucsf.edu/). The search can be limited by searching a particular species or taxon, e.g. mammalia only, thus increasing the speed. However, when looking for a homologous sequence the species should not be defined. The modification of cysteine residues, if any, should be included otherwise the number of peptides matched to the theoretical list will be decreased, producing a worse **hit.** If no cysteine modification has been carried out, and if the protein originates from a gel sample, then much of this residue will have been converted to acrylamide-modified Cys.

Unmatched masses should be re-searched since sometimes two or more proteins run together on electrophoresis. Note the **delta p.p.m.** (the difference between the

theoretical and the experimental mass of a particular peptide) which should be low and consistent. This gives an indication of whether the result is genuine. If an internal calibration has been performed the **mass accuracy** parameter can be set to 20 p.p.m. For a close external calibration this should be set to 50 p.p.m. If a hit is not found with the first search this parameter can be increased.

Different protein databases can be searched:

- MSDB is a non-identical database maintained by the Proteomics Department at Imperial College, London.
- NCBInr is a non-identical database maintained by NCBI for use with their search tools BLAST and Entrez. The entries have been compiled from GenBank CDS translations, PIR, SWISS-PROT, PRF and PDB.
- SwissProt is a database that is ideal for peptide mass fingerprint searches in which sequences are non-redundant, rather than non-identical; therefore there may be fewer matches for a tandem MS search than from a comprehensive database, such as MSDB or NCBInr.
- dbEST is the division of GenBank that contains 'single-pass' cDNA sequences, or expressed sequence tags, from a number of organisms.

NCBInr is the largest database while Swiss Prot is smaller. However Swiss Prot provides the most information with the protein hits. During a Mascot search, the nucleic acid sequences are translated in all six reading frames. dbEST is very large and is divided into three sections: EST_human, EST_mouse and EST_others. Nevertheless, searches of these databases take far longer than a search of one of the non-redundant protein databases. An EST database should only be searched if a protein database search has failed to find a match. If it is known that the protein is not larger than e.g. 100 kDa then the **mass range** should be limited to prevent false hits. Although the search will be refined by limiting to a particular mass range of the intact protein, the possibility of subunits or fragments must be considered. Some information on the isoelectric point of a protein will also be known for a 2D gel sample but this should also be treated cautiously.

If a number of larger size peptides are seen in the digest then the **missed cleavages** parameter should be increased. Typically this is set to 1 or 2.

If the possibility of post-translational or other modification is uncertain then the top three options should be selected, i.e. acetylation of the N-terminus, oxidation of Met, and conversion of Glu to pyro-Glu. If phosphorylation of S, T or Y is selected when not suspected this may lead to false hits. More than one amino acid can usually be listed in the box (e.g. STY 80 to select any phosphorylation).

The list of peptide masses should be input to four decimal places if possible. In the initial search, use masses from the higher signal intensity peaks and set the **minimum number of peptides** low compared to number of masses in the peptide list. To increase the specificity of the search this number can be increased. If no hits are found then this number can be decreased in subsequent passes. Be sure to select whether the fragment and precursor ions have been calculated from monoisotopic or average masses.

Deisotoping software is available to artificially remove the ^{13}C peaks arising from chemically identical peptides but which arise from the presence of the ^{13}C isotopic

form of carbon. This simplifies the spectrum but more importantly this will ensure that the search algorithm will not be not confused and attempt to find two or more distinct peptides that each differ by 1 Da. This is particularly valuable when analysing peptide mixtures since overlapping isotope clusters are thus identified correctly and only the genuine ^{12}C peaks are reported.

If the resolution of the mass spectrum is not sufficient to resolve individual isotope peaks then the average mass is often reported. This is still the case with larger polypeptides and proteins (see Fig. 9.14) but in modern instruments, the all ^{12}C, one ^{13}C, two ^{13}C (etc.) peptide forms can be resolved (see Fig. 9.13).

Various software (including commercial software packages such as SEQUEST) is available to use the information on the fragment ions obtained from a tandem MS experiment to search protein (and DNA translation) databases to identify the sequence and the protein from which it is derived. Once the protein has been identified one can view the full protein summary and link to protein structure, Swiss 2D PAGE, nucleic acid databases, etc.

9.8 SUGGESTIONS FOR FURTHER READING

Aebersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, **422**, 198–207. (There are also a number of other, very informative proteomics reviews in this issue between pages 193 and 225.)

Breitling, R., Pitt, A.R. and Barrett M.P. (2006). Precision mapping of the metabolome. *Trends in Biotechnology*, **24**, 543–548. (Review of the study of metabolic networks in complex mixtures using the high resolving power of FT-ICR and Orbitrap MS to discriminate between metabolites of near identical mass that differ perhaps by only 0.004 Da.)

Glish, G.L. and Burinsky, D.J. (2008). Hybrid mass spectrometers for tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry*, **19**, 161–172. (Review of the development of hybrid mass spectrometers and a comparison of the particular applications of each type.)

Hah, S.S. (2009). Recent advances in biomedical applications of accelerator mass spectrometry. *Journal of Biomedical Science*, **16**, 54. (This is an Open Access article which reviews all aspects of this subject.)

Han, X., Aslanian, A. and Yates, J.R. 3rd (2008). Mass spectrometry for proteomics. *Current Opinion in Chemical Biology*, **12**, 1–8. (An excellent introductory review of the new instruments with a comparative table of their performance characteristics, such as mass resolution, mass accuracy, sensitivity, m/z range and main applications. Also covers quantitative proteomics and description of the latest jargon such as shotgun and top–down proteomics.)

Nita-Lazar, A., Saito-Benz, H. and White, F. M. (2008). Quantitative phosphoproteomics by mass spectrometry: past, present, and future. *Proteomics*, **8**, 4433–4443. (This reviews methods for identification of this important and widespread post-translational modification. This volume, no. 8 issue 21, 4367–4612 is a Special Issue on signal transduction proteomics which includes other informative articles on this subject.)

Websites

The ExPASy (Expert Protein Analysis System) server of the Swiss Institute of Bioinformatics (SIB) contains a large suite of programs for the analysis of protein sequences, structures and proteomics as well as 2D PAGE analysis (2D gel documentation and 2D gel image analysis programmes). The ExPASy suite of programmes is at http://us.expasy.org/tools/

Glyocomod and GlycanMass are found at http://us.expasy.org/tools/glycomod/ and http://us.expasy.org/tools/glycomod/glycanmass.html respectively.

Deltamass is a database of protein post-translational modifications at http://www.abrf.org/index.cfm/dm.home which can be accessed to determine whether post-translational modifications are present. There are hyperlinks to references to the modifications.

There is also a prediction program 'findmod' for finding potential protein post-translational modifications in the ExPASy suite at http://expasy.org/tools/findmod/

Information and protocols for sample clean-up are found at http://www.millipore.com/catalogue/module/c5737 and http://www.nestgrp.com/protocols/protocol.shtml#massspec

Products for Phosphorylated Peptide and Protein Enrichment and Detection IMAC columns and chromatography at http://www.piercenet.com/files/phosphor.pdf and GelCode Phosphoprotein Staining Kit. http://www.piercenet.com/