APPLICATIONS OF MASS SPECTROMETRY IN THE ANALYSIS OF PROTEINS, MODIFIED PEPTIDES AND CHIRAL ISOMERS.

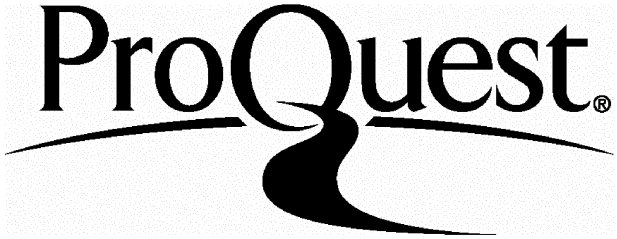
Thesis for the degree of Doctor of Philosophy

by

STEVEN ALAN HOWELL

The School of Pharmacy,

University of London.



Dedicated to Ivy and Marjorie for their constant support and encouragement, and to the Most High God for their love, faithfulness and mercy.

Acknowledgements

I am indebted to Prof. Mike Baldwin for his patient supervision and to Dr Alastair Aitken who kindly took on a supervisory role in Mike's absence.

I would also like to express my gratitude to Dr Sheren Abdulrahmen, Dr David Carter, Dr Mark Harrison, Dr John Langley, Dr Tracey Madden and Dr Kevin Welham for their friendship and expertise. In particular I thank Kevin for stepping in as academic supervisor. Likewise I am indebted to my colleagues at the National Institute of Medical Research, in particular those in the Laboratory of Protein Structure.

The technical assistance of the following was greatly appreciated:- Mr Alan Harris and Miss Sheila Laithwell (NIMR) for performing Edman sequencing and amino acid analysis respectively; Miss Yasmina Patel (NIMR) for SDS-PAGE analysis; Dr Michelle Learmonth (Thrombosis Research Institute) for BNPS-skatole digestion and Western blots; Dr Brian Green (VG Analytical) for the electrospray and high mass liquid SIMS analyses; Finnigan MAT for the laser desorption analysis; Dr Brian Coles (Middlesex Hospital) for sequencing the thymosins.

The species studied in this thesis were obtained from a number of sources detailed below:-

(a) Isoforms of bovine annexin V, Prof. R.Donato (Univ. of Perugia).

(b) *S.obliquus* & *A.variabilis* plastocyanins, Prof.A.G.Sykes (Univ. of Newcastle)

(c) Rat thioesterase II tryptic peptides, Dr A.Slabas (Unilever plc)

(d) Thymosin N-terminal peptides, Dr A.Haritos (Univ. of Athens)

(e) Myristoylated $pp60^{src}$ peptide, Dr Torsen Saermark

(e) Pyroglu-phe-pro-amide, Dr Z.Khan (NIMR)

(f) Phosphorylated MARCKS peptide, Dr B.Amess (NIMR)

(g) Lipoylated fragments of bovine E2 PDC, Dr S.Yeaman (Univ. of Newcastle)

(h) Dialkyl tartrates, Dr F.Winkler (Univ of Tubingen)

The award of a research scholarship by the School of Pharmacy is gratefully acknowledged.

ABSTRACT

The current expansion in the application of mass spectrometry (MS) to determine the primary structures of proteins and peptides is reflected in the bulk of this thesis. At the protein level, electrospray and fast atom bombardment (FAB) MS have been integrated with conventional Edman sequencing to establish the sequences of two isoforms of annexin V co-expressed in bovine brain. Using a similar approach the sequences of plastocyanins from the cyanobacterium *Scenedesmus obliquus* and the blue green alga *Anabaena variabilis* have also been ascertained.

Peptide studies focus on several post translationally modified species. N-terminal acetylation has been proven to exist in thioesterase II from rat mammary gland. The sequences of thymosin β_{11} and a novel isoform β_{12} have been determined and shown to be acetylated at their N-termini. The less common modification of N-terminal myristoylation is encountered in the analysis of a $pp60^{src}$ derived peptide. The formation of pyroglutamate has been proven to occur in a novel tripeptide present in human seminal fluid, together with C-terminal amidation. A study is described to identify a similar species present in rabbit prostate complex by FAB sequencing.

Phosphorylation is a major post translational modification having an essential role in cell regulation, the phosphorylation of a peptide derived from a major substrate of protein kinase C is described.

The sequences of peptides derived from bovine heart pyruvate dehydrogenase complex possessing lipoic acid prosthetic modifications have been determined. The final studies of modified peptides focus on the use of FAB-MS to characterise synthetic peptides, with important advantages noted over conventional protocols.

The remainder of the thesis concerns a study in the detection of chirality using

FAB-MS. The differentiation of dialkyl tartrate enantiomers due to dimer effects is described, however attempts to resolve enantiomers of threonine and serine derivatives were unsuccessful.

Contents

List of tables and figures

Abbreviations

ATP	adenosine triphosphate
CaBP	calcium binding protein
CAPS	3-cyclohexylamino-1-propane sulphonic acid
cDNA	complementary DNA
CI	chemical ionisation
CID	collision induced dissociation
CNBr	cyanogen bromide
DEAE	diethylaminoethyl
DET	diethyl tartrate
DIPT	diisopropyl tartrate
DMF	dimethyl formamide
DMT	dimethyl tartrate
EDTA	ethylene diamine tetra acetic acid
EI	electron impact
FAB	fast atom bombardment
FD	field desorption
FI	field ionisation
FMOC	fluorene methoxy carbonyl
FTICR	fourier transform ion cyclotron resonance
LD	laser desorption
MS	mass spectrometry
PD	plasma desorption
PDC	pyruvate dehydrogenase complex
pGlu	pyroglutamyl
PTH	phenylthiohydantoin
PVDF	polyvinylidinedifluoride
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SIMS	secondary ion mass spectrometry
TRH	thyrotropin releasing hormone
Tris	tris(hydroxymethyl)aminomethane

CHAPTER ONE

INTRODUCTION TO MASS SPECTROMETRY

1.0 Overview

In this chapter an introduction to the theory of mass spectrometry is given covering the formation, separation and detection of gas phase ions.

1.1 Modes of Ionisation

1.1.1 Electron Impact (EI)

This was one of the earliest modes of obtaining ions in the gas phase. As outlined in figure 1.1 the principle of operation is to heat a tungsten filament to produce an emission of electrons and to direct these electrons in a concentrated beam which traverses a chamber into which vapourised sample is introduced. The interaction between gaseous molecules and electrons results in the process of ionisation in approx. 1% of sample molecules (Rose and Johnstone, 1982):

$$M + e \rightarrow M^{+\cdot} + 2e$$

Other processes occurring include electron capture which produces negatively charged ions in approx. 0.01 % of the analyte:

$$M + e \rightarrow M^{-\cdot}$$

The high vacuum in the ion source ($\approx 10^{-5}$ mbar) ensures that there are no collisional processes between ions/molecules. It has been found that ion yield increases directly with electron energy and reaches a plateau around 50-60 eV (Harrison and Tsang, 1972) for the purposes of reproducibility an ion energy of 70eV has been adopted as standard. On collision a fraction of this energy is transferred to the sample molecules, however this is often sufficient to cause fragmentation to take place giving rise to a mass spectrum. Often the fragmentation process is so facile that the

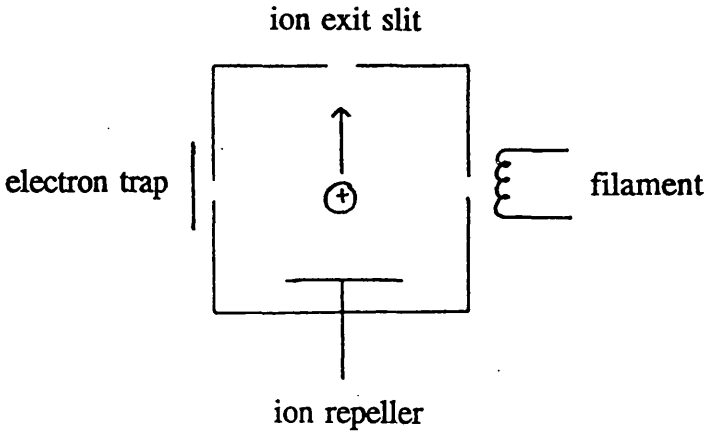


Figure 1.1 Schematic of EI source

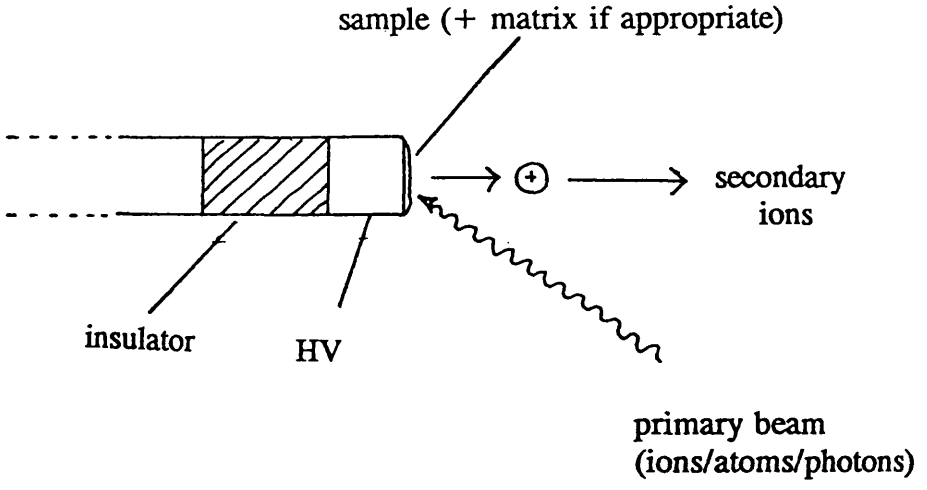


Figure 1.2 General schematic of desorption ionisation processes

molecular ion is not observed. Electron impact mass spectra are quite reproducible at 70eV and within a reasonable range of temperatures - the reproducibility of spectra has been of great practical importance in that it has led to the compilation of standard libraries of mass spectra, which are invaluable in the identification of unknown samples.

A detailed discussion of fragmentation is beyond the scope of this brief introduction, in general the molecular radical cation dissociates to form even electron cations by the loss of radicals. Rearrangement processes also occur, resulting in the loss of neutral molecules (McLafferty, 1980; Budzikiewicz *et al*, 1967).

1.1.2 Chemical Ionisation (CI)

In 1966, Field and Münson reported the generation of molecular ions by the transfer of protons from strong proton donor gas phase ions (Brønsted acids), produced from a "reagent gas" by conventional EI ionisation, to the sample under analysis. Protons are transferred with far lower increases in internal energy of the sample (Brønsted base) than with EI and as a consequence this technique affords a far gentler means of ionisation with correspondingly less fragmentation. Commonly employed reagent gases include methane, isobutane and ammonia which due to the differing acidity of their reagent ions offer different degrees of selectivity in the samples which can be ionised (Harrison, 1983). A typical reaction scheme involving a sample M, utilising methane as reagent gas, would be as follows:

$$CH_4 + e \rightarrow CH_4^{+\cdot} + 2e$$

$$CH_4^{+\cdot} + CH_4 \rightarrow CH_5^+ + CH_3^{\cdot}$$

$$CH_5^+ + M \rightarrow CH_4 + MH^+$$

To prevent the sample from being ionised by the high energy 70eV electron beam it is usually analysed at a concentration ratio of 1:1000 with respect to the reagent gas. Facilitation of the ion-molecule reaction scheme outlined above is achieved by the utilisation of a gas-tight chamber which incorporates a tiny slit for the egress of ions into the analyser region. As with EI this method is restricted to volatile samples which are not thermally labile.

1.1.3 Field ionisation/desorption (FI/FD)

By generating very large positive electric fields of 10^7 - 10^8 V/cm, on an electrode having a tip of high curvature, ionisation of gaseous samples (Field Ionisation) or of involatile samples deposited on the (heated) electrode (Field desorption) can be induced. The process of ionisation is one of extraction of an electron from the sample molecule to the positive electrode, resulting in the positively charged species being accelerated away into the analyser. Ion yield increases exponentially with the electric field strength, E, which can be related to the voltage, V, applied to a point with radius of curvature, r, by the equation $E = V/r$. Thus an improvement in the technique was forthcoming with the observation that fine needles, made of carbon for example, having radii far smaller than the original surface could be grown on the anode (Reddy *et al*, 1977).

As the region close to the ionising electrode is of relatively high pressure ion/molecule reactions can occur with the result that $[M+H]^+$ ions are usually formed as opposed to $M^{+\cdot}$. Fragmentation is generally much lower than in EI.

1.1.4 Secondary Ion Mass Spectrometry (SIMS)

By bombarding an involatile sample on a metallic surface with a KeV stream of primary ions such as cesium, emission of secondary ions can be effected from the sample (Benninghoven and Sichtermann, 1978). The technique has been successfully adopted in a range of applications ranging from depth profiling of metal surfaces to analysis of lunar rocks (Burlingame *et al*, 1990). However its widespread use is limited due to the extensive fragmentation incurred by the majority of organic samples and the short lifetime samples can exist before extensive damage from the irradiating plasma beam.

1.1.5 Laser Desorption

By applying an intense short pulse of high energy photons to a thin layer of analyte deposited on a metal surface (held at high potential) in the presence of inorganic salts (doped if necessary) cationated molecular ions may be desorbed. The mechanism of ion formation is somewhat dependent upon the laser power employed. Thus with high intensity pulses it was found that the wavelength of the laser beam does not markedly affect the spectra (Posthumus *et al*, 1978), whereas a low intensity laser beam did show a wavelength dependence related to the infra red absorption bands of the analyte (Stoll & Röllgen, 1979). The former mechanism would appear to be essentially one of thermal desorption with associated ion/molecule reactions. However the range of molecules which can be effectively desorbed by laser desorption without extensive fragmentation is somewhat limited.

A major breakthrough came with the discovery that certain organic compounds with high extinction coefficients could be used as matrices for LD, acting by

transferring energy to the analyte. The technique can produce relatively soft ionisation of laser transparent species with little fragmentation. Various organic matrices have been employed, such as cinamminic acid derivatives and caffeic acid, enabling the use of both UV and IR lasers (Hillenkamp, 1990). The use of fine metal particles suspended in glycerol has also been advocated as a matrix system (Tanaka, 1987). Results have been impressive with species having molecular weights of as high as 236,000 (Catalase) being detected (Hillenkamp and Karas, 1990). The main drawback of the technique at present appears to be the low resolution attainable.

1.1.6 Californium-252 Plasma Desorption (PD)

The radionuclide ^{252}Cf decays to form about 40 different pairs of fission products (eg $^{142}Ba^{18+}$ and $^{106}Tc^{22+}$), which are emitted in opposite directions with MeV energies (see figure 1.3). When one ion of a fission pair impinges on sample deposited on a nickel foil matrix rapid heating occurs producing a local temperature of about 10,000K. Rapid volatilisation of sample is effected before decomposition can take place and ionisation occurs through ion-pair formation and ion/molecule reactions, largely to produce molecular or quasi molecular ions with relatively little fragmentation (McFarlane and Torgerson, 1976). Dimers can be formed which can fragment to give positive and negative ions with or without proton transfer. The other ion from the fission pair is used to trigger a timing circuit which enables the desorbed ions to be timed on reaching a "stop" detector in a time of flight mass spectrometer (and hence assigned a mass, see section 1.2.6). Ions with mass to charge ratios of up to 45,000 daltons have been reported (Jonsson *et al*, 1989) - prior to the development of matrix assisted LD and electrospray, PDMS was the ionisation

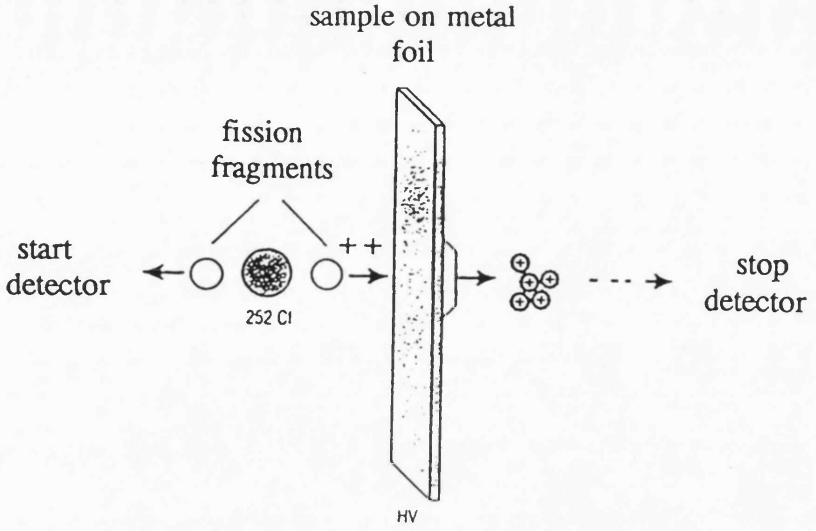


Figure 1.3 252-Californium desorption ionisation process

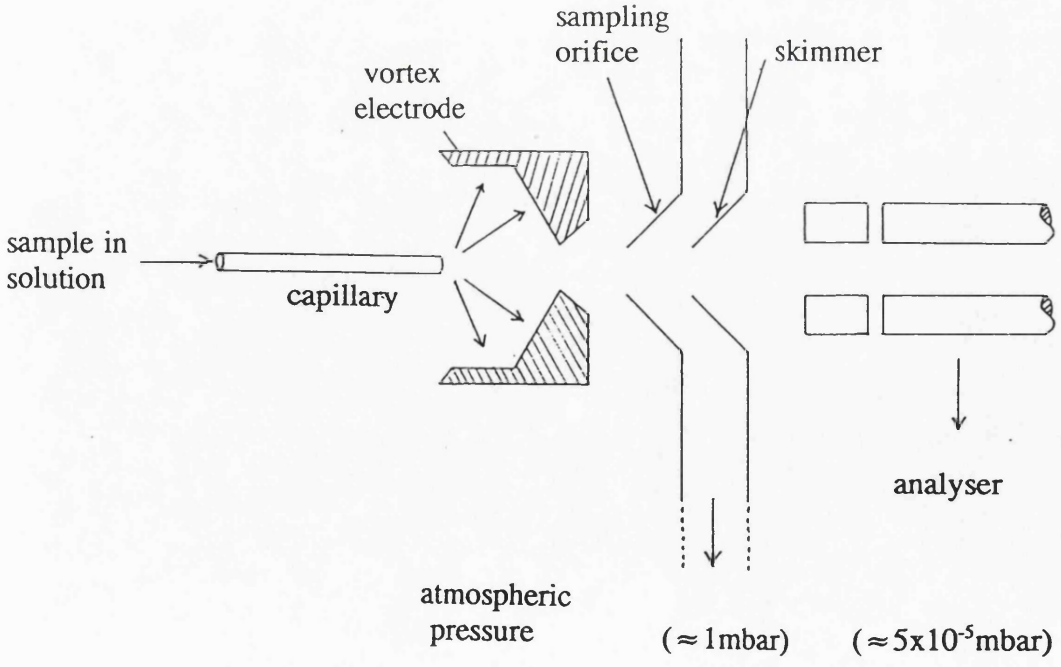


Figure 1.4 Electrospray ionisation schematic

method of choice for high mass analysis. Sample purity is important in this technique as salt impurities can cause quenching of molecular ion species. Protocols for simple sample preparation have therefore been developed to remove contaminating species (Roepstorff, 1990). The accuracy of molecular weight determination appears to be in the order of 0.1% with resolution in the order of a few hundred. Sensitivity is very high with picomolar requirements, very little of the sample is actually used in the analysis and unused sample can be recovered.

1.1.7 Fast Atom Bombardment (FAB)

This technique emerged as a development of SIMS with the adoption of a liquid matrix for sample dispersal and also with the use of an irradiating beam of atoms rather than ions (Barber *et al*, 1981). (Initially it was thought that the use of an ion beam was partly responsible for the observed damage to samples under SIMS conditions and hence an irradiating beam of neutral atoms was used. It was later realised (Aberth and Burlingame, 1984) that an ion beam could be used, hence the name liquid SIMS was applied to this technique). This adoption of a liquid matrix was due to the observation that involatile liquids such as glycerol gave long lived spectra when bombarded with an uncharged keV beam of xenon - and furthermore that samples dispersed in glycerol also gave relatively long lived spectra in which protonated pseudomolecular ions were frequently observed. Involatile and thermally labile organic species, rather than proving to be troublesome to analyse, in fact generally proved to be more easy to analyse than relatively non polar species (a degree of complementarity is therefore observed with EI for example). As magnetic sector technology advanced to produce high mass range capabilities of up to 25,000

u (at 6kV accelerating voltage) FAB was demonstrated to provide meaningful molecular ion data for high mass species such as small proteins (Barber and Green, 1987).

The mechanism of FAB is still a matter of some debate, salient principles may be outlined as follows:

(a) In the absence of cationic impurities the pseudomolecular ion (if observed) is the protonated species $[M + H]^+$ and fragmentation is generally mild.

(b) Species which form ions in solution appear to be sputtered directly into the analyser (Cooks and Busch, 1983). In the case of non-ionic species many more neutrals than ions are thought to be sputtered into the gas phase (Derrick and Sundqvist, 1987; Williams *et al*, 1987).

(c) When pairs of aromatic amines (A and B) are analysed such that the gas phase basicity of A is greater than B, whereas the solution phase basicity of A is less than B, it is A which is preferentially ionised - the degree of preference being related to the difference in gas phase basicities (Sunner *et al*, 1986). (This conclusion was strengthened when differences in surface activity where also considered (Lacey and Keough, 1989)). From studies using a split-probe FAB has been described as a CI type process occuring in a high pressure region near the sample surface (Munster *et al*, 1987).

(d) In the analysis of pairs of long chain amine species of differing chain lengths, but similar gas phase basicities, a strong preference for the ionisation of the more hydrophobic species was observed (Lacey and Keough, 1989). FAB is thus a technique in which hydrophilic species can be suppressed in the presence of more hydrophobic species.

The quantification of a given species typically requires the synthesis of a deuterated analogue as, for example, in the analysis of platelet activating factor and its lyso metabolite (Mallet, 1990).

A large range of liquids have been tested as matrices since the inception of FAB (Gower, 1985) . The qualities of a good matrix being, in general, an ability to withstand the high vacuum, ability to disperse analytes, chemical inertness to analytes and production of a low background of matrix ions (Depauw, 1991). The most popular matrices adopted include glycerol, thioglycerol and *meta* nitro-benzyl alcohol. Matrices are commonly acidified as this can increase ion yields with certain samples (Shiea and Sunner, 1991).

1.1.8 Electrospray

One of the most important recent advances in ionisation technology has come with the rediscovery of an electrohydrodynamic technique (Dole *et al*, 1968) which exploits a fundamental property of mass spectrometers, namely the measurement of the mass to charge ratio of an ion. The main development has been the application of this phenomenon to measure the molecular weights of biological materials such as peptides and proteins which can carry large charges (Fenn, 1989). The principles of the technique are outlined in figure 1.4 in which the analyte is pumped in an acidic solution at a rate of a few microlitres a minute through a fine bore metal needle held at a high potential (2-5 kV) and the resulting spray, which is formed at atmospheric pressure, directed through an orifice into the analyser. The charged droplets then evaporate to produce droplets of a few microns diameter which have high charge densities, from which analyte ions are desorbed with varying degrees of protonation

having a range of mass to charge ratios. There are two unknown quantities for a given peak within the resulting mass spectrum, namely the charge state of the ions and the molecular weight of the sample. If a peak of mass m_1 has n units of positive charge then an adjacent peak in the same series of (lower) mass m_2 within this series will have a charge state of (n + 1), thus the molecular weight (M) can be described by the following equations:

$$m_1 = [M + n] / n$$

$$m_2 = [M + (n + 1)] / (n + 1)$$

$$\therefore \quad M = [(m_1 - 1)(m_2 - 1)] / (m_1 - m_2)$$

A computer algorithm to search for other pairs of adjacent peaks will then give a number of solutions for M, providing for a determination of molecular weight with up to 0.01% accuracies commonly achieved. This converts to being able to determine the molecular weight of a 10kDa protein to within 1-2 mass units - hence the interest this technique has excited! This technique was developed almost simultaneously on quadrupole and magnetic sector instrumentation (Smith *et al*, 1990; Gallagher *et al*, 1990). Electrospray has also been coupled to quadrupole ion trap (Van Berkel *et al*, 1991) and FT-ICR (Henry *et al*, 1991) analysers. With the latter the resolution was sufficient to separate peaks due to protein ions containing 13C isotopes thus enabling mass and charge assignment to be based on the mass separations between such ions. A comparison of electrospray with FAB, PDMS and matrix assisted LD has been presented (Mann, 1990).

1.2.0 Analysers

1.2.1 Introduction

Having produced gaseous ions the next step is to measure their mass to charge ratios. As gas phase ions show pronounced mass dependent phenomena in electric and/or magnetic fields it is the application of these fields which underlies the various technologies now available for ion discrimination.

There are various factors which are important to consider when evaluating these different techniques such as (a) the accuracy of mass measurement, (b) the mass to charge range available, (c) the speed with which a scan can be recorded, (d) the resolution attainable (usually defined as $M/\Delta M$ with a 10% valley), (e) the degree of ions transmitted from the ion source to the detector (ie the sensitivity) and (f) suitability for tandem mass spectrometry. These factors will be considered in the following discussions of the major types of analyser available where appropriate.

1.2.2 Magnetic sector and double focussing devices

A typical normal geometry mass spectrometer is depicted in figure 1.5. Ions are emitted from the ion source under the influence of an accelerating voltage of 4-8,000 V as a divergent beam which is then passed through two parallel radial plates across which an electric field (E) is applied. To negotiate this sector ions (of mass m and z units of charge e) must experience an electric force equal to the centrifugal force encountered in describing an arc of radius R:

$$zeE = mv^2 / R \qquad (2.1)$$

The kinetic energy of the ions is defined by the accelerating voltage V:

$$zeV = \tfrac{1}{2}mv^2 \qquad (2.2)$$

Combining (2.a) and (2.b) gives the equation of motion:

$$R = 2V / E \qquad (2.3)$$

The energy focussed beam is then passed through a perpendicularly applied magnetic field (B) such that ions describe radii of curvature (r) dependent on their centrifugal force being balanced by the applied force due to the magnetic field:

$$Bzev = mv^2 / r \qquad (2.4)$$

Combining with (2.b) gives the equation of motion through the magnet:

$$m/ze = B^2r^2 / 2V \qquad (2.5)$$

Ions can then be detected either by array detection or can be focussed at a fixed radius of curvature onto a point detector. Sequential focussing onto a point detector can be achieved by scanning B or V. Hysteresis (eddy currents) is produced if currently available magnets are scanned at rates faster than 0.1 s/decade, although air-core magnets free from hysteresis have been designed (Bateman *et al*, 1985). Scanning V results in decreased sensitivity as V is decreased and can also lead to the ion beam being defocussed (Cottrell and Greathead, 1986).

It can be seen from equation 2.6 that the mass range of the instrument can be extended by increasing B or r and decreases as V is increased. Magnetic fluxes of up to 1.8 T can be obtained using pure iron or low carbon steel, whilst fluxes of up to 2.35T can be achieved using vanadium permandur alloy (2%V, 49%Fe, 49%Co) as pole pieces on a laminated low carbon steel yoke (Cottrell and Greathead, 1986). Increasing r results in a decrease in sensitivity as also does a decrease in V. Typically r = 30cm and V = 8,000V to give a mass range of up to about 3,000. The largest mass range analysers commercially available typically have r = 60-70cm and scan up to 15,000 D at 8,000V.

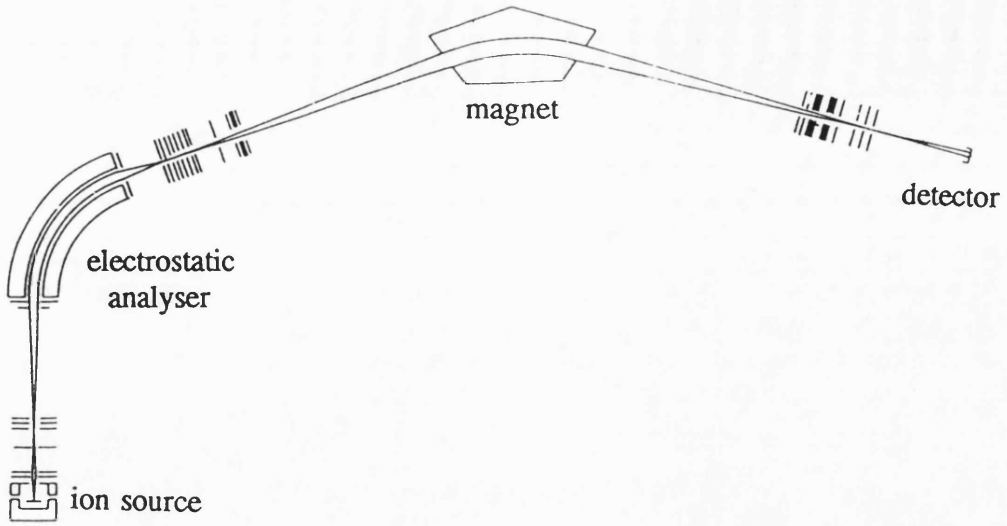


Figure 1.5 Magnetic sector double focussing mass spectrometer

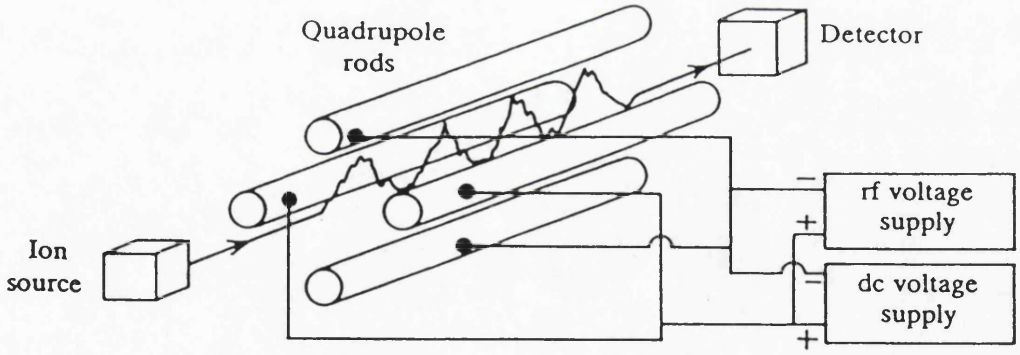


Figure 1.6 Quadrupole mass spectrometer schematic

By using mechanical slits to reduce beam divergence, usually situated a few cm after the ion source and a few cm before the detector, resolutions of the order of 100,000 are possible although at much reduced transmissions. For accurate mass measurement a resolution of 10,000 can typically be achieved at a transmission of 5%. In a reverse geometry instrument the order of the electric and magnetic sectors is reversed to give a BE configuration, which has greater versatility for the study of metastable ions (see section 1.2.7).

1.2.3 Quadrupole

Low kinetic energy gas phase ions (< 100eV) can be separated on a mass to charge basis when passed through electric fields produced by using four parallel rods. The rods, which are of circular or hyperbolic cross section, are paired electrically as in figure 1.6 and consist of a direct voltage component U and a superimposed high frequency voltage $V_0\cos(\omega t)$. Ions follow complex trajectories due to the fields (Lawson and Todd, 1972), the equations of motion for ions of mass m and charge z travelling in the z plane, through rods of separation $2r_0$, being of the Mathieu type:

$$m\left(\frac{d^2x}{dt^2}\right) + \frac{2z(U + V\cos\omega t)x}{r_0^2} = 0 \qquad m\left(\frac{d^2y}{dt^2}\right) - \frac{2z(U + V\cos\omega t)x}{r_0^2} = 0$$

the general form being:

$$\frac{d^2x}{d\gamma^2} + (a + 2q\cos 2\gamma)x = 0$$

where x = displacement, $\gamma = \omega t/2$ and the coefficients a and q are given by

$$a = \frac{8zU}{mr_0^2\omega^2} \qquad q = \frac{4zV}{mr_0^2\omega^2}$$

By plotting a graph of a versus q a stability diagram can be produced as shown in

figure 1.7. (In theory there are other regions where oscillations are stable in both x and y planes, the region depicted has the lowest values of a and q and is the one adopted in practice in quadrupole filter design). For varying ratios of a/q different size windows of m/z values are stable. At the apex of the stability area the window is narrow enough to allow ions of value m_1 to be transmitted whilst m_2 and m_3 do not pass through the filter. At the apex $a/q = 0.334$, and as $a/q = 2U/V$, by scanning U and V at a ratio of 0.167 a scan of m/z is effected in which ions pass through the filter sequentially. For increasing values of m/z there is a decrease in velocity of the ions and a corresponding increase in their residence time in the quadrupolar region (and hence exposure to the electric fields). This has two major effects: (i) For $a/q = 0.334$ unit resolution is attained at a given m/z value (up to a maximum of about 2000) and (ii) sensitivity decreases as m/z increases. The reason for this is that ions not travelling through the centre of the rods experience fringe fields which cause unstable oscillations, thus as the residence time increases as m/z increases more ions are lost than at lower m/z values. This can be circumvented to some degree by passing the ion beam emergent from the source through a narrow circular aperture so that ions are directed away from the fringes. As the quadrupole does not suffer from the hysteresis encountered with magnets scan rates in the order of ms/decade are attainable. This fact, as well as the low operating voltage of the source, the relatively high pressures which can be tolerated and the rugged design, is why the quadrupole is the most popular analyser for many hyphenated chromatographic techniques, such as capillary GC-MS. Currently quadrupoles have m/z ranges up to 4000, with transmission typically better at low masses than with double focussing machines.

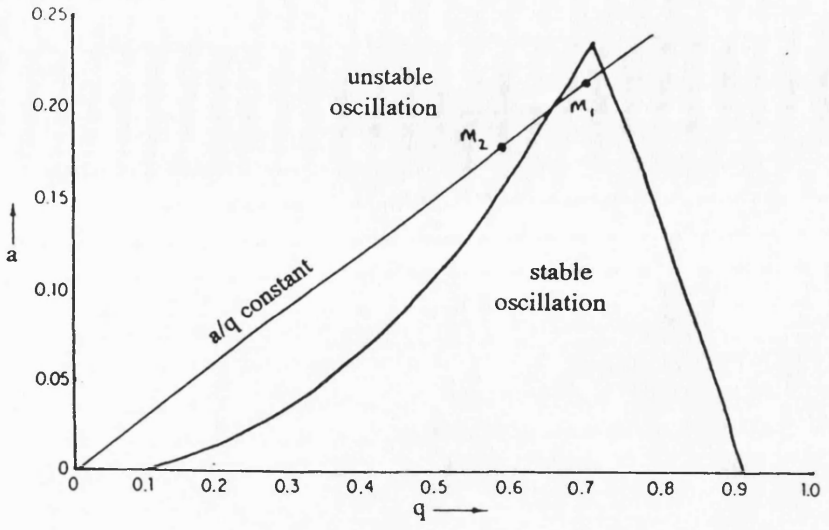


Figure 1.7 Stability diagram of quadrupole analyser

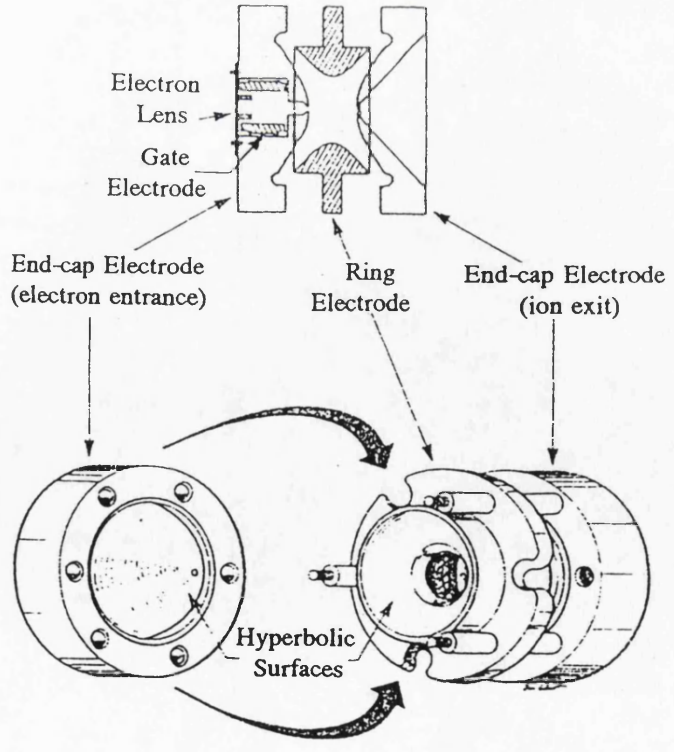


Figure 1.8 Quadrupole ion trap schematic

1.2.4 Quadrupole ion trap

As can be seen from figure 1.8, the term quadrupole in this device is a misnomer because the central electrode is a continuous taurus. Ions are generated in the centre of the device and experience electric fields as in a standard quadrupole with similar Mathieu equations of displacement. (The graph of a versus q has a different relief than for a quadrupole due to the asymmetry imposed by the taurus). The main difference to a standard quadrupole is that stable ions are "trapped" between the electrodes. Ions are detected by raising the electric field such that ions are sequentially forced into unstable trajectories thus hitting the electrodes or colliding with a detector outside the trap (Nourse and Cooks, 1990). The mass resolution of the device is enhanced by utilising an inert bath gas, such as helium (which leads to a reduction in the spread of the kinetic energies of trapped ions) and is of the order of 1,000. The sensitivity of the device is excellent with attomole limits reported in favourable cases (Cox *et al*, 1992). By missing a small window when the electric field is raised the ion trap can be used in an MS/MS mode to isolate ions which can subsequently be used for MS/MS, with the capability of MS^n analysis (Cox *et al*, 1992), however parent ion and neutral loss scans cannot be performed.

External ion sources can be fitted to enable injection of ions into the ion trap and there is also the potential for hybridisation with other analysers. A mass range of around m/z 650 is commonly quoted although extension to values as high as 45,000 have been reported (Nourse and Cooks, 1990).

1.2.5 Time of flight

This form of analyser relies on the principle that when ions of different m/z

value are accelerated by the application of a high voltage V then, as their kinetic energies are equal, their velocities are mass dependent and given by $zeV = \frac{1}{2}mv^2$ (equation 2.1). Thus the time taken to traverse a distance l is mass dependent and given by:

$$t = l \cdot \sqrt{\frac{m}{2zeV}}$$

To enable the timing of ions from their source to a detector necessitates a pulsed mode of operation, either of the ionisation or of the accelerating voltage. A timing circuit is activated with each pulse for a fixed time value dependent on the mass range required to be covered. Scan times are typically extremely fast and in the order of microseconds. It is apparent that there is no theoretical mass limit with the time of flight methodology, if a sufficient time gap between pulses is adopted then gas phase ions of any mass can be analysed. It is this fact which has led to this analyser being the one of choice for high mass studies utilising 252-Californium plasma desorption and Laser desorption, both of which incidentally can be operated on a pulsed basis. Transmission rates are very high, which has led to femtomole detection levels of certain species (Opsal *et al*, 1985). The main drawback with this detector is that low resolving powers are obtained in the order of a few hundred. Depicted in figure 1.9 is a "reflectron" in which resolution is improved by "bunching" ions of similar momenta by means of an applied electric field at the reflection point (Gohl *et al*, 1983).

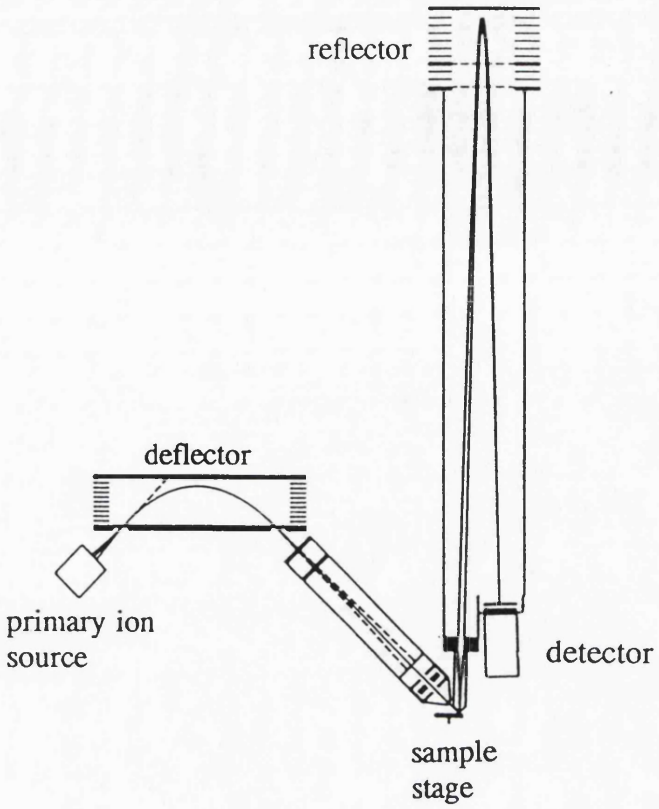


Figure 1.9 SIMS "reflectron" time of flight mass spectrometer

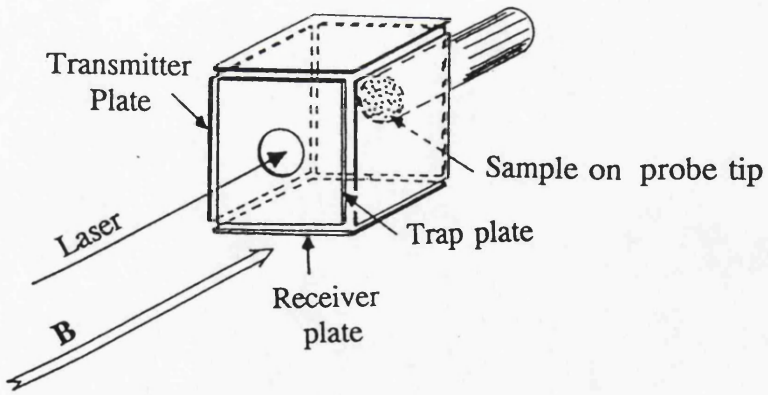


Figure 1.10 Laser desorption FT-ICR schematic

1.2.6 Ion Cyclotron Resonance

This analyser is operated on a similar principle as the quadrupole ion trap, namely that of trapping ions in stable orbits. The heart of the device is a small cell of varying geometry, one being cubic with three pairs of plates of differing functions arrayed as in figure 1.10 (Marshall and Grosshaus, 1991). Under the influence of the magnetic field, B, ions move in circular orbits and are trapped in the cell by applying a potential to the trap plates. The cyclotron frequency, ν_c, of ions of mass m and charge q is given by the approximate relationship $\nu_c = qB/2\pi m$. Fourier transformation is the preferred mode of operation rather than sequential raising of electric fields to perform scans. In one FT-ICR method after a set trapping time of 1-100ms in which ion/molecule reactions can occur, ions are excited by the application of a fast radiofrequency sweep to the face opposite pair of transmitter plates. By resonance absorption ions are brought into large circular orbits which induce so-called image currents in the receiver plates. These currents decrease with time due to ion/molecule collisions, fourier transformation of the time-domain signal results in a frequency-domain spectrum which can be converted to a mass spectrum. By applying a frequency sweep of sufficient power that is missing a small notch of frequencies, all ions except those of a selected m/z ratio may be excited to collide with the walls of the analysis cell. The remaining ions can be used for experiments such as MS/MS or even MS^n. Very high resolving powers are obtainable with figures of around 60,000 having been reported (Henry *et al*, 1991). However low analyser pressures are required in the order of 10^{-9} Torr to achieve this. If an external ion source is utilised a bath gas of helium is often used, however this results in a loss of resolution. Sensitivities are often excellent, down to femtomole levels, especially

when an internal ion source is used.

1.2.7 Tandem mass spectrometry

With the exceptions of the quadrupole ion trap and the ion cyclotron resonance analysers tandem mass spectrometry is usually performed on multi-sector instruments. Double focussing magnetic sector instruments can be used to analyse daughter ions by linking the electric and magnetic fields at constant ratio in the so-called B/E linked scan (Millington and Smith, 1977). Reverse geometry instruments can also analyse daughter ions by means of MIKES scans (McLafferty, 1980) in which parent ions are selected with the magnet and the electric sector scanned. Poor resolution and difficulties assigning masses accurately (Neumann and Derrick, 1984) may limit the usefulness of the latter technique. In these methods formation of metastable ions can be enhanced by the use of gas filled collision cells located in the appropriate field free regions to effect collisionally activated dissociation (CAD). A disadvantage that both techniques suffer from is a high parent ion acceptance window. Three and four sector instruments (an example of the latter is shown in figure 1.11) with various combinations of electric and magnetic sectors have been developed to overcome this problem and to provide higher resolution. Hybrid instruments incorporating quadrupole analysers have also been adopted (Yost and Boyd, 1990). Pure quadrupole based tandem mass spectrometers are available in the form of triple quadrupoles in which the second quadrupole is used as a collision cell.

CAD (Hayes and Gross, 1990) is by far the most important means of inducing fragmentation in tandem mass spectrometry. A theory for the collisional activation of macromolecules has recently been proposed (Uggerud and Derrick, 1991).

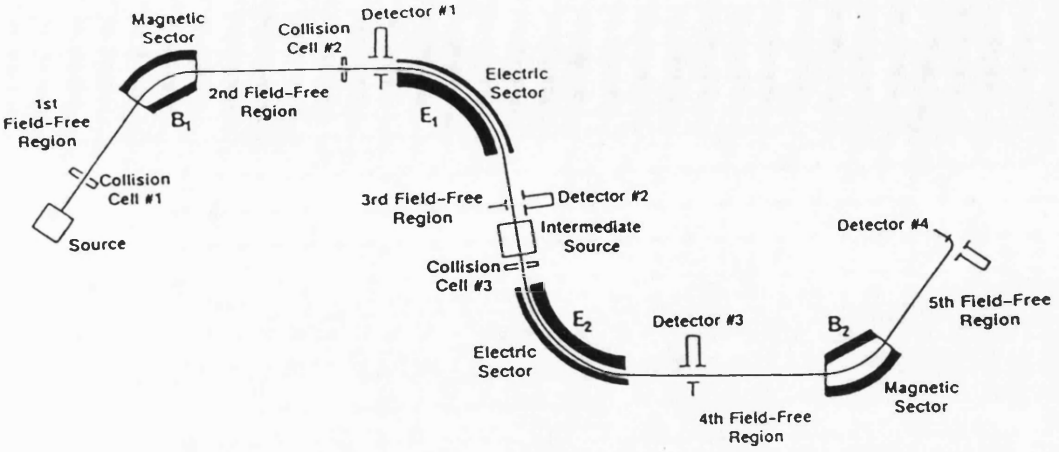


Figure 1.11 Four sector tandem mass spectrometer (VG ZAB-4F)

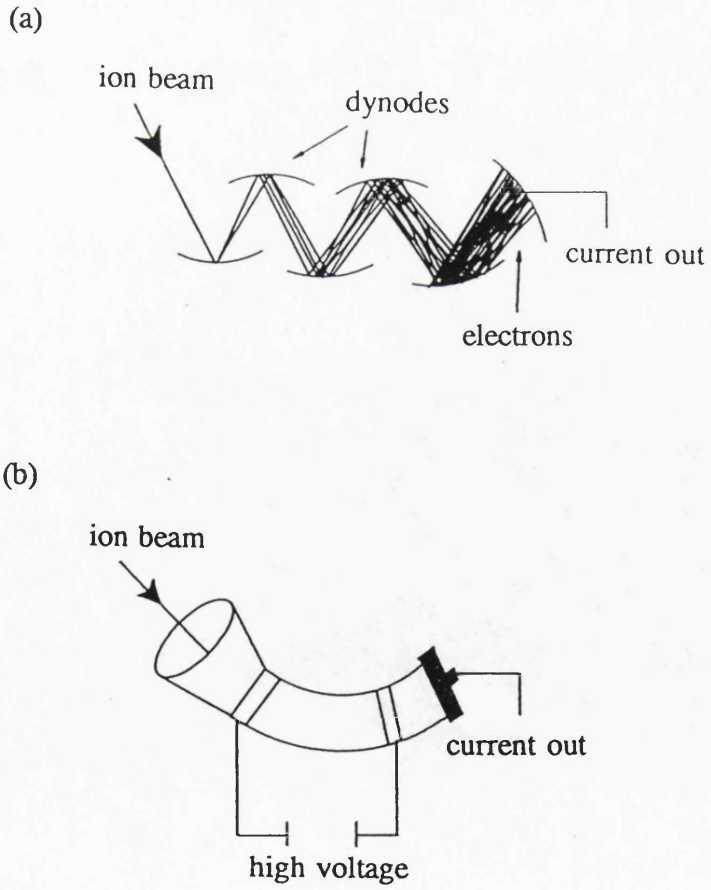


Figure 1.12 Electron multiplier devices (a) discrete stage (b) channeltron

Typically helium is the collision gas chosen, due to its minimising losses due to scattering or neutralisation compared with gases such as argon or oxygen. Apart from the use of a collision gas, photons (Hunt *et al*, 1985), electrons (Aberth and Burlingame, 1990) and solid surfaces (Cody and Freiser, 1979) have also been used to effect dissociation of parent ions.

1.3.0 Detectors

The most commonly adopted means of detecting the gas phase ion beam which has passed through a filtering device are described in the following sections. The prime requisite of the detector is clearly to obtain the highest signal to noise ratio recording of the ion current. This signal may then be passed to a fast response time direct recorder, such as an ultra violet detector, or to an analogue-digital converter for subsequent computerised data analysis.

1.3.1 Plate detector

The simplest device is one on which the ion current is collected on a metal plate and then passed to an amplifying device. A complicating factor in measuring ion currents is that when high energy ion beams (such as are encountered with magnetic sector or time of flight machines) impinge upon a metal surface then secondary electrons are emitted from the plate. This effect can lead to measurements of ion currents being in error by several hundreds of percent. As the energy of these electrons is in the order of a few electron volts then the application of a negative potential of -50V, for example, is sufficient to repel the electrons back to their origin (Evans, 1990). Ion currents as low as 10^{-16}A can be detected, however the electrical

noise encountered at this limit demands a slow response time in the amplifier, which is not compatible with fast scanning analysers. The faraday cage detector is an extension of this design in which the detecting plate takes the form of a rectangular cage having a narrow aperture.

1.3.2 Electron multiplier

The direct measurement devices are too insensitive for most analytical applications, for which a multiplier unit is required. The most commonly used is the electron multiplier, of which their are two basic designs - discrete and continuous stage. The discrete type operates on the principle of amplifying the incoming ion by the successive emission of secondary electrons in a cascade effect as depicted in figure 1.12. The exit is held at a higher potential than the entry stage and so the secondary electrons are accelerated through to this anode before being fed to the amplifier. The discrete plates are commonly manufactured of beryllium oxide coated onto copper as this combination gives good secondary ion yields. Gains of up to 10^7 are obtained with 16-20 stages and applied voltages of 1000-4000V. The continuous electron multiplier, or "channeltron", amplifies on the same cascade effect principle and utilises a curved lead-glass tube across which the voltage is applied. The curved shape is adopted to minimise feedback effects due to the ionisation of residual gas molecules.

1.3.3 Post acceleration detectors

The yield of secondary electrons produced when an ion impinges on a surface has been shown to be proportional to the velocity of the ion rather than its kinetic

energy (Hedin *et al*, 1987), and that for bovine insulin (Mr = 5733 Da) the threshold value is 1.8 x 10^4 ms^{-1}. From the relationship $zeV = \frac{1}{2}mv^2$ (eqn. 2.2) with an accelerating voltage of 10,000V the velocities of ions of m/z values 500 and 5000 are calculated as 6.2 x 10^4 ms^{-1} and 2.0 x 10^4 ms^{-1} respectively. Thus for high mass ions the mean velocity of the ions is rather close to the threshold value quoted for insulin. This has led to the development of methods to accelerate ions after their emergence from the mass filter. One design (which is incorporated in figure 1.13) has been to impinge the ion beam on a highly polished aluminium electrode held at high negative potential of up to -20kV. Secondary electrons produced are then accelerated, due to the negative potential, towards the electron multiplier at velocities which are greater than the impinging ion beam, with the gain especially noticeable at high masses.

1.3.4 Photo-multiplier

By the incorporation of a photocathode at the entrance to an electron multiplier a device which acts as a photon amplifier is effected. One immediate advantage over the standard electron multiplier is that this apparatus can be sealed in a glass case and evacuated, thus prolonging the life of the multiplier. The photons needed to actuate the release of electrons from the cathode can be produced by passing an ion beam through a scintillator. To prevent the (non conducting) scintillator from acting as a capacitor it is coated with aluminium, for example, to conduct away the charge. However, if the ion beam impinges on the aluminium coated scintillator directly, the result is that the aluminium is sputtered away and much of the ion beam does not pass through to the scintillator. Thus a conversion dynode is employed at high negative potential to produce secondary electrons which produce very little sputtering of the

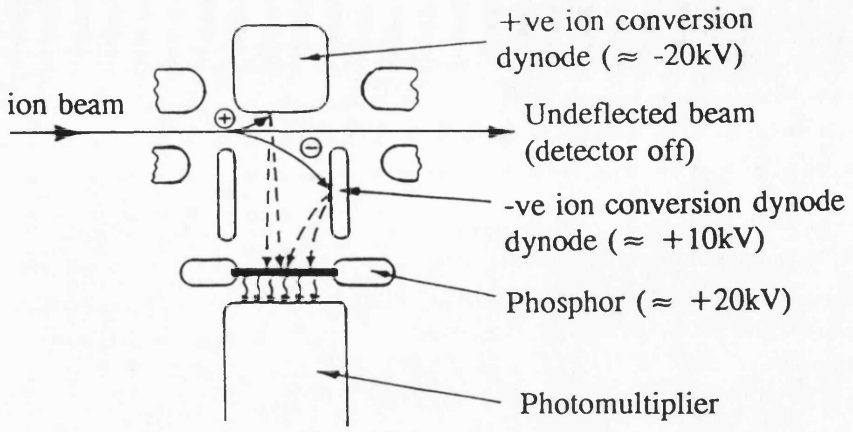


Figure 1.13 Photomultiplier device (Bateman *et al*, 1985)

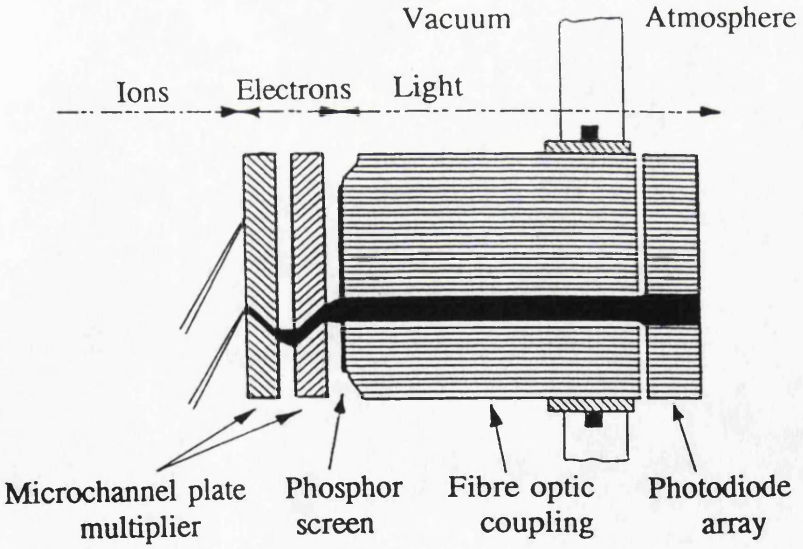


Figure 1.14 Array detector (Cottrell and Evans, 1987)

aluminium and readily produce scintillation. It cannot be used to detect negative ions, a design enabling the detection of positive and negative ions (Bateman and Butt, 1987) is depicted in figure 1.13.

1.3.5 Array detector

The relative inefficiency of array detection using a photoplate led to the development of point detectors based on analyser scanning techniques. Recently however there has been a return to methods of array detection, the main impetus being to improve detection efficiencies for large mass/charge measurements and for tandem MS experiments. Two types of array detector are currently commercially available. The photodiode array depicted in figure 1.14 was developed first and is comprised of 1024 separate microchannels which enable coverage of 4% of the mass range at a resolution of 1-2,000. The device can be operated in a stepped fashion to cover larger mass ranges. An increase in ion collection efficiency of a hundred fold is claimed compared with conventional point detectors (Cottrell and Evans, 1990). The more recently developed PATRIC (Position And Time Resolved Ion Counter) array device works on the principle of detecting secondary electrons emitted from a microchannel plate directly onto a plate of 54 conducting strips (Pesch *et al*, 1989). This plate is mounted to allow the secondary electron cloud emitted from the microchannel plate to fall as a spot over a number of strips. Using charge sensitive amplifiers the position of the centre of the spot and its intensity are recorded. About 8% of the mass range can be covered at a resolution of up to 7500, with sensitivity similar to the photodiode array (Evans, 1990).

CHAPTER TWO

INTRODUCTION TO THE ANALYSIS OF PROTEINS AND PEPTIDES

2.0 Overview

In this chapter current methodology for determining the primary structures of proteins and peptides is discussed. The use of mass spectrometry is described and also the purification requirements when analysing biomolecules.

2.1.0 Biological expression of proteins

Proteins are synthesised on polyribosomes due to processes which are initiated within the nucleus of a cell (Alberts *et al*, 1989). Small sections of the DNA, termed operons, are transcripted to produce messenger RNA which then leaves the nucleus as a complementary to the DNA template. The mRNA then binds to polyribosomes prior to reaction with the transfer RNA-amino acid complexes, which bind to the mRNA through their anti-codons and thus bring amino acids into line. The formation of peptide bonds between amino acids results in the production of a polypeptide chain from which the tRNA molecules minus their amino acids are jettisoned. After this process of "translation" the newly synthesised protein may then be acted upon by a number of enzymes which produce structural modifications. However, the term "post- translational modification" has come to be used in its broadest sense, namely any structural difference between the protein produced and the theoretical protein structure deduced by a simple interpretation of the genetic code of the DNA encoding it (Wold, 1981). Thus this definition also includes modifications at the transcriptional level, such as RNA editing (see section 2.2.2), as well as amendments which are thought to occur co-translationally as the polypeptide chain is growing on the polyribosomes. For example, the action of acyl-hydrolases is thought to be a co-translational process (Farries *et al*, 1991).

2.2.0 Peptide and protein structural elucidation

The two techniques which are classically used in the determination of primary structure are discussed in this section, these being the Edman degradation of proteins and the determination of the cDNA sequence with the subsequent translation of the codons to deduce the sequence of the expressed protein.

2.2.1 Edman degradation

Historically this was the earlier of the two main modes of analysis, being introduced in the 1960s (Edman and Begg, 1967). The scheme, outlined in figure 2.1, is to form a phenylthiocarbamoyl derivative of the primary amine (or secondary in the case of proline) N-terminus of the peptide and then to selectively cleave the peptide bond between the modified N-terminal residue and the neighbouring amino acid using a strong anhydrous acid, such as trifluoroacetic acid. The released N-terminal residue is converted to a stable phenylthiohydantoin derivative in the presence of strong aqueous acid before being extracted using an organic solvent such as butyl chloride. The extracted amino acid derivative may then be identified and quantified using reverse phase HPLC on the basis of its retention time. The truncated peptide is then retreated with phenylisothiocyanate and the cycle repeated to provide primary structure of the peptide. The analysis may terminate before the whole structure of the peptide is determined due to either the amount of PTH-amino acid derivative being too small to be detected by HPLC analysis or due to "wash out" of the peptide in certain cases, which will be discussed shortly. Most modern sequencing is done using automated equipment for which the efficiency of the procedure is 90-95% per cycle (Hewick *et al*, 1981). With an initial sample size of

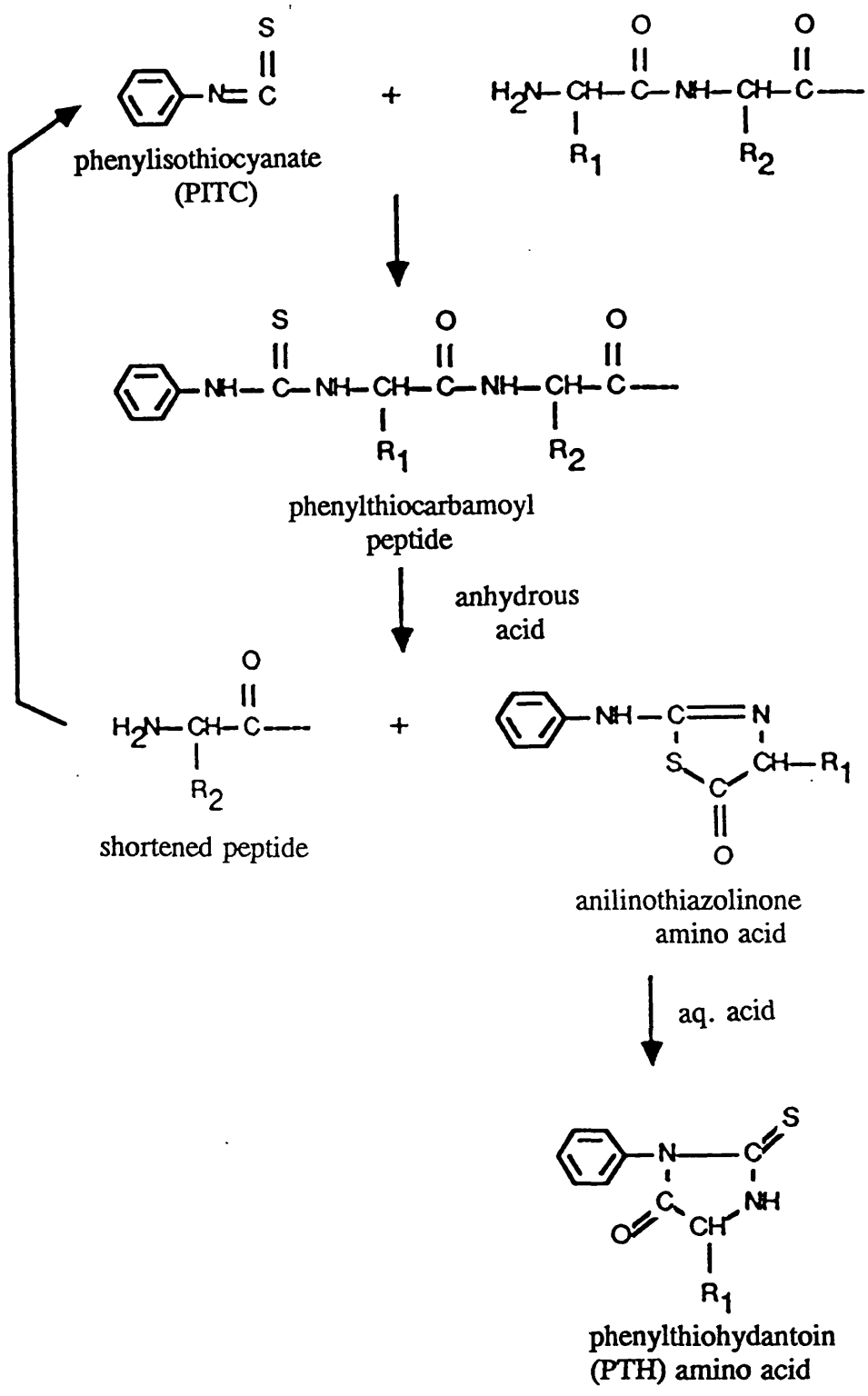


Figure 2.1 Outline of the Edman degradation peptide sequencing protocol

100pmol this converts to an ability to determine 30-40 residues, with each cycle typically taking one hour. There are two main methods of retaining the peptide in the reaction vessel, in "gas phase" sequencing this is achieved using a polybrene disc to which the peptide binds electrostatically, whereas in the "solid phase" approach covalent binding of the carboxyl terminus is utilised. With small hydrophobic peptides bound to polybrene, such as may be encountered as the C-terminus is reached, there may be insufficiently strong electrostatic bonding to prevent extraction along with the phenylthiohydantoin amino acid derivative. This is the condition referred to earlier as "wash out", a problem not encountered when the peptide is covalently bound. By the use of two or more protein cleaving agents (discussed in section 2.3.0) peptide fragments may be generated which overlap and hence enable the primary structure to be determined. One limitation of the technique is that it will not proceed if the N-terminus is blocked (Geisow and Aitken, 1989). The other main limitation is that acid labile post translational modifications of the protein will be lost in the acid dependent steps in the procedure. Another potential problem is that if a modified PTH-amino acid elutes at the retention time of a commonly occurring PTH-amino acid then it is highly probable that it will be assigned incorrectly.

2.2.2 Gene sequencing

The structure of a gene expressed protein may be determined indirectly by the translation of the codons of the relevant stretch of DNA (Alberts *et al*, 1989). Messenger RNA or complementary DNA is usually the genetic material that is used in sequence analysis. The main strategies for sequencing involve using a polymerase enzyme to make a copy of the RNA/DNA sample from the four common nucleotides

whilst a small percentage of modified nucleotides are present which prevent chain extension on incorporation. These modified nucleotides are radiolabelled or tagged with chromophores and the cDNA/RNA into which they are incorporated separated chromatographically (Maxam and Gilbert, 1980; Smith *et al*, 1986) and the sequence of individual nucleotides determined. The sequence of triplets can then be translated to determine the amino acid sequence of the encoded protein as groups of three bases specify a particular amino acid. The translation of the genetic code to derive the protein sequence is not quite as straightforward as it may appear to be, for example the determination of the reading frame is not always apparent. The phenomenon known as "RNA editing"(Benne *et al*, 1986), in which for example cytosine may be deamidated to produce uracil in mitochondrial messenger RNA (Covello and Gray, 1989), can result in errors in interpreting the genetic code if mitochondrial DNA is used (Gualberto *et al*, 1989). However, the major problem with the gene sequencing protocol is that post translational modifications cannot be determined. All newly synthesised proteins contain methionine at the N-terminus, for the majority this "initiator" methionine is either acetylated or removed by enzyme action (Arfin and Bradshaw, 1988). The resultant N-terminal amino acid may also be subject to acetylation (or similar blocking) or proteinase action. The C-terminus of a protein is also subject to pruning by carboxypeptidase action to produce truncated species. Other post translational modifications, such as phosphorylation and glycosylation, for example, are formed in complex processes often involving more than one enzyme and cannot be simply deduced from consensus sequence strategies (Aitken, 1990). It is apparent from the preceding discussion that a definitive structure for a protein cannot be obtained from knowledge of the gene encoding it. At best a deduced structure

may be highly probable, at worst it may simply be incorrect to some degree due to the presence of unsuspected post translational modifications. Many such modifications profoundly affect the function of proteins, their elucidation is therefore of great importance.

2.3.0 Role of proteases and chemical cleavage agents

Frequently proteins or large peptides are broken down into smaller peptides, by the use of agents which cleave peptide bonds, often of a specific nature, prior to subsequent analysis. Commonly used proteases and chemical agents are outlined in tables 2.1 and 2.2 together with the specificity of the cleavages they produce. Most enzymes have an optimal pH range for activity, these are detailed where appropriate. It must be noted that most enzymes do not cleave when the C-terminal bond is to proline. Typically disulphide bonds are reduced and alkylated prior to cleavage, a commonly used means being to reduce with dithiothreitol and alkylate with 4-vinylpyridine (Friedman *et al*, 1970; Amons, 1987). Guanidine hydrochloride or urea are also often added to denature proteins and thus to expose all potential cleavage sites. Following cleavage the resultant mixture of peptides may be analysed directly or separated prior to analysis. The method most often used for separation of mixtures is reverse phase HPLC, which also removes the cleaving agent employed and any other contaminants (Hunkapiller, 1984; Simpson *et al*, 1988). Typically a C_8 or C_{18} analytical column is employed with a gradient system operated under a linear programme extending from 100% solvent A (water/0.1%TFA) to 60% solvent B (acetonitrile/0.1% TFA)

Enzyme	Specificity (C-terminal side of)	Comments
Trypsin	Lys, Arg	pH 7.5-8.5
Chymotrypsin	Trp,Tyr,Phe,Leu,Met	pH 7.5-8.5
Pepsin	Phe, Met, Leu, Trp	pH 2-3
Staph. aureus V8	Glu-Xaa	Bicarbonate buffer pH 8
	Glu-Xaa, Asp-Xaa	Phosphate buffer pH 7.8
Arg-C proteinase	Arg	pH 7.5-8.5
Subtilisin	Non-specific	pH 7.5-8.5
Papain	Non-specific	Exclude thiol reagents. pH 7.5-8.5
Thermolysin	N-terminal side of Leu, Ile,Met,Phe,Val,Ala	pH 7-8

Table 2.1 Commonly used proteases (Xaa = Hydrophobic residue)

Chemical	Specificity	Comments
Cyanogen Bromide	C-terminal side of Met residues	pH 1.5-3.5. Met converted to homoserine or homoserine lactone
BNPS-Skatole	C-terminal side of Trp residues	pH 2-3
Hydroxylamine	Asn-Gly bonds	pH 9

Table 2.2 Chemical cleavage agents

2.4.0 The use of mass spectrometry in the determination of protein and peptide primary structures

The aforementioned techniques clearly have limitations in the degree to which the primary structure can be determined, especially is this true in the detection of post-translational modifications. The main role for mass spectrometry which is emerging would appear to be as a complementary technique to these "core" technologies. One of the main uses of mass spectrometry, particularly with the emergence of electrospray and matrix assisted laser desorption, is in the confirmation of a deduced structure by an accurate mass determination of the protein. (By "accurate" is meant within a few daltons, for small proteins). The determination of molecular weight can thus highlight the presence of post-translational modifications. In a recent review over one hundred and fifty such modifications of the "common twenty" amino acids were reported (Wold, 1981). The use of mass spectrometry in this important area is discussed in chapters 4 and 5.

The determination of molecular weights of high mass species, such as polypeptides and proteins, deserves some consideration due to isotope envelopes and resolution limits. For masses below about 3-4 kDa the best approach would appear to be determine the isotopic distribution of the species and to calculate the mass of the most abundant peak of the envelope from monoisotopic atomic weights. For higher masses the most abundant peak usually corresponds to the mass calculated using average atomic weights. Monoisotopic and average masses of amino acid residues are listed in table 2.3.

Amino Acid	Single letter code	Residue Mass Monoisotopic	Average
Alanine	A	71.037	71.079
Arginine	R	156.101	156.188
Asparagine	N	114.043	114.104
Aspartic acid	D	115.027	115.089
Cysteine	C	103.009	103.144
Glutamic acid	E	129.043	129.116
Glutamine	Q	128.056	128.131
Glycine	G	57.021	57.052
Histidine	H	137.059	137.142
Isoleucine	I	113.084	113.160
Leucine	L	113.084	113.160
Lysine	K	128.095	128.174
Methionine	M	131.040	131.198
Phenylalanine	F	147.068	147.177
Proline	P	97.052	97.117
Serine	S	87.032	87.078
Threonine	T	101.048	101.105
Tryptophan	W	186.079	186.213
Tyrosine	Y	163.063	163.170
Valine	V	99.068	99.133
Homoserine	HSe	101.047	101.105
Pyroglutamic acid	-	111.032	111.100

Table 2.3 Mass values for amino acid residues (NH-CHR-CO)

2.5.0 Use of different ionisation modes in the analysis of proteins and peptides

The last 10 years have seen an explosion of growth in the application of mass spectrometry in the analysis of peptides and proteins. The advent of FAB as a means of analysing underivatised peptides initiated this response and lately further spurts of growth have been forthcoming with the development of matrix assisted laser desorption and electrospray as modes of ionisation suitable for the analysis of intact proteins. In the following sections the application of the major modes of ionisation in this field are discussed.

2.5.1 Electron Impact

For many years EI provided the only means of analysis. Application of the technique was severely hindered by the need for extensive derivatisation of the peptide under investigation and also by the low mass range of early instrumentation of around 1000 Da (Morris, 1980). In practice hexapeptides were the largest peptides that could be studied. Sample sizes required are huge by modern standards, with up to 100 nmol being required due to the lengthy derivatisation procedures. Typically peptides are derivatised to form deuterated O-trimethylsilyl-polyamino alcohols (Kelley *et al*, 1975) or N-acetyl-N,O-permethylated peptides (Morris, 1972). Fragmentation tends to be rather extensive, but can be used to determine the amino acid sequence of the sample in favourable cases.

2.5.2 Chemical Ionisation

Due to the "softer" nature of this mode of ionisation compared with EI, more intense signals may be obtained for the pseudomolecular ion with less marked

fragmentation and thus complementary data is acquired to the EI data. Fragmentation tends to be at the amide bonds producing N-terminal acylium ions or C-terminal ammonium ions (Kiryushkin *et al*, 1971). Negative ion CI has also been adopted successfully in the analysis of peptide derivatives, often providing different and complementary fragmentation to that obtained under positive ion conditions (Hunt *et al*, 1976)

2.5.3 Field Desorption

For many years this was a popular technique for the analysis of underivatised peptides. The softness of the ionisation leads to little fragmentation and the predominant formation of the pseudomolecular ion. Between 0.1 - 2 nmoles of peptide are required for FD analysis. Tandem mass spectrometry has been employed to sequence peptides (Matsuo *et al*, 1981). The use of FD has been largely superceded by FAB due to the relative ease of use of the latter, however it has been noted that FD can be used to provide data for very hydrophobic peptides which are insoluble in the matrices used for FAB analysis (Stults, 1990).

2.5.4 Plasma Desorption

Plasma desorption has been widely used in the molecular weight determinations of peptides and proteins (Sundqvist *et al*, 1984; McFarlane, 1990), up to as large as 35kDa in the case of chymotrypsin (Craig *et al*, 1987). It is a relatively soft method although a degree of fragmentation is evident for low molecular weight samples. This fragmentation rarely provides much in the way of interpretable sequence related data. One application of the technique is in the molecular weight

mapping of peptides derived from an intact protein by enzymatic digestion (see section 2.3.0), with this process having been carried out *in situ* on nitrocellulose coated films (Chait and Field, 1986). One drawback with PDMS is that the width of peaks obtained limits the accuracy of molecular weight determination. A figure of 0.1% is generally quoted as the degree of accuracy of molecular weight determination of peptides with sensitivity typically in the picomolar range (Mann, 1990).

2.5.5 Laser Desorption

The desorption of peptides by a sub microsecond burst of photons from a laser source has been used to a limited degree as a mode of analysis. Peptides of up to 1-2kDa have been reported as amenable to the production of pseudomolecular ions (often cationated), although in general little interpretable structural fragmentation is observed (McCrery *et al*, 1982; Wilkins *et al*, 1985). The importance of laser desorption has dramatically increased in the field of protein/peptide analysis due to the advent of the matrix assisted methodology (Karas and Hillenkamp, 1988). Using photons at 266nm from a Neodynium/Yttrium-Aluminium garnet laser, with cinnaminic acid derivatives as matrices, for example, molecular weight data for a number of proteins have been established (Hillenkamp and Karas, 1990). Catalase, having a molecular weight of 236kDa, is quoted as the highest monomeric protein analysed, with the largest ion detected being that of the trimeric subunit of urease at 275kDa. The mass accuracy is similar to that of plasma desorption, and again a figure of 0.1% is quoted (Mann, 1990). Sample consumption is also similar to PD, being in the low picomolar range.

2.5.6 Secondary Ion MS

SIMS has not been used to a large extent in the analysis of peptides due to the excessive fragmentation/damage produced. The use of SIMS in the analysis of amino acids and peptides has been reported (Benninghoven and Sichtermann, 1978), however very little structurally relevant fragmentation was observed.

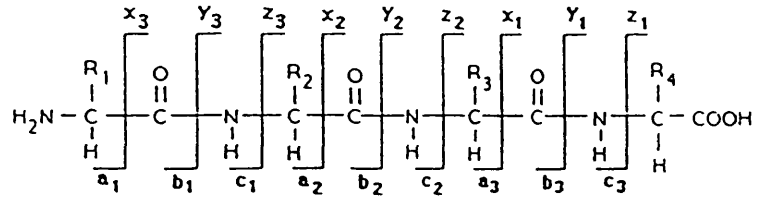
2.5.7.0 Fast Atom Bombardment

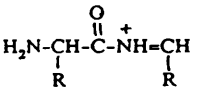
FAB has been probably the most widely used ionisation method for the analysis of peptides and small proteins. The technique generally affords intense pseudomolecular species (protonated and/or cationated) with sample requirements generally in the 0.1-1 nmol range. Matrices adopted include glycerol and thioglycerol, which work well for more hydrophilic samples and meta-nitrobenzyl alcohol which works well with hydrophobic peptides (Gower, 1985; Meili and Seibl, 1984). Strong acids are often added to these matrices to enhance sensitivities (Sheia and Sunner, 1990). In the analysis of mixtures of peptides there is a well documented suppression effect of hydrophilic species by hydrophobic peptides, which limits the usefulness of the "FAB mapping" strategy (Morris *et al*, 1983) in which a small protein is digested (usually with trypsin or cyanogen bromide) and molecular weight data for the resulting smaller peptides obtained without the use of chromatographic separation techniques. Modest amounts of salt contamination can be tolerated in FAB analysis, although in some cases this can give rise to spreading the pseudomolecular ion current over a number of peaks (eg $[M+Na]^+$, $[M+K]^+$ etc.). The FAB spectra of a number of small proteins up to 25kDa have been determined with accuracies in the order of 0.01% (Barber and Green, 1987). To enhance sensitivities for

hydrophilic peptides various derivatisation procedures have been advocated including methyl or hexyl esterification (Knapp, 1990; Falick and Maltby, 1989) and acetylation (Biemann, 1990).

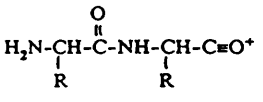
2.5.7.1 Sequencing of peptides by FAB

At present FAB provides the best method for sequencing peptides of the various ionisation methods. The original theoretically derived fragmentation scheme (Roepstorff and Fohlmann, 1984) has been modified in the light of experience (Johnson *et al*, 1987) and is depicted in figure 2.2, which also includes fragment ions due to side chain losses. Amino acid derived immonium ions and "internal" ions consisting of two or more amino acids are also noted due to their being observed experimentally. An early paper on the application of FAB ionisation reported the determination of the sequence of bradykinin from the normal spectrum (Barber *et al*, 1981), however this is a rare occurence and in general few sequence ions are usually observed in peptide spectra. Far superior data is obtained by using a tandem mass spectrometric method (Biemann and Martin, 1987). The simplest of these (although not strictly a tandem method) is the B/E linked scan on a double focussing magnetic sector instrument, preferably with a collision cell in the first field free region. Good daughter ion spectra of peptides can be obtained in the CAD mode with a parent ion transmission of 50-80% (Heerma *et al*, 1983). The main disadvantage of the B/E analysis of peptides is the large parent ion acceptance window, which can result in matrix ions being passed through. MIKES scanning on reverse geometry instruments has also been used to generate daughter ion spectra with a collision cell in the second field free region (Tomer *et al*, 1984). Hybrid three sector instrumentation has been

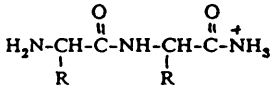




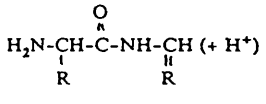
(a)



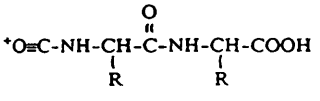
(b)



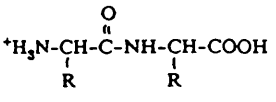
(c)



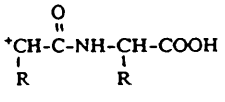
(d)



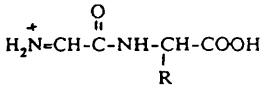
(x)



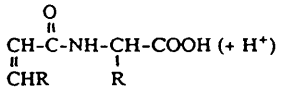
(y)



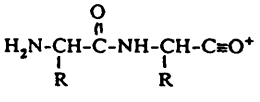
(z)



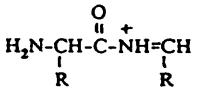
(v)



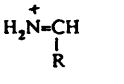
(w)



internal acyl



internal immonium



amino acid immonium

Figure 2.2 Peptide fragment ions observed using FAB ionisation

adopted in peptide analysis, with a significant improvement noted in resolution and accuracy compared with the two sector approach (Yost and Boyd, 1990). Triple quadrupole tandem mass spectrometry has been utilised for peptide sequencing (Hunt *et al*, 1985). The lower energy of the parent ions of around 100 eV generally results in less fragmentation than with the high energy collisions encountered with magnetic sector instruments (Hayes and Gross, 1990). A disadvantage of this is that *d*, *v*, and *w* series ions are not formed and so leucine and isoleucine cannot be distinguished. Also peptides above a molecular weight of 2,000 Da do not appear to fragment under the low energy conditions. The highest quality daughter ion spectra are obtained with four sector instrumentation (Poulter and Taylor, 1989). Using a BEEB configuration instrument Biemann and co-workers have succeeded in the complete sequencing of glutaredoxin from rabbit bone marrow (Hopper *et al*, 1989) and a number of thioredoxins (Johnson and Biemann, 1987). A present limitation in high energy CID sequencing is that peptides of larger molecular weight than around 2500 do not produce significant fragmentation (Hayes and Gross, 1990).

It has been suggested that basic amino acid residues are the major sites of protonation and that these direct the observed fragmentation (Bieman, 1990). To support this contention peptides containing N-terminal lysine or arginine were shown to be dominated by the N-terminal series *a,b,c,d* ions, whilst the presence of a C-terminal lysine or arginine gave rise to predominantly C-terminal *x,y,z,v,w* ions. By substituting specific amino acids in a series of enkephalins it has also been shown that the intensities of conserved ions in the CID daughter ion spectra were very similar unless arginine or lysine were replaced, suggesting that these basic amino acids direct fragmentation (Dass and Desiderio, 1987). The formation of charged derivatives at

the N- or C-terminus exhibit the same directing effects (Johnson *et al*, 1988; Wagner *et al*, 1991). A practical consequence of these observations is the recommendation that trypsin be used as the enzyme of choice in proteolytic digest, as the majority of the resulting peptides will contain C-terminal lysine or arginine . An interesting study has been carried out on a synthetic peptide containing a number of regularly spaced lysine residues. The resulting daughter ion spectrum was largely unintelligible, however when these basic residues were acetylated to produce far less basic moieties the CID spectrum was easily interpreted (Biemann, 1990).

2.5.8 Electrospray

Electrospray is ideally suited for the molecular weight determination of peptides and proteins due to the feasibility of their producing multi-charged species. The impact on the biochemical community has been quite marked, and as the great superiority of an electrospray molecular weight determination, having an accuracy of about 0.01%, compared with the 5% achievable with SDS-PAGE (Carr *et al*, 1991) becomes more keenly appreciated it is expected that this impact will further intensify. An accuracy of 0.01% translates into being able to determine the molecular weight of a 100kDa protein to within 10 Da. This is sufficient to enable the verification of the primary structure of a sequenced protein or the detection of any post translational modifications. The amount of sample required is in the low picomolar range. Since its emergence in the field of protein chemistry electrospray has been used in a range of applications such as the characterisation of recombinant proteins (Van Dorrselaer *et al*, 1990), detection of mutant proteins (Green *et al*, 1990) and the determination of protein "ragged ends" (Witkowska *et al*, 1990)

2.6.0 Compatibility of sample preparation with mass spectrometric analysis

It has been said that "the conventional protocols of any biochemistry and molecular biology research laboratory are replete with practices and problems that can thwart the best MS laboratory" (Burlingame *et al*, 1990) A rather sobering thought which emphasises the need for the consideration of mass spectrometric requirements in the initial stages of protocol design. Typically encountered problem areas are briefly outlined in the following paragraphs with possible remedial steps noted.

(i) The use of salt solutions, for example in the preparation of buffers or in ion exchange chromatography can result in an unacceptably high concentration of salt(s) in the sample. This problem is often exacerbated if an evaporation process is used to increase sample concentration. The simplest remedies are to use volatile species such as ammonium bicarbonate or N-ethylmorpholine in making up buffers and to arrange the order of chromatographic steps, if possible, such that ion exchange is not the final method used. Small Sephadex size exclusion columns (eg PD10 of Pharmacia) can be used to clean up salt contaminated samples, however their use is limited to samples of molecular weight greater than 2000 Da. The use of 18-crown-6 ether as a matrix additive has been reported to largely reduce peptide cation formation in liquid SIMS analysis (Orlando, 1992)

(ii) The use of detergents or surfactants (eg SDS, Tween, Triton or polyethylene glycol) even in seemingly low concentrations can create difficulties.

(iii) The use of strong acids or chlorinated solvents with plastic apparatus will often leach out plasticiser(s), such as octyl phthalates, to some degree. Unfortunately octyl phthalates are highly surface active and run readily under FAB conditions.

(iv) Scrapings of tlc plates can give rise to regions of the mass spectrum being

dominated by silica derived ions. Selective solvation of the analyte followed by centrifugation and analysis of the supernatant may help to minimise this problem. With high sample loadings a procedure for removing the analyte from the tlc plate surface using a probe tip covered with adhesive tape has been described (Chang *et al*, 1984).

(v) One of the most effective means of sample purification is the use of reverse phase HPLC with volatile buffers and where present in a purification protocol it should be adopted as the final step if possible.

(vi) The use of vacuum centrifugation can cause problems, if proteins or peptides are evaporated to dryness irreversible binding to plastic tubes can occur, especially at the temperatures of 30-40°C which are typically encountered with this apparatus. Lyophilisation is usually a more efficient method of sample concentration.

CHAPTER THREE

STRUCTURAL ELUCIDATION OF ANNEXIN AND PLASTOCYANIN SPECIES

3.0 Introduction

In this section two studies are described in which mass spectrometry has been utilised in an integrated approach with more conventional techniques in the determination of the primary structures of endogenous proteins. In the first study the structural elucidation of two isoforms of annexin V present in bovine brain is discussed from first principles and a striking vindication of accurate molecular mass determination of proteins noted as opposed to SDS-PAGE measurements. The second study concerns the confirmation of plastocyanin sequences from two algal sources, for which the primary structures were already largely known.

3.1.0 Elucidation of the primary structures of isoforms of bovine annexin V

3.1.1 Introduction

Calcium plays a major role in cellular regulation and is mediated by a range of calcium binding proteins. Many belong to the superfamily of EF-hand proteins, exemplified by calmodulin, troponin C, parvalbumin and proteins of the S-100 family (Kligman, 1988). More recently, a new family of calcium/phospholipid and membrane binding proteins has been identified (Klee *et al*, 1988; Crompton *et al*, 1988). Although these proteins contain repeated domain structures, an EF hand is not predicted from sequence data. They are characterised by the so-called endonexin fold - a highly conserved cluster of 17 amino acids which is repeated several times in individual sequences. This family has been named "annexins" (Crumpton and Dedman, 1990) . There are many members of this group (annexins I to VIII) and while multiple functions have been proposed including inhibition of phospholipase A_2

and blood coagulation, as well as a role in cell growth regulation, no biological role for any isoform has yet been clearly defined. The distinct differences in tissue distribution suggest specialised function(s) for the individual proteins.

The purification of two calcium-dependent phospholipid- and membrane-binding proteins from bovine brain has recently been described (Donato *et al*, 1990). They co-purify on phenyl-Sepharose in the presence of calcium but can be separated on DEAE-Sepharose. By SDS PAGE they migrate at 33 and 37 kDa, respectively-hence the names CaBP33 (*Ca*lcium *B*inding *P*rotein *33*kDa) and CaBP37 have been adopted. The two proteins are immunologically related to each other but immunologically distinct from other calcium-dependent phospholipid- and membrane-binding proteins (Donato *et al*, 1990). Other proteins have also been observed of this molecular weight range from bovine tissue that co-purify, including intestinal mucosa, lung, and adrenal medulla (Pepinsky *et al*,1988). It has been thought that the lower molecular weight form might be a proteolysis product of the larger protein (Bonstead *et al*, 1988). Recently a 32 kDa doublet was purified from bovine brain by calcium/phospholipid-dependent affinity chromatography (Bazzi and Neslestuen, 1991). This doublet comigrates with CaBP33/CaBP37 mixture in SDS gels and Western blot analyses revealed strong immunological cross-reactivity between this 32 kDa doublet and the CaBP33/CaBP37 mixture (Prof R.Donato, unpublished results).

3.1.2 Experimental

The proteins were provided by Prof R.Donato (University of Perugia, Italy) having been isolated from bovine brain by a combination of anion-exchange chromatography steps according to a method previously described (Donato, 1990).

Homogeneity was confirmed by the absence of other proteins on analysis of a heavy loading of protein by SDS PAGE. Proteins were run on 12.5% polyacrylamide gels containing SDS (Laemmli, 1970) and visualised by Coomassie Blue staining.

Sequence analysis was performed on peptides from cyanogen bromide (CNBr), tryptic and BNPS-skatole digests. Some CNBr peptides were subdigested by Arg-C proteinase and trypsin (see section 2.3.0 for details on the specificity of cleaving agents). Samples for cleavage with CNBr were reduced with 0.2M 2-mercaptoethanol (2h at 30°C), lyophilised and redissolved in 50% v/v formic acid containing CNBr (approximately 100 fold excess over methionine residues). The samples were incubated under nitrogen for 24h at 4°C in the dark after which the solution was diluted to 5% formic acid and lyophilised. Trypsin digestion was carried out at a ratio of 1:30 protease:substrate in 0.1M N-ethylmorpholine acetate buffer, pH 8.5 for 2h at 37°C. Arg-C digestion was carried out with 0.5 units of protease and 0.5 to 2 nmol substrate in 0.2M ammonium bicarbonate at 37°C for 2 hrs. BNPS-skatole digestion was carried out in the presence of a 50-fold excess of reagent in 50% v/v glacial acetic acid for 4h at 47°C and the fragments separated by SDS PAGE followed by electroblotting onto Problott (PVDF). Bands were cut from the membrane and transferred to the protein sequencer.

Peptides from CNBr cleavage were redissolved in HPLC buffer A (0.1% aqueous TFA and insoluble peptides and fragments removed by high speed bench centrifugation. These were later separated by SDS PAGE and electroblotted onto Problott (PVDF). The soluble peptides were purified by reverse phase HPLC using a water / acetonitrile gradient (0 to 50% in 0.1% v/v TFA) over a period of 55 min.

on a Vydac reverse-phase C_{18} column (25 x 0.46cm). Peaks (A_{215}) were collected, concentrated and the peptides sequenced on Applied Biosystems 470A gas-phase and 477A pulsed liquid-phase protein sequencers using recommended reagents with 120A on-line phenylthiohydantoin (PTH)-amino acid analyzers.

Peptides from BNPS-skatole and CNBr digests were separated by SDS PAGE and transferred onto PVDF membrane using 10% methanol / 10mM CAPS in a semi-dry blotter (Novablot, Pharmacia) with a current of 0.8 mA/cm^2 (Matsuidara, 1987). After blotting, peptides were visualised with Coomassie blue, cut out and transferred to the protein sequencer.

Structural analysis of N-terminal peptides was performed by liquid SIMS. Candidate peptides for N-terminal analysis which did not yield sequence data by automated micro-sequencing were air dried to 1-2μl on the stainless steel probe and mixed with 1μl of thioglycerol matrix containing 1% v/v TFA. Spectra were recorded using multi-channel analysis in the positive ion mode on a VG 70-250 SE mass spectrometer at an accelerating voltage of 8kV using a caesium ion gun operated at 20kV. The B/E linked scanning mode of the mass spectrometer was utilised in the elucidation of the N-terminal tryptic peptide sequence. Parent ions produced by liquid SIMS ionisation of 2 nmol of peptide dissolved in 2μl of glycerol/1% TFA matrix were collisionally activated in the first field free region of the mass spectrometer using helium gas and at a pressure which gave 80% transmission for m/z 393 of caesium iodide.

Electrospray mass spectrometry was performed on a prototype VG Bio-Q mass spectrometer. The protein samples were gel-filtered, to remove low molecular weight contaminants such as inorganic salts, using disposable PD10 Sephadex columns

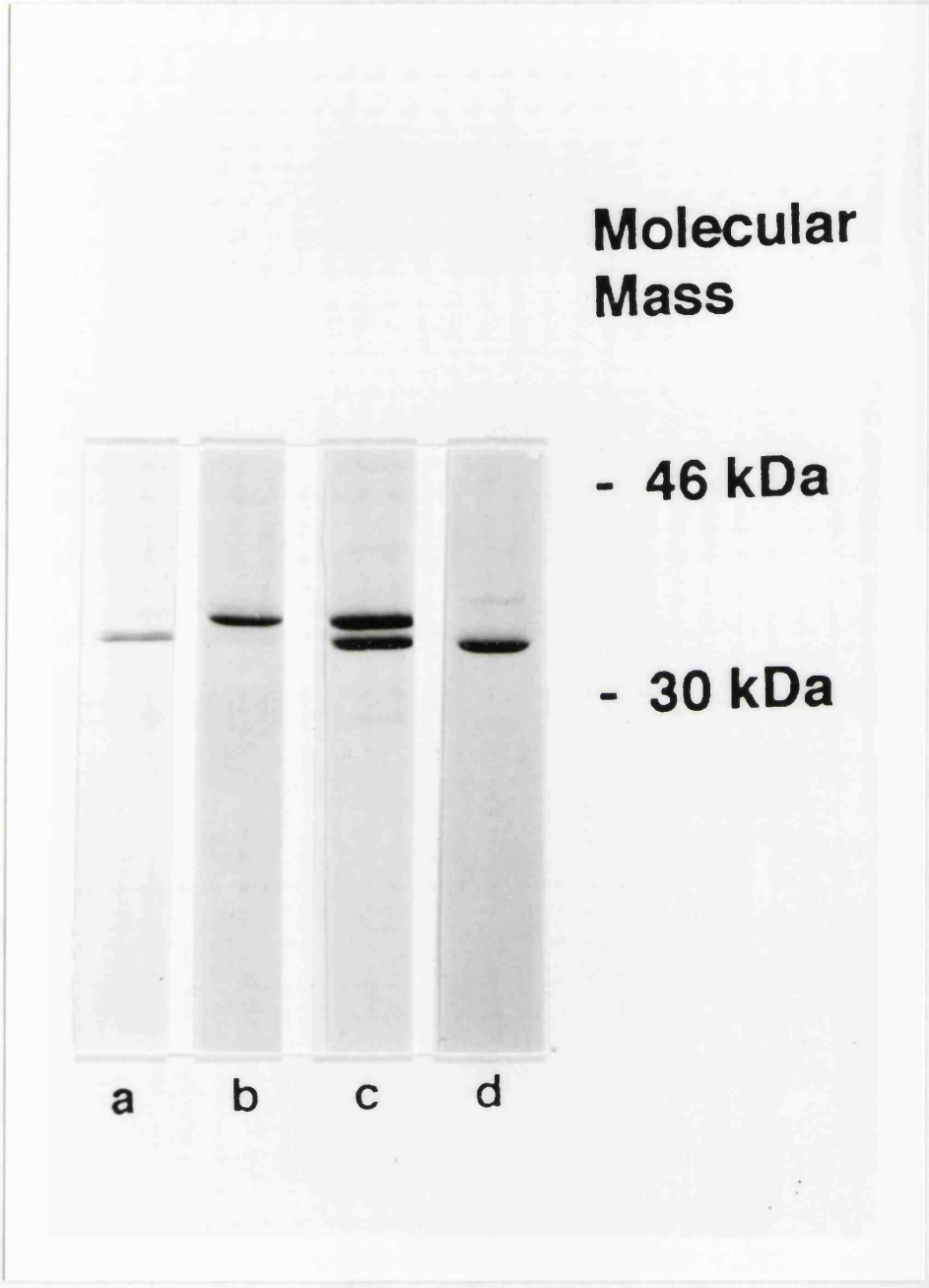


Figure 3.1 12.5% SDS PAGE analysis of (a) CaBP33 (b) CaBP37 (c) CaBP33/CaBP37 mixture (d) rat annexin V

(Pharmacia). Samples (0.2-10μg) were introduced intothe source in a solution of aqueous methanol (50%) containing 1-25% acetic acid at a concentration of 1-50pmol/μl. The sample was delivered to the nebulizer head of the electrospray source via a 20μl Rheodyne injection valve and pumped with an LKB syringe pump at 2μl/min. A potential difference of 3 kV was maintained between the capillary tip and the circular electrode, which was at a distance of 2cm. Data were acquired in the multichannel analysis mode of the data system over a mass range of 700-1500 Da. The masses were calibrated from the multiply charged ions due to a separately introduced sample of horse heart myoglobin (average M_r 16,950.5).

Matrix assisted laser desorption spectrometry was carried out on a Finnegan-MAT Lasermat mass spectrometer coupled with a UV nitrogen laser. Approximately 10 pmol of protein were analysed using sinapinic acid as the matrix.

3.1.3 Results and Discussion

Figure 3.1 shows the SDS PAGE analysis of CaBP33 and CaBP37, together with a sample of rat annexin V. A difference in molecular weight of approx 4kDa is evident between the two proteins whilst rat annexin V appears to have a similar molecular weight to CaBP37.

Samples of each protein were subjected to CNBr degradation, peptides that were soluble in 0.1% aqueous TFA were subjected to reverse phase HPLC analysis and the insoluble fragments were separated on SDS PAGE, then electroblotted onto PVDF, prior to sequence analysis. Samples of each protein were also subjected to cleavage at the single tryptophan residue to give two fragments which were separated by SDS PAGE and electroblotted for gas phase Edman sequencing.

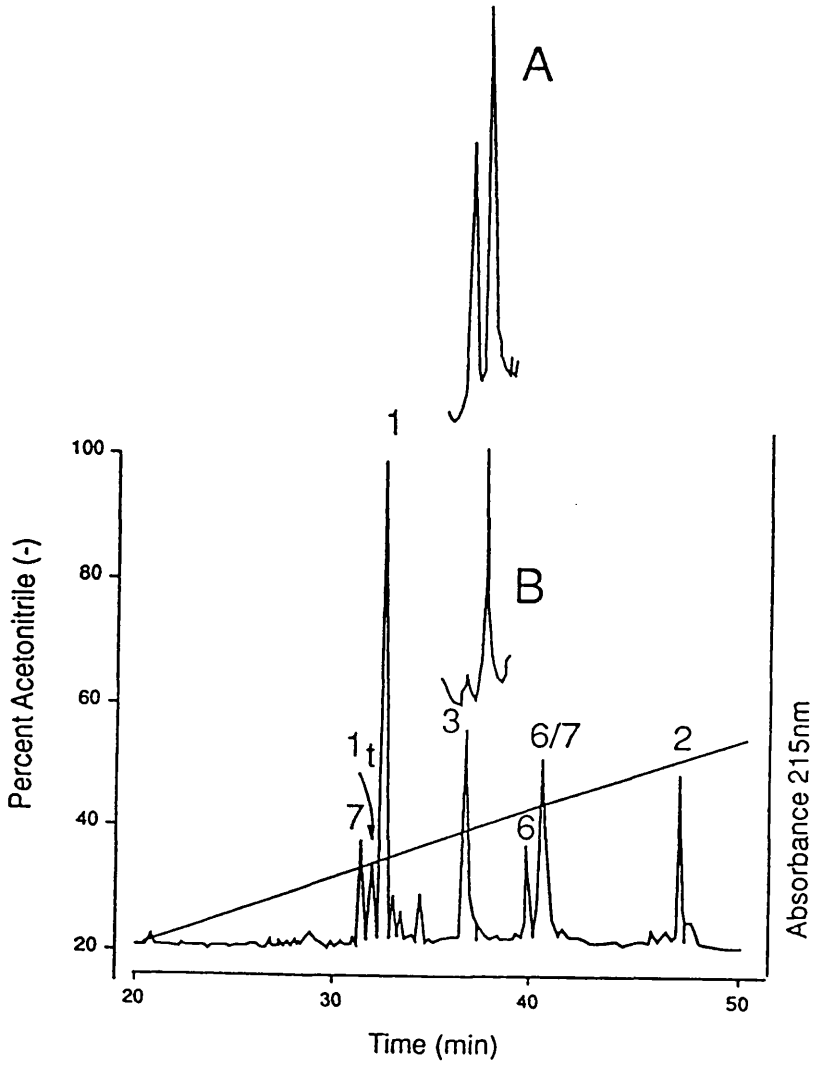


Figure 3.2 HPLC profile of CNBr digest of CaBP33. Differences shown inset for (A) CaBP33/37 mixture (B) CaBP37

The insoluble peptides were found to consist mainly of peptides CN4 and CN5, and of incompletely digested fragments. Cleavage of the Met_{213}-Thr_{214} bond with CNBr was low, as is commonly encountered (Aitken, 1984) therefore fragments with the amino terminus of peptide CN5 were in low yield.

When the CNBr peptides from each protein were separated by reverse phase HPLC, an almost identical pattern was seen for both (see Figure 3.2) with a shift in only one peptide, CN3. Digest mixtures of both proteins contain a minor peak designated CN1t. Subdigestions of certain CNBrpeptides were carried out with trypsin and Arg-C proteinase, the HPLC profiles of the arg-C digested N-terminal peptides (CN1) are shown in Figure 3.3. Sequence analysis of intact and subdigested CN3 from both proteins revealed one difference at position 125 i.e. Lys in CaBP33 and Glu in CaBP37. Analysis of peptide CN2 revealed one difference at position 36, serine in CaBP33 and threonine in CaBP37, which is a conservative substitution.

The molecular weight difference between the two proteins observed by SDS PAGE analysis could be explained by a post translational modification, such as glycosylation. To investigate this possibility the proteins were subjected to electrospray mass spectrometry (see Figure 3.4). Despite the apparent molecular weight difference of 4kDa on SDS PAGE, the results indicated that the two proteins differ by approximately 9Da, CaBP33 having an experimentally determined mass of 35, 996.2 $\pm$ 5.5Da and CaBP37 having a mass of 36,004.9 $\pm$ 3.2 Da. This surprising data suggests that the structures of the two isoforms are very similar and that the 4kDa difference noted by SDS-PAGE analysis is entirely misleading.

Determination of the structures of the N-termini was the only remaining problem. Amino acid analysis of the N-terminal peptide CN1 from a CaBP33/37

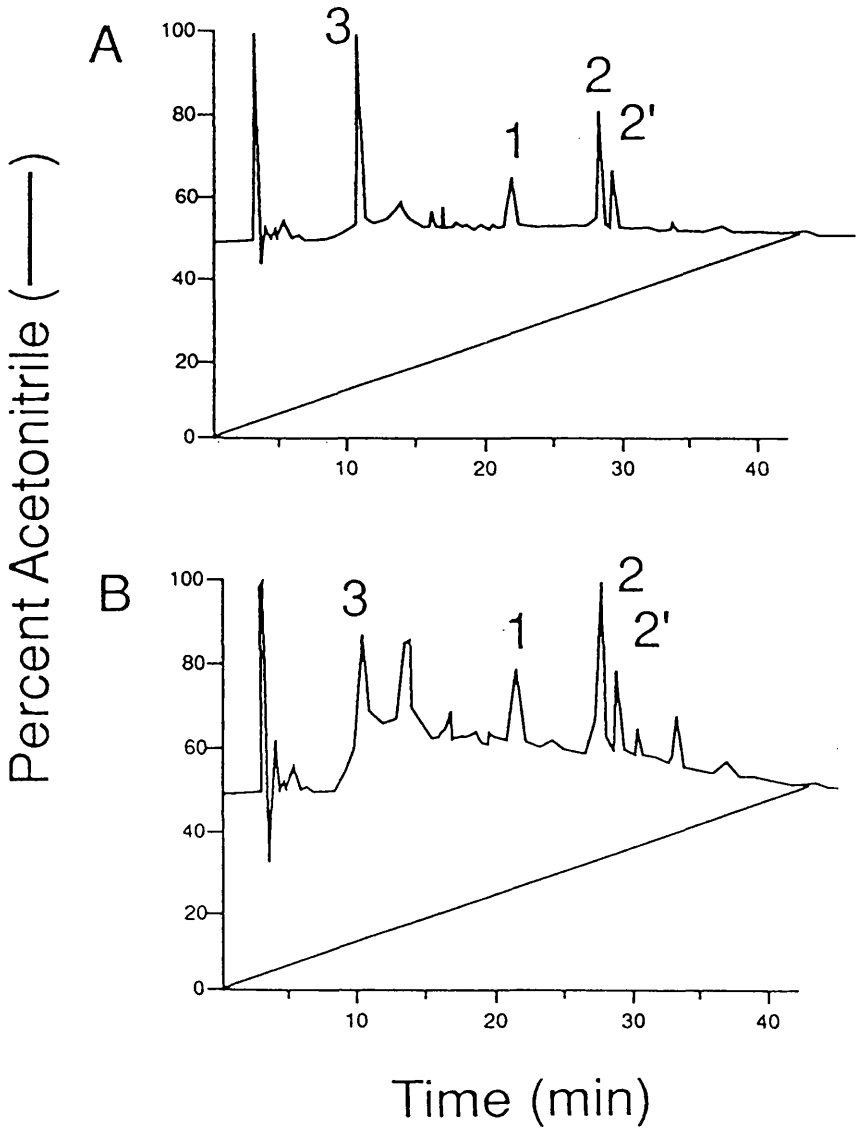


Figure 3.3 HPLC profile of arg-C subdigested CNBr fragment CN1 for (A) CaBP33 (B) CaBP37 (numbering of peaks as in figures 3.7 and 3.8)

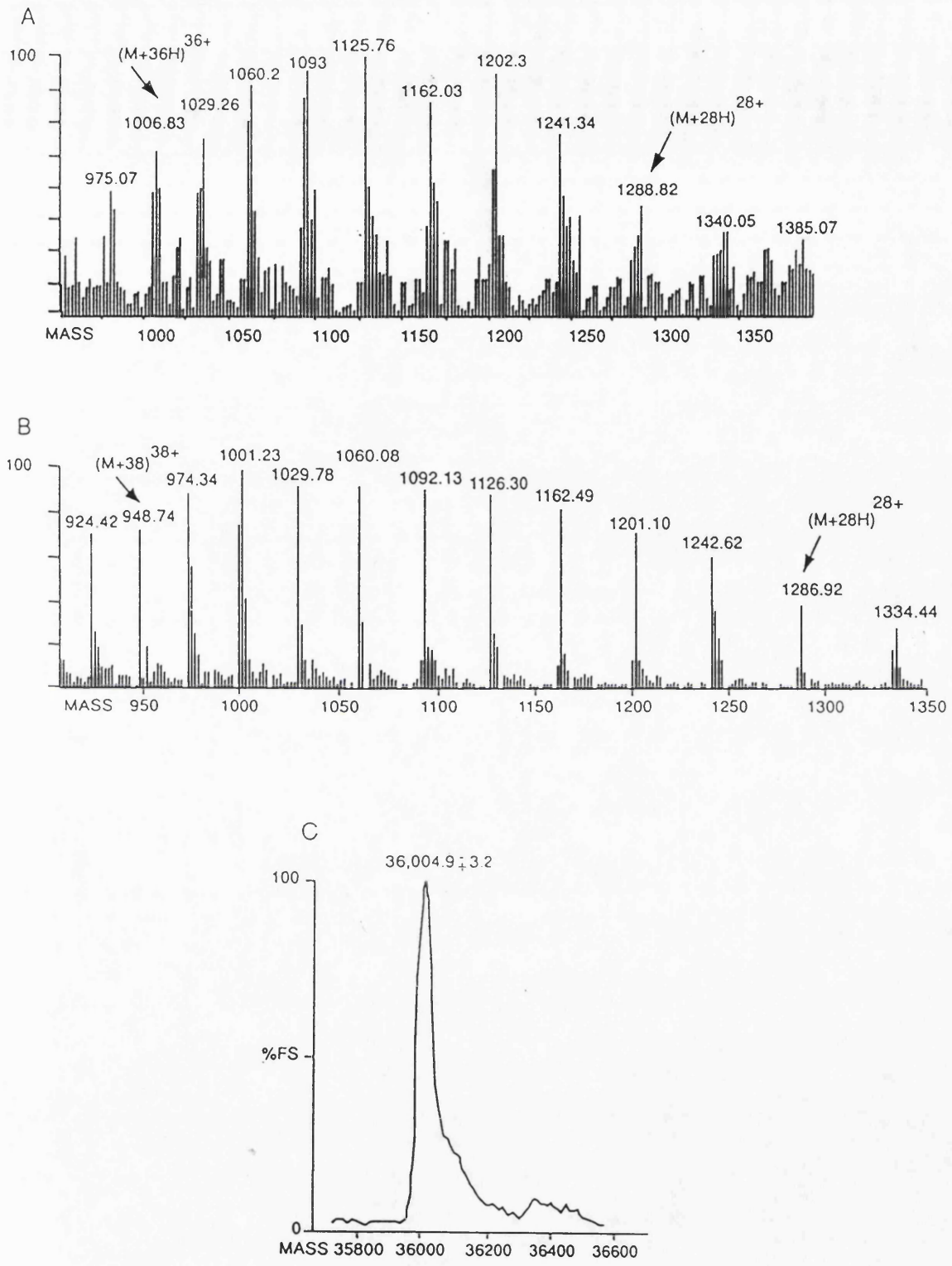


Figure 3.4 Electrospray spectra of (A) CaBP37 (B) CaBP33 (C) deconvoluted spectrum of CaBP37

mixture gave the composition $Asx_{3.3}$, $Glx_{3.1}$, $Gly_{2.2}$, $Arg_{2.9}$, $Thr_{2.0}$, $Ala_{4.7}$, $Pro_{1.0}$, $Val_{1.8}$, $Leu_{1.8}$, $Phe_{1.9}$, $Lys_{1.0}$, $Met_{1.0}$ (in the form of homoserine). Peptide CN1 from each of the two proteins was subjected to liquid SIMS mass spectrometry, each gave a molecular ion of $(M + H)^+ = 2960$ as shown in figure 3.5(a). This molecular weight is 42 Da higher than that calculated for the above amino acid composition, suggesting that both CaBP33 and CaBP37 were N-acetylated. The N-terminally blocked peptide CN1t was also analysed (see figure 3.5b) and gave a molecular ion of 2990, corresponding to the Ala to Thr substitution in CN1 noted in the sequencing of trypsin subdigested CN1 (see figure 3.7).

By subtracting the amino acids present in the fragments of the arg-C digest of CN1, CN1R2-R4, the composition of the blocked N-terminal peptide is found to be Glx, Arg, Ala, Val and Leu. To determine the amino acid sequence a B/E linked scan was performed at a parent ion m/z of 628 corresponding to an acetylated peptide having the derived amino acid composition. Collisionally activated dissociation was effected using helium gas and the resulting spectrum is shown in figure 3.6(a).Due to the specificity of the arg-C protease used to effect cleavage, arginine must be the C-terminal residue. The spectrum was then searched for C-terminal ions predicted by assigning each of the remaining four amino acids at position 2 (numbering from the C-terminus). As can be seen from figure 3.6(b) leucine appears to be present at this position with the x_2,y_2 and z_2+1 ions noted in the spectrum, in contrast to the other amino acids where the predicted ions are not seen to the same extent. In the same fashion the amino acid at residue 3 is seen to be valine, asparagine is present at position two and N-acetylated alanine at the N-terminus. N-terminal acetylation is further evidenced by the y_5 ion at m/z 585. Having deduced the sequence the full

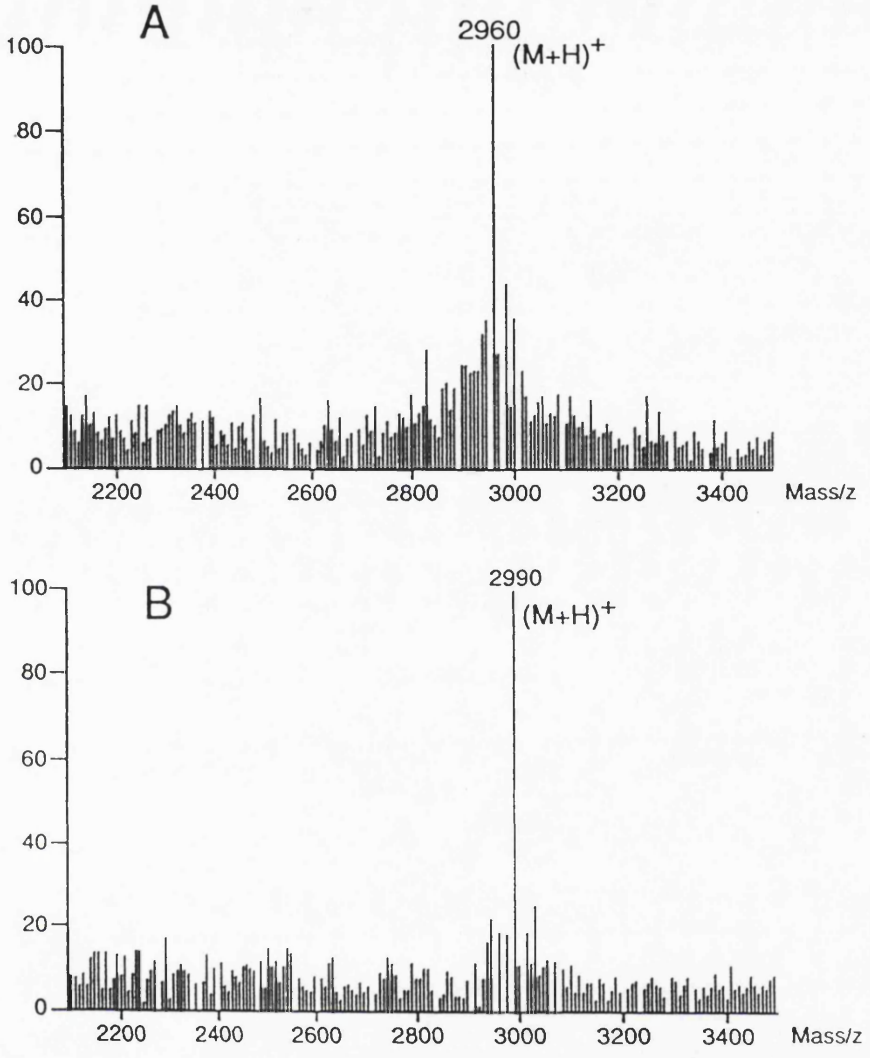


Figure 3.5 Liquid SIMS spectra of (A) N-terminal cyanogen bromide peptide (CN1) of a CaBP33/37 mixture (B) Peptide CN1t from CaBP33/37 containing an Ala to Thr substitution

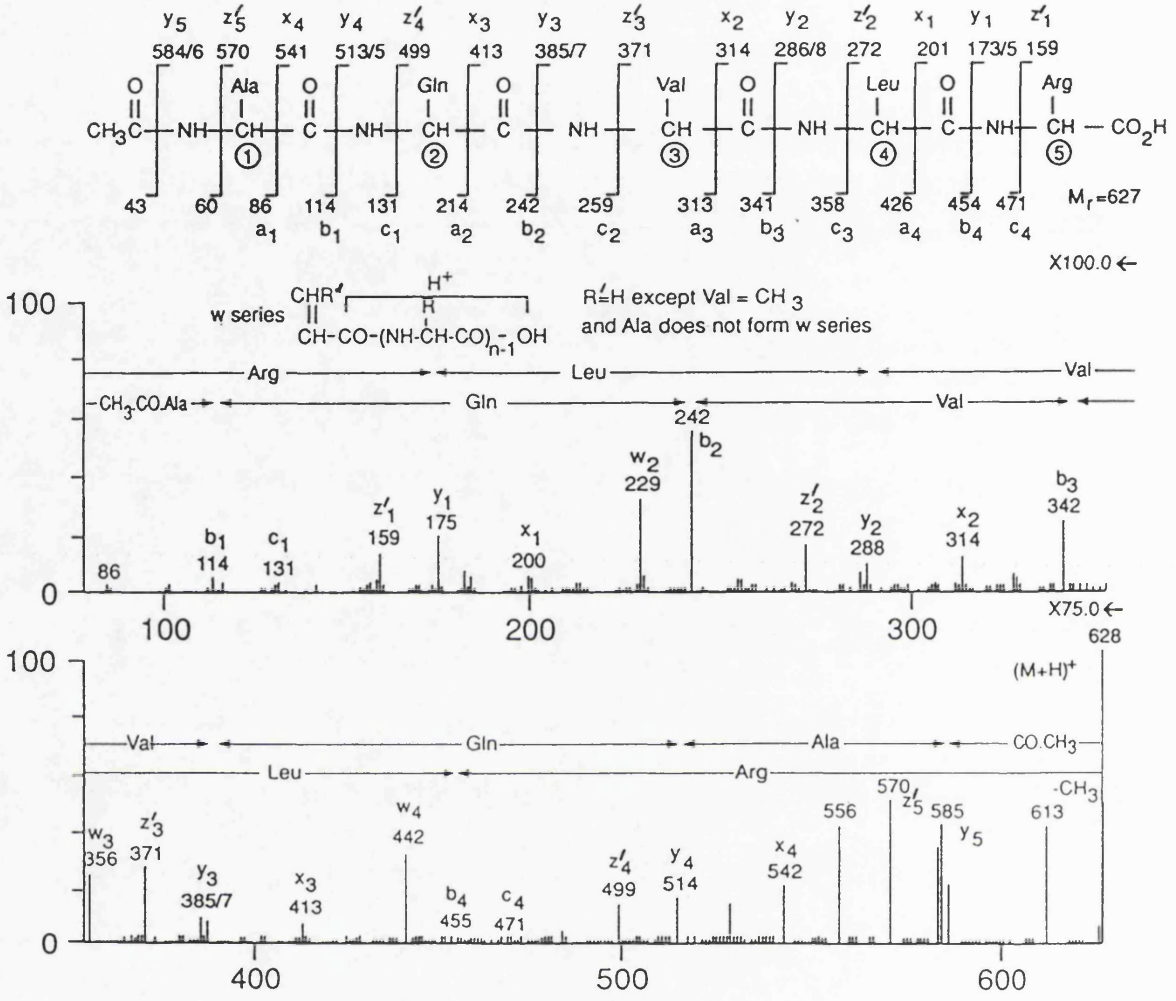


Figure 3.6 B/E linked scan liquid SIMS spectrum of fragment CN1T1 (N-terminus) of a CaBP33/37 mixture.

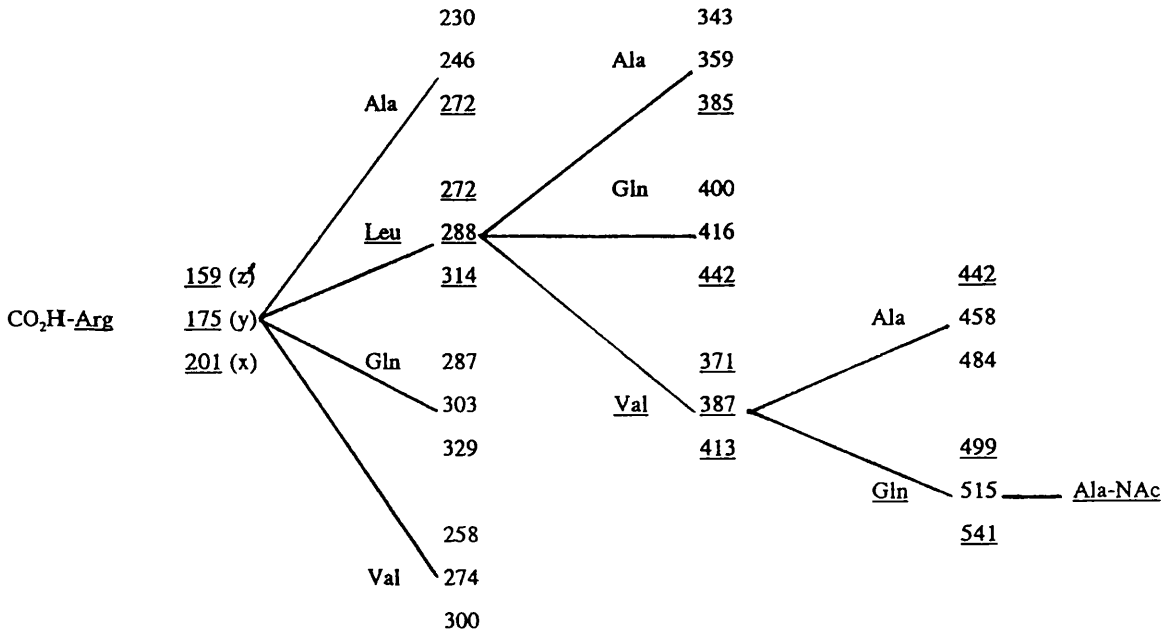


Figure 3.6b Deduction of N-terminal peptide sequence from known amino acid composition (m/z values observed in the mass spectrum are underlined)

suite of theoretical fragment ions, including the N-terminal fragments and those due to side chain loss, can be checked for in the spectrum. The N-terminal series is seen to be rather incomplete, as may be expected with a C-terminal arginine residue, however the complete sequence can be described from the *b* series of ions. A complete series of *w* ions is also seen (given that valine does not form this series). No *v* or *d* ions are seen, the latter observation not being surprising due to the absence of any basic amino acids other than the C-terminal arginine.

Figures 3.7 and 3.8 show a summary of the sequence data from CaBP33 and CaBP37 respectively. The calculated masses for these sequences are in excellent agreement with the experimental values obtained by electrospray noted earlier - CaBP33 has a calculated mass of 35,985 Da using average chemical weights (exp. = 35,996) with that of CaBP37 being 36,000 Da (exp = 36,005). A search of the European Molecular Biology Laboratory protein sequence database showed the proteins to be isoforms of annexin V. The sequence of CaBP33 is identical to that of bovine intestinal mucosa annexin V (Pepinsky *et al*, 1988) except for the presence of Tyr at position 281 in place of Phe. This could be due to the presence of allelic forms such as the situation seen here at position 9, or could be a tissue specific difference. CaBP33 and CaBP37 sequences are shown aligned with the rat and human sequences of annexin V in figure 3.9.

Annexin V was first isolated from human placenta and was independently termed endonexin II (Schlaepfer *et al*, 1987) and the placental anticoagulant protein 1 (Funakoshi *et al*, 1987). Many of the annexins are substrates for protein kinase C and protein tyrosine kinases (Weber *et al*, 1987).

The three dimensional structure of chicken annexin V (Dr D.Waller

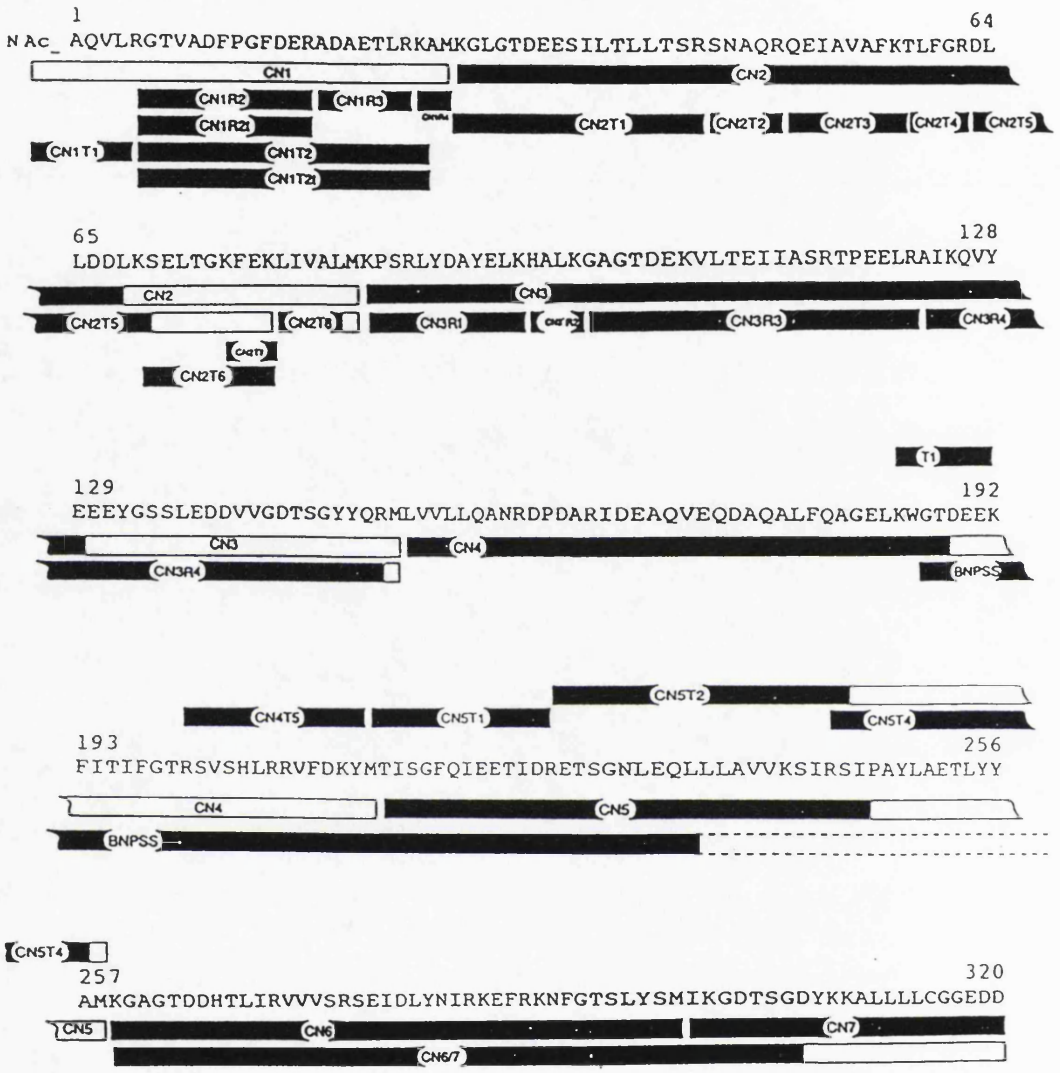


Figure 3.7 Sequence alignment of digested fragments of CaBP33 (CN=Cyanogen bromide, T=Trypsin, CNT=Trypsin subdigest of CN fragment, BNPSS=BNPS Skatole)

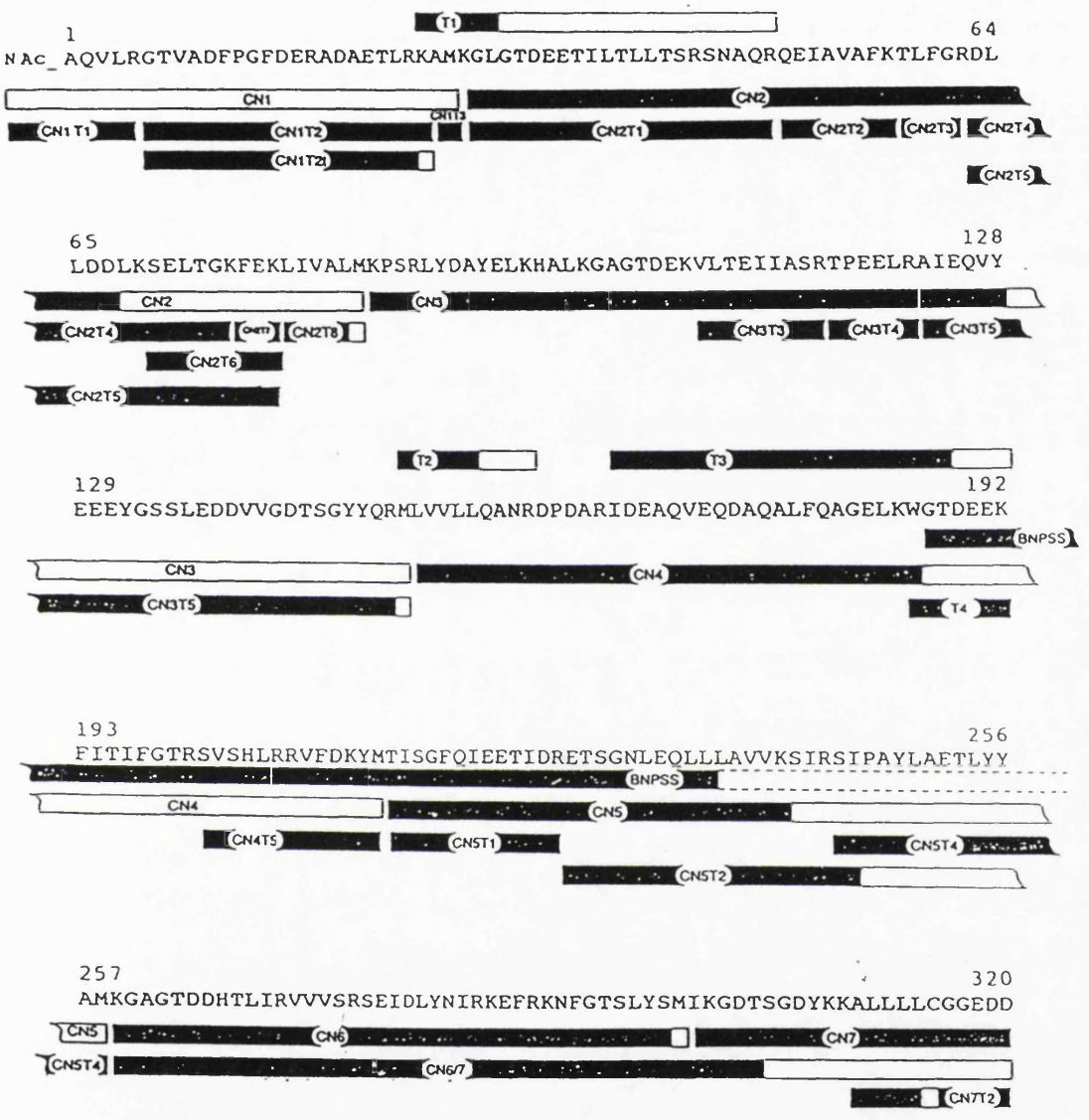


Figure 3.8 Sequence alignment of digested fragments of CaBP37 (CN=Cyanogen bromide, T=Trypsin, CNT=Trypsin subdigest of CN fragment, BNPSS=BNPS Skatole)

```
                1                                                              64
CaBP33  NAc.    AQVLRGTVADFPGFDERADAETLRKAMKGLGTDEESILTLLTSRSNAQRQEIAVAFKTLFGRDL
CaBP37                  t                          I
annex5human     ........T..........................................SA...........
annex5rat       .●●.....T..S...G.....V............D...N...A.......Q..EE.........

                65                                                            128
                LDDLKSELTGKFEKLIVALMKPSRLYDAYELKHALKGAGTDEKVLTEIIASRTPEELRAIKQVY
                                                                           .E.
                ........................................N.......................
                VN.M....................................D.....................A.

                129                                                           192
                EEEYGSSLEDDVVGDTSGYYQRMLVVLLQANRDPDARIDEAQVEQDAQALFQAGELKWGTDEEK

                ....................................G...........................
                ......N............................TA..D....L...................

                193                                                           256
                FITIFGTRSVSHLRRVFDKYMTISGFQIEETIDRETSGNLEQLLLAVVKSIRSIPAYLAETLYY

                ..............K.................................................
                ....L....................................N......................

                257                                                           320
                AMKGAGTDDHTLIRVVVSRSEIDLYNIRKEFRKNFGTSLYSMIKGDTSGDYKKALLLLCGGEDD

                ...............M........F..........A........................●EDD
                ...............I........F..........A............................
```

Figure 3.9 Sequences of CaBP33 and CaBP37 showing homology with annexin V

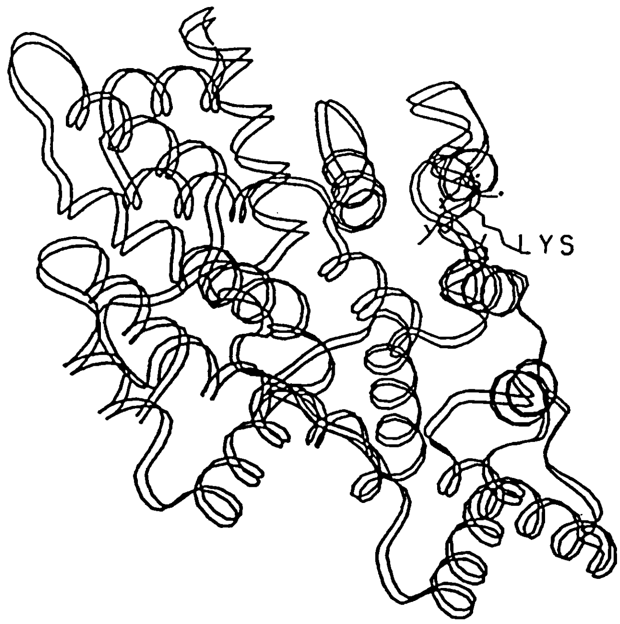


Figure 3.10 The position of Lys-125 in the ribbon diagram of chicken annexin V

unpublished results) which has a high homology with CaBP33/37 is depicted in figure 3.10 in which Lys-125 is seen to be located on the surface of the protein. In CaBP33 and CaBP37 Lys-125 and Glu-125 respectively will also reside on the protein exterior. This may have implications for any differing biological activities of the two isoforms.

Subsequent to determining the structures of these isoforms the matrix assisted laser desorption spectrum of CaBP37 was obtained as shown in figure 3.11, giving rise to a mass determination of 36,373 Da. This result is rather disappointing being an error of approximately 373 Da, corresponding to a 1% inaccuracy. It has been noted (Hillenkamp, 1991) that many proteins desorb with the formation of matrix adducts in this technique. An important question is thus raised as whether a sinapinic acid adduct(s) is contributing to the molecular weight value noted above. If the value of 36,373 represents the adduction of one sinapinic acid residue (M_r = 224 Da) per molecule of protein, this would give rise to a revised experimental molecular weight of 36,149 Da and an error of +149 Da on the true value. Two residues would represent an error of -75 Da. When the electrospray error of 5 Da is compared with these figures it is clear that in this particular problem the electrospray data is of much superior accuracy than the laser desorption analysis.

In conclusion it has been shown that bovine brain contains two isoforms of annexin V. Both proteins are N-terminally acetylated and they differ by 4000 Da apparent molecular weight on SDS PAGE but by only 15Da according to the amino acid sequence which is confirmed by electrospray mass spectrometry. This difference is due to the substitution of Lys_{125} and Ser_{36} in CaBP33 for Glu_{125} and Thr_{36} in CaBP37 respectively. (This is a difference of 15Da, within the error of ± 9 in the

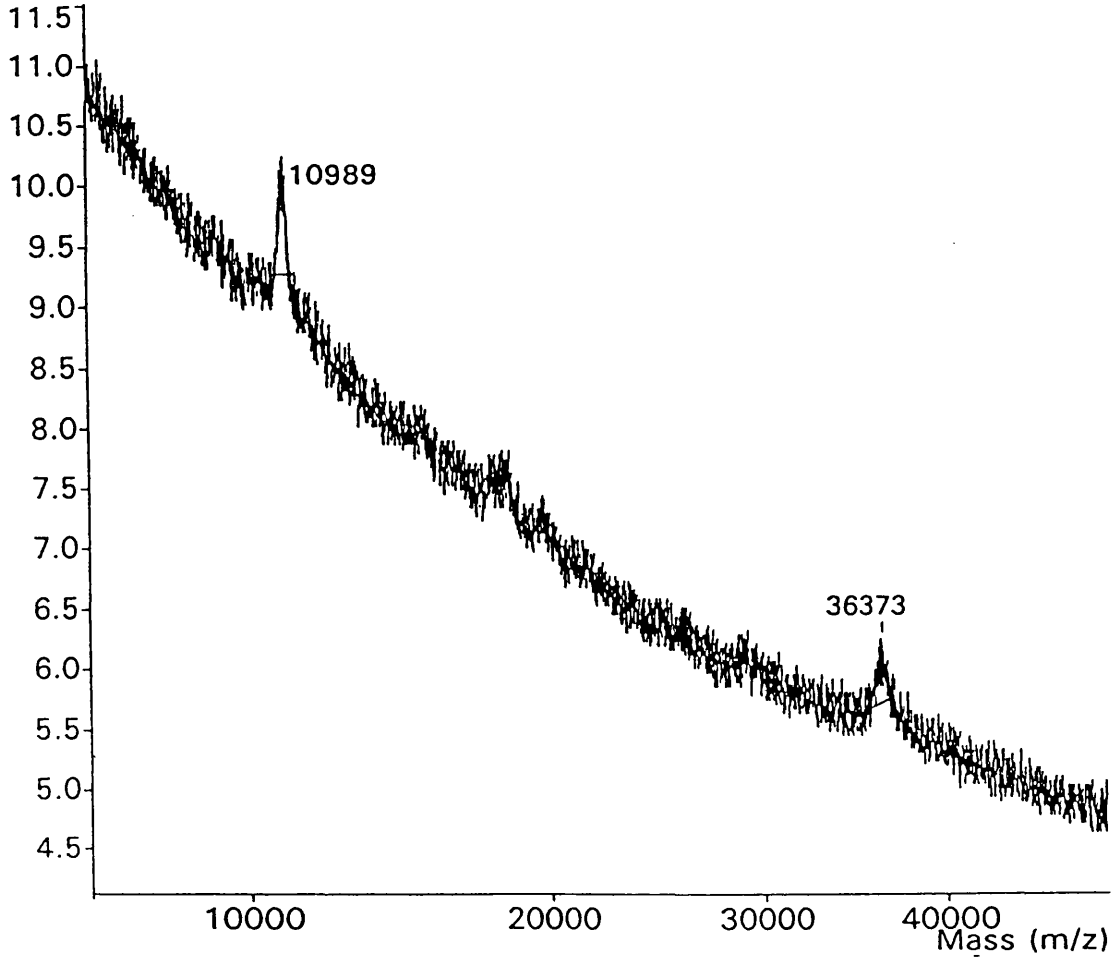


Figure 3.11 Laser desorption spectrum of CaBP33/37 mixture

mass spectrometry result).

A similar marked change in molecular weight due to a Lys → Glu substitution has also been seen in p21 *ras* proteins (Fasano *et al*, 1984). Mutant p21 *ras* proteins were also shown to exhibit a large shift in migration position on SDS PAGE due to a Glu to Lys substitution . The rat and human annexins V which migrate at about 32k on SDS PAGE have a Lys at position 125. In the cDNAs for annexin V in many species (Pepinsky *et al*, 1988; Funakoshi *et al*, 1987) $Lysine_{125}$ is coded by AAA. Therefore the sequence change seen for CaBP37 could be explained by a point mutation AAA → GAA (Glu).

3.2.0 Confirmation of the primary structures of plastocyanin from *Scenedesmus obliquus* and *Anabaena variabilis* cyanobacteria

3.2.1 Introduction

Plastocyanin is a metalloprotein of approximately 100 amino acids containing redox active copper in a tetrahedral ligand environment which gives rise to its intense blue colouration (λ_{max} = 597nm) when oxidised (Penfield *et al*, 1981). It is found to occur in higher plants, green algae and in cyanobacteria, whilst there is also evidence of its presence in other eukaryotic algae such as euglenoids (Aitken, 1988). The role of plastocyanin is to transfer electrons between cytochrome f and the P700 component of photosystem I in the oxygen evolving process of photosynthesis (Cramer *et al*, 1985). Sequence data has been determined for a number of species (Sykes, 1985) including the tertiary structure of poplar plastocyanin by x-ray crystallography (Freeman, 1983). The amino acid sequence of plastocyanin from the

```
                        5                   10                  15
A.v  Glu Thr Tyr Thr Val Lys Leu Gly Ser Asp Lys Gly Leu Leu Val
S.o          Ala Asn Val Lys Leu Gly Ala Asp Ser Gly Ala Leu Val

                        20                  25                  30
A.v  Phe Glu Pro Ala Lys Leu Thr Ile Lys Pro Gly Asp Thr Val Glu
S.o  Phe Glu Pro Ala Thr Val Thr Ile Lys Ala Gly Asp Ser Val Thr

                        35                  40                  45
A.v  Phe Leu Asn Asn Lys Val Pro Pro His Asn Val Val Phe Asp Ala
S.o  Trp Thr Asn Asn Ala Gly Phe Pro His Asn Ile Val Phe Asp Glu

                        50                  55                  60
A.v  Ala Leu Asn Pro Ala Lys Ser Ala Asp Leu Ala Lys Ser Leu Ser
S.o  Asp Ala Val Pro Ala  -   -  Gly Val Asn Ala Asp Ala Leu Ser

                        65                  70                  75
A.v  His Lys Gln Leu Leu Met Ser Pro Gly Gln Ser Thr Ser Thr Thr
S.o  His Asp Asp Tyr Leu Asn Ala Pro Gly Glu Ser Tyr Thr Ala Lys

                        80                  85                  90
A.v  Phe Pro Ala Asp Ala Pro Ala Gly Glu Tyr Thr Phe Tyr Cys Glu
S.o  Phe  -   -  Asp Thr  -  Ala Gly Glu Tyr Gly Tyr Phe Cys Glu

                        95                  100
A.v  Pro His Arg Gly Ala Gly Met Val Gly Lys Ile Thr Val Ala Gly
S.o  Pro His Gln Gly Ala Gly Met Val Gly Thr Ile Val Gln
```

Figure 3.12 Amino acid sequences of plastocyanins from *Anabaena variabilis* and *Scenedesmus obliquus* (with homology shown).

eukaryotic alga *Scenedesmus obliquus* has been determined and is detailed in figure 3.12 (R.P Ambler, unpublished results quoted in Sykes (1985)). An accurate determination of the molecular weight of the protein would confirm the proposed sequence. The amino acid sequence of the protein from the cyanobacterium *Anabaena variabilis* has previously been reported (Aitken, 1975) and is also shown in figure 3.12, confirmation of the sequence and of the absence of post-translational modifications was sought from the application of electrospray mass spectrometry.

3.2.2 Experimental

Plastocyanins, extracted from *A. variabilis* and *S. obliquus* according to standard procedures (Aitken, 1975) were a gift from Prof A.G.Sykes (University of Newcastle). The proteins had been eluted in 10 mmol phosphate buffer from a DEAE cellulose column in the final step of the purification procedure and a trace of potassium ferricyanide added to prevent reduction.

Because of the possible effects on the sensitivity of the electrospray measurements of the phosphate buffer, the proteins were purified by gel filtration through PD-10 sephadex columns (Pharmacia) into water. The columns were washed with a concentrated solution of potassium ferricyanide to oxidise the reducing agent "merthiolate" in which the PD-10 columns are stored. One blue coloured band was collected in each case, showing that the plastocyanin samples were in the oxidised forms (the reduced forms are colourless). After filtration the proteins were lyophilised prior to analysis.

Electrospray data were obtained on approximately 100 pmol of protein using a prototype VG Biotech Bio Q mass spectrometer operating at a nozzle voltage of

5000V. Proteins were dissolved in 10μl 1:1 methanol/water (v/v) and injected into a flowing 1:1 methanol/water stream containing 1% acetic acid, from an LKB syringe pump, entering the mass spectrometer source at a flow rate of 5μl/min. Both sources of plastocyanin were also analysed after dissolving in the previous solvent containing 5mM EDTA (Ethylene Diamine Tetra Acetic acid).

For further analytical work on the C-terminus of *S.obliquus* plastocyanin the protein was digested with cyanogen bromide to cleave at Met-97. Approximately 50μg of lyophilised protein was dissolved in 40% aqueous formic acid and cyanogen bromide added (100 fold excess over methionine residues). After flushing with nitrogen the digest was incubated in the dark for 24hrs at 4°C. The reaction was terminated by dilution to 5% formic acid and subsequently lyophilised. The sample was redissolved in 50μl water and centrifuged at high speed for 10 minutes, using a bench top microfuge, to precipitate out most of the larger of the two expected fragments. The supernatant was then used for further analysis.

Liquid SIMS mass spectra were obtained on a VG-ZAB-SE mass spectrometer operated at 8kV accelerating voltage and a caesium ion gun voltage of 35kV. For the high mass studies, spectra were acquired in the multichannel analysis mode under linear scan conditions with a scan time of 15s and a resolution of 1000. The low mass data were acquired over the mass range 1000-40 Da at a scan time of 4s/decade and with the cesium gun operated at 20kV. 10μl of the supernatant from the cyanogen bromide digest was air dried on the probe tip to which 1μl of thioglycerol had previously been applied as the matrix.

Edman sequencing of the remainder of the supernatant was carried out using an Applied Biosystems 470 gas phase automated sequencer.

3.2.3 Results and discussion

Under the FAB conditions employed (8keV Xe) no observable signals were generated for both sources of plastocyanin. In contrast under liquid SIMS conditions, with an incident beam of 35keV Cs+ ions, both proteins exhibited mass spectra. The spectrum of *S.obliquus* plastocyanin, depicted in figure 3.13(a), consists of four relatively sharp peaks at 10,102 Da, 10,146 Da, 10,181 Da and 10,252 Da superimposed upon an assymetric component which stretches over several hundred Daltons. The superimposition of relatively sharp pseudomolecular ions upon a broad assymetric background signal has been noted as a common feature of the liquid SIMS spectra of small proteins (Green, 1988). The expected mass of copper containing plastocyanin from the sequence in figure 3.12 is 10,042.4 Da, which is about 60 Da less than the measured value of 10,102 Da, representing a rather large error. With similar quality liquid SIMS data to that of figure 3.13(a) Green (1988) noted an experimental error of +11 Da for the molecular weight of lactoglobulin-A.

The spectrum obtained for plastocyanin from *A.variabilis*, shown in figure 3.13(b), was of rather poor quality consisting of a broad "hump" with no clearly defined peaks superimposed. Manual centroiding of this broad signal results in a mass measurement of 11,190 Da. This value is 23 Da above the theoretical value of 11,167.2 Da.

Due to the large errors observed for both proteins when using liquid SIMS more accurate mass measurement was sought from the application of electrospray ionisation. As can be seen in figure 3.14 the spectra obtained were of far superior resolution and signal to noise ratio than the liquid SIMS results. Different series of ions are evident in both spectra labelled A,B,C etc.. With *S.obliquus* the differences

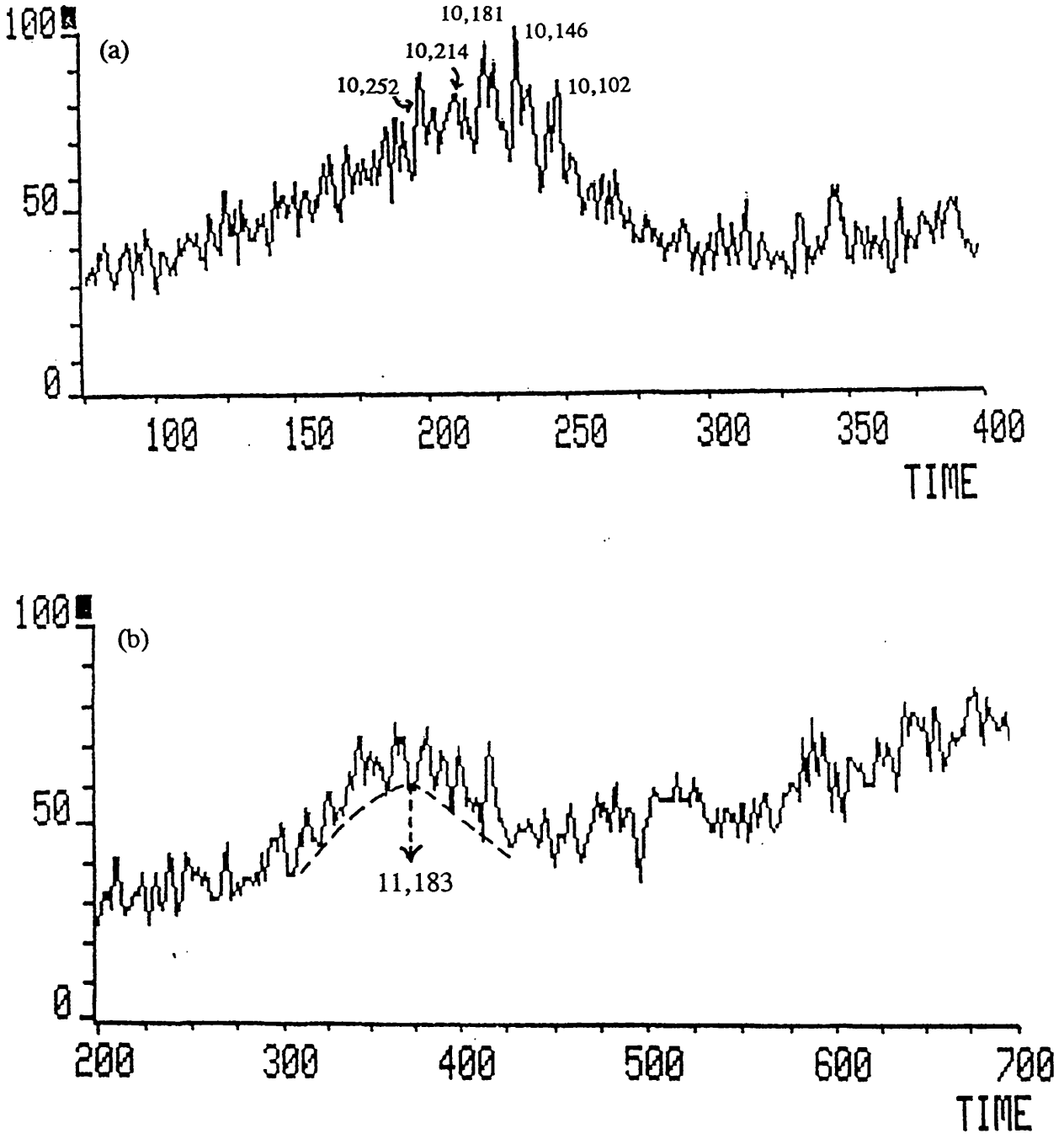


Figure 3.13 Liquid SIMS spectra of (a) *Scenedesmus obliquus* (b) *Anabaena variabilis* plastocyanins

in mass between these series were 37.0 (B-A), 38.1 (C-B), 39.6 (D-C) suggesting that B,C and D represent potassium adducts. The A series appears to represent the fully protonated plastocyanin, with an experimental molecular weight of 10,103.1 $\pm$ 2.2 Da. This plastocyanin was also run in the presence of EDTA to complex and remove copper, forming the *apo* protein. The addition of EDTA was found not to alter the masses of the A,B,C,D ions, from which it is concluded that the spectra obtained without the addition of EDTA already represented the *apo* plastocyanin - the copper had been lost in the electrospray process. From the sequence of figure 3.12 the expected mass of *S. obliquus apo*-plastocyanin is 9978.9 Da - which is 124 Da less than the measured value of10,103.1 Da from electrospray analysis - well outside the 1-2 Da error expected by electrospray analysis at this mass range (Carr *et al*, 1991). No explanation for this discrepancy was forthcoming until correspondence with Prof. R.P.Ambler revealed that his unpublished sequence for *S.obliquus* plastocyanin had been misquoted by Sykes in his 1986 review. The residue at position 100 in figure 3.12, which should have been lysine, had been incorrectly assigned as threonine. Ambler also mentioned that there was evidence that *S.obliquus* plastocyanin seemed to possess an extra valine at position 101 and there was also some uncertainty as to whether there was an extra tyrosine residue present at position 17. The rest of the sequence was known with a high degree of confidence to be correct (Kelly, 1975). With the Thr-Lys correction the expected mass of the protein is 10,006.9 Da, whilst an extra valine residue gives rise to a calculated molecular weight of 10,105.9 Da. The presence of an extra tyrosine yields a molecular weight of 10,169.9 Da. The molecular weight of the valine containing species differs by only 2.8 Da from the experimental electrospray measurement and it was concluded that *S.obliquus*

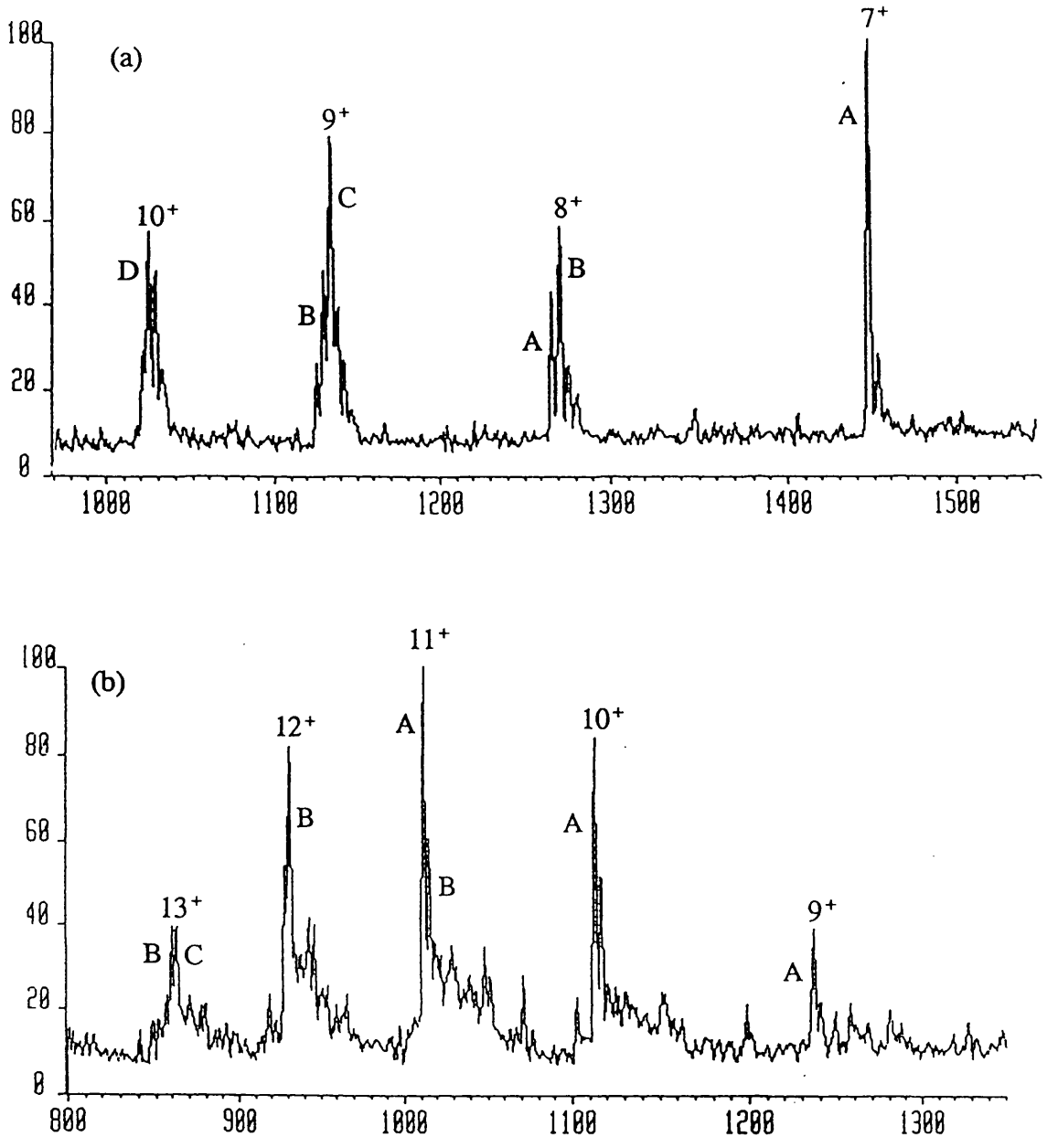


Figure 3.14 Electrospray spectra of (a) *Scenedesmus obliquus*, (b) *Anabaena variabilis* plastocyanins

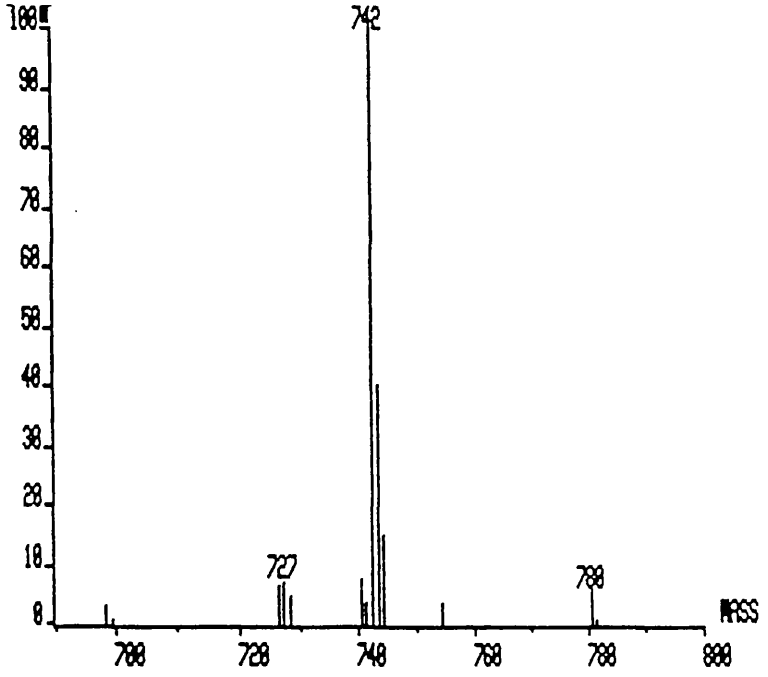


Figure 3.15 Liquid SIMS spectrum of C-terminal cyanogen bromide fragment of *S.obliquus* plastocyanin

plastocyanin contains this additional valine residue, but does not possess an extra tyrosine.

As additional proof of the proposed amino acid sequence a cyanogen bromide digest was carried out to cleave the molecule at the only methionine in the protein at residue 97. A fraction of the lyophilised digest was analysed by liquid SIMS, as shown in figure 3.15, to give a protonated pseudomolecular ion of 742 Da, consistent with the calculated molecular weight of 741 Da for the valine containing cyanogen bromide peptide with expected structure Val-Gly-Lys-Val-Ile-Val-Gln. A B/E scan was attempted after running the ordinary spectrum but unfortunately no data was forthcoming.

Gas phase Edman sequencing gave the expected corrected C-terminal sequence of ^{98}Val-Gly-Lys-^{101}Val-Ile-Val-Gln. There was no evidence of carboxypeptidase processing or microheterogeneity at the C-terminus from either the liquid SIMS data or the sequencing data, consistent with the electrospray results. The sequencing run also provided the sequence of the first eighteen residues of the large N-terminal fragment (present as a minor component) giving the sequence Ala-Asn-Val-Lys-Leu-Gly-Ala-Asp-Ser-Gly-Ala-Leu-Val-Phe-Glu-Pro-Ala-Thr- and thus provided further proof of the absence of tyrosine at position 17.

In the electrospray spectrum of plastocyanin from *A.variabilis* the A series of ions gave rise to an experimental molecular weight of 11,102.2 $\pm$ 1.3 Da. The B and C series were again seen to be potassium adducts as with *S.obliquus* (B-A = 38.2 Da). Again the addition of EDTA did not result in a shift in mass of any ions and thus 11,102 Da represents the molecular weight of *apo* plastocyanin. This mass is in excellent agreement with the calculated value of 11,103.7 Da confirming the

integrity of the sequence in figure 3.12 and the absence of any post translational modifications.

It is usual for the maximal charge state in electrospray spectra to be equal to or less than the total number of basic sites (ie the sum of Arg, Lys, His residues plus the N-terminus if this is not acetylated). Thus the spectrum of plastocyanin from *A.variabilis* is typical with fourteen basic sites and a maximal charge state of 13^+, whilst with *S.obliquus* the unusual phenomenon is noticed of the maximal charge state of 10^+ being higher than the number of basic sites, which is seven. In extensive documentation of the electrospray spectra of over a hundred proteins and peptides, Smith and co-workers (1990) note this effect in about 10% of species, although no explanation is offered. The base peaks in the the electrospray spectra of both proteins are due to the A series, at 7^+ for *S.obliquus* and 11^+ for *A.variabilis*. It is noted that for charge states higher than these the relative intensity of the A series ions drops sharply and the degree of potassium adduction increases as the charge state increases.

In the light of the primary structures deduced subsequent to electrospray analysis it is possible to re-evaluate the high mass liquid SIMS data. For *S.obliquus* plastocyanin the measured mass of 10,102 Da compares well with the mass of apo-plastocyanin (10,106 Da), being in error by 4 Da. The higher mass peaks appear to represent potassium adducts, but there is no evidence of the presence of copper as also noted with the electrospray spectra. In contrast the experimental mass of 11,190 Da for *A.variabilis* represents a difference of +88 Da to the mass of apo-plastocyanin. It is unlikely that this large error is due to the presence of copper due to it not being observed for *S.obliquus* plastocyanin. It is more likely, due to the lack of clearly defined peaks, that the liquid SIMS mass measurement for *A.variabilis*

represents an average of various adducts and small molecule losses of the protein.

In conclusion, the primary sequences of plastocyanins from the eukaryotic alga *Scenedesmus Obliquus* and the cyanobacterium *Anabaena variabilis* have been corrected and confirmed respectively compared with the sequences noted by Sykes (1985). The use of electrospray mass spectrometry was found to be essential in drawing these conclusions, whilst the use of liquid SIMS as an ionisation method for high mass analysis was found to be of limited use.

CHAPTER FOUR

DETERMINATION OF POST-TRANSLATIONAL MODIFICATIONS (I): N-TERMINAL BLOCKING

4.0 Introduction

As noted in section 2.1.0 a post-translational modification may be defined in its broadest sense as having occured if there is any alteration in the primary structure of a protein when compared with the sequence derived from the codons of the DNA gene. These modifications are centred on the termini of the protein (amino and carboxyl groups) and on the side chains of the constituent amino acids, with over 150 different modifications having been reported (Wold, 1981; Yan *et al*, 1989). The major classes of modifications and the major analytical methods used for their determination are as follows:

(i) N-terminal blocking

This is the most commonly encountered modification and is discussed in detail in section 4.1.0

(ii) Phosphorylation

Perhaps the next most commonly occuring modification is that of phosphorylation of hydroxylated amino acids, notably serine and threonine. This subject is discussed in chapter 5.

(iii) Sulphation

Sulphation of tyrosine residues is thought to be a non-reversible process occurring in around 1% of proteins (Aitken, 1990). The biological effects of this modification are not evident in many cases, although a reported effect is to cause the secretion of proteins and their proteolytic processing (Huttner, 1987). There is only one sulphotransferase known at present, having a consensus substrate sequence requirement of *neg-neg-Tyr-neg-neg* where *neg* represents a negatively charged

residue (Hortin *et al*, 1986). Tyrosine-O-sulphate is acid labile and therefore identification is incompatible with Edman analysis. A tlc method has been developed for the identification of sulphotyrosine (Karp 1983). A review of the application of MS in this field has been made (Gibson and Cohen, 1990) in which it is noted that negative ion spectra should be acquired due to the lability of the sulphate moiety in positive ion mode.

(iv) Sulphide bond determination

The formation of sulphide bonds between cysteine residues to form a cystine bridge has a profound effect upon the tertiary structure of a protein. The assignment of such bridges presents a major analytical challenge. The usual approach is to determine the primary structure of the reduced protein using classical techniques and hence the location and number of cysteine residues. Reanalysis of the oxidised form may then yield information as to which cysteine residues are involved in cystine formation (Thannhauser *et al*, 1985). Notable success in the assignment of cystine residues has been achieved through the use of fast atom bombardment MS (Morris and Pucci, 1985; Smith and Zhou, 1991)

(v) C-terminal structure

Confident assignment of the C-terminal sequence of a protein presents a common problem with the Edman sequencing approach. The main problem concerns the possibility of "ragged ends" at the carboxy terminus due to the action of carboxypeptidases (Stryer, 1981). Without the careful monitoring of PTH-amino acid yields (in the case of the analysis of two or more components) or isolation and sequencing of all proteolytic C-terminal peptides, then these modifications will not be detected. A particular difficulty encountered in the gas phase sequencing approach

is that hydrophobic C-terminal fragments without a polar amino acid residue will be extracted from the polybrene sampling disc, as discussed in section 2.2.1, thus leading to an incorrectly terminated primary structure. Sequencing from the C-terminus is not yet routine by chemical means (Bailey and Shively, 1990) and is often not practical by enzymatic means using carboxypeptidases (Stults, 1990). The advent of electrospray and matrix assisted LD make possible the detection of ragged ends in intact proteins. Electrospray was recently used to show that a recombinant sample of p24, the viral capsid protein of HIV, contained contaminating proteins with truncated C-termini (Carr *et al*, 1991).

Other C-terminal modifications include amidation and the recently discovered glycosyl-phosphatidylinositol membrane anchor (Cross, 1987)

(vi) Prosthetic groups

This class of modification embraces permanent covalent attachment of an essential cofactor for enzyme function and is discussed in chapter 5.

(vii) Glycosylation

Glycosylation of asparagine, serine and threonine residues in the Golgi body of eukaryotic cells constitutes one of the major post-translational modifications. The main biological roles of glycoproteins include cell adhesion, action as cell surface receptors and protein targetting (Kornfeld and Kornfeld, 1985; Rademacher, 1988). Serine and threonine form O-glycosidic bonds with N-acetyl galactosamine, whereas asparagine forms an N-glycosidic bond with N-acetyl glucosamine. A consensus sequence has been recognised for N-glycosidic asparagine glycosylation, being *-Asn-Xaa-Thr/Ser-*, where Xaa is any amino acid (Kornfeld and Kornfeld, 1985). Edman sequencing of glycoproteins is a poor method of analysis due to a number of factors

including the ready β-elimination of the carbohydrate chain under acidic conditions, poor extraction of the PTH-derivative and N-terminal blocking due to O→N acyl shift (Aitken, 1990). Mass spectrometry is being increasingly used in this field, frequently the carbohydrate moiety is released from the protein and then permethylated to enhance sensitivity of detection (Dell, 1990)

Gene sequencing cannot be used to determine the presence of these modifications directly. Due to the harsh conditions employed and the limitations of basing identification on retention times Edman degradation is often unsuitable for detecting modifications and mass spectrometry has emerged as being often the analytical method of choice. Powerful analytical techniques such as NMR and x-ray crystallography require large amounts of sample, in the order of milligrams, which renders them unsuitable for most endogenous studies.

In the rest of this chapter the use of MS in studies in which N-terminal blocking was evident is described together with the broader biochemical context of each study.

4.1.0 N-terminal blocking

N-terminal blocking represents one of the major classes of post-translational modification, indeed it is thought that most intracellular proteins are blocked. The most common blocking groups are acetyl, myristoyl and pyroglutamyl although methyl, glucaronyl and ketoacyl groups are also encountered, albeit very rarely (Wold, 1981). The biological role of this modification is not clear, although evidence suggests an important role in prolonging the *in vivo* half life of a protein (Bachmair

et al, 1986).

4.1.1 Detection of N-terminal blocking groups.

One of the most common methods of detection is to isolate the N-terminally blocked amino acid by chemical or enzymatic cleavage to form a small peptide, whose amino acid composition is then found. The chromatographic retention time is then compared with an authentic sample of the amino acid modified with the suspected blocking group (Randhawa and Smith, 1987).

Mild hydrolysis can sometimes be adopted to modify the N-terminus to a form which can be analysed by Edman degredation. For example pyroglutamic acid can be thus converted to glutamic acid. However this modification is of limited applicability.

Other techniques which have been used include the use of acyl hydrolases to remove N-terminal amino acids thus rendering the peptide amenable to Edman degradation (Farries *et al*, 1991). However this approach is limited to small peptides as the enzyme is ineffective otherwise.

In this chapter various studies are described which illustrate the integration of mass spectrometry as the means of identifying the presence of three of the most common N-terminal modifications noted above.

4.2.0 N-terminal acetylation

Acetylation of protein N-termini represents by far the major form of N-terminal blocking, with the claim being made that perhaps 80-90% of intracellular

eukaryotic proteins possess this modification (Wold, 1981). Newly synthesised proteins often contain N-acetylated initiator methionine residues, these may be removed by aminoacyl hydrolase and the newly formed N-terminal amino acid may also be acetylated (Arfin and Bradshaw, 1988; Wold, 1984). There is an apparent preference for the acetylation of particular amino acids with 35-50% of known acetylated proteins being acetyl-Ser, 27-33% acetyl-Ala, 5-8% acetyl-Gly, 5-6% acetyl-Met, 5-6% acetyl-Thr, 2-3% acetyl-Val and 1-3% acetyl-Asp (Aitken, 1990). The biological function of acetylation is not clear although there is evidence that acetylated proteins have a longer half-life in the cell, perhaps due to protection from ubiquitin mediated proteolysis (Bachmair, 1986). In this sub-section three examples of N-acetyl determination are discussed.

4.3.0 Primary structure of the N-terminus of Fatty acid thioesterase II

4.3.1 Introduction

In mammals *de novo* synthesis of fatty acids in lactating mammary glands is catalysed by a multienzyme complex, fatty acid synthetase, which elongates acetyl coenzyme A by successive additions of two carbon units derived from malonyl coenzyme A. The growing acyl chain is covalently bound, via a thioester linkage, to the 4'-phosphopantetheine moiety of the fatty acid synthetase (Phillips *et al*, 1970). Chain termination and release of the free fatty acid product is achieved by hydrolysis of the thioester bond by a component of the synthetase complex termed thioesterase I. The chain length of the released fatty acid is usually C_{16}. However in rats, and several other species, medium chain fatty acids (C_8, C_{10}, C_{12}) are

prominent constituents of the fatty acids in milk (Knudson *et al*, 1976). These result from the interaction of the fatty acid synthetase with a second thioesterase, thioesterase II, which is not an integral component of fatty acid synthetase but exists as a separate protein of about 30,000 Da (Libertini *et al*, 1978). The structure of rat thioesterase II has been determined by sequencing of the cDNA and also from the isolated protein (Safford *et al*, 1987; Randhawa and Smith, 1987). The protein contains a blocked N-terminus, which was tentatively assigned as acetyl due to HPLC co-elution of the N-terminal cyanogen bromide fragment with authentic acetyl-homoserine. It was the aim of this study to provide a definitive assignment of the blocking group.

4.3.2 Experimental

The N-terminal tryptic peptide of thioesterase II was provided by Dr A.Slabas. The intact protein was isolated by a published method (Slabas *et al*, 1983) and tryptic peptides obtained as follows. The protein (20nmol) was reductively alkylated using iodoacetamide, dialysed into 0.1M-ammonium bicarbonate/10mM-$CaCl_2$, pH 8.0, and digested with two equal additions of tosylphenylalaninechloromethyl ketone (TPCK)-treated trypsin at 37°C (final enzyme/substrate ratio, 1:50) added at 0 and 8 hours for a total incubation of 24 hours. Tryptic peptides were separated on a Gilson HPLC system using a Vydac C_{18} reverse phase column (4.6mm x 250mm). The column was equilibrated in 0.1% (v/v) trifluoroacetic acid before sample loading. Fractions were eluted from the column with a 5-65% gradient of 90% (v/v) acetonitrile-0.1% (v/v) trifluoroacetic acid at a flow rate of 1ml/min (Fig 4.1).

In order to identify the N-terminal peptide, having a theoretical sequence met-

Glu-Thr-Ala-Val-Asn-Ala-Lys, a computer programme, developed by Dr T.Marlow, was utilised to order the elution position of peptides under reverse phase HPLC conditions using the hydrophobicity indices for amino acids and from considerations of chain length (Creighton, 1983). The predicted order of elution of peptides expected from a tryptic digest of thioesterase II is listed in table 4.1. Amino acid analysis was carried out subsequent to the overnight hydrolysis of peptides in 6M HCl at 110°C in the vapour phase under nitrogen and analysed using an Applied Biosystems 420 amino acid analyser. Gas phase sequencing was carried out on an Applied Biosystems 477A gas phase sequencer. Fast Atom Bombardment mass spectra were recorded in the positive ion mode on a VG ZAB-SE mass spectrometer at an accelerating voltage of 8kV using glycerol/thioglycerol/acetic acid (2:2:1 by vol.) matrix. A standard Iontech fast atom bombardment source was used to generate an 8keV xenon beam.

4.3.3 Results and discussion

The computer programme indicated that the required peptide should elute as one of the earliest fragments on reverse phase HPLC. Amino acid analysis confirmed this, with fragment 6 giving the expected composition of the N-terminal peptide (Ala_2, Asn, Glu, Lys, Met, Thr, Val). When subjected to gas phase sequencing no sequence data was obtained. As N-terminally blocked peptides cannot participate in the Edman degradation, this confirmed that the N-terminal peptide had indeed been identified. The averaged fast atom bombardment spectrum exhibited a pseudomolecular ion, $[M + H]^+$, at 905.7 Da (Fig 4.2) which gives rise to an experimental molecular weight of 904.7 Da. This differs by 42.3 Da from the

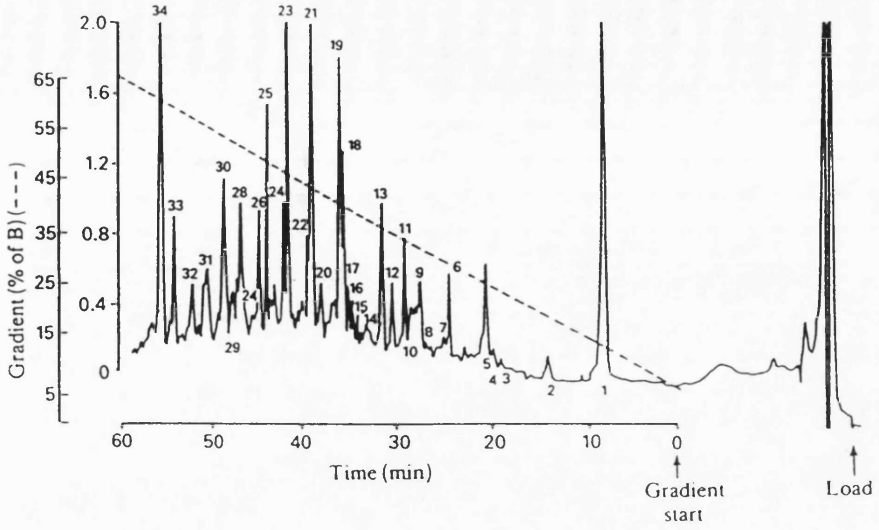


Figure 4.1 HPLC profile of trypsin digested Thioesterase II

Predicted elution no.	Sequence	Hydrophobicity
1 (1)	EK	0
2	SPR	0
3	NEK	0
4	ETR	0
5 (6)	METAVNAK	1.5
6	LAGR	1.8
7	YK	2.3
8	ADAGVVK	3.0
9	WGQK	3.4
10	CLELS	3.6
11 (5)	NYIAK	4.8
12	HHLLDFGGTPK	6.1
13	INDSLEVEHAVR	7.3
14	FIFDKPSK	7.5
15	HLIEDQVLR	7.6
16	DIEGWQDLTSGK	7.7
17	MFIPLLK	8.6
18	NLNCLYQNDAVFK	11.4
19	ALLSLDITGFLGSEDTIK	14.7
20	MEPLHIFVGASAPHST SRPQVPDLNELTEEQVR	14.9
21	LICFPWAGGGSIHFAK	15.2
22	FDVHMLPGDHFYLMKP DNENFIK	17.4
23	LGEPFANDIYQIDEI VTALLPIIQDK	24.2
24	AFAFFGHSFGSYIALITALLLK	24.5

Table 4.1 The theoretical peptides of a tryptic digest of thioesterase II and their predicted elution order (actual position in parenthesis).

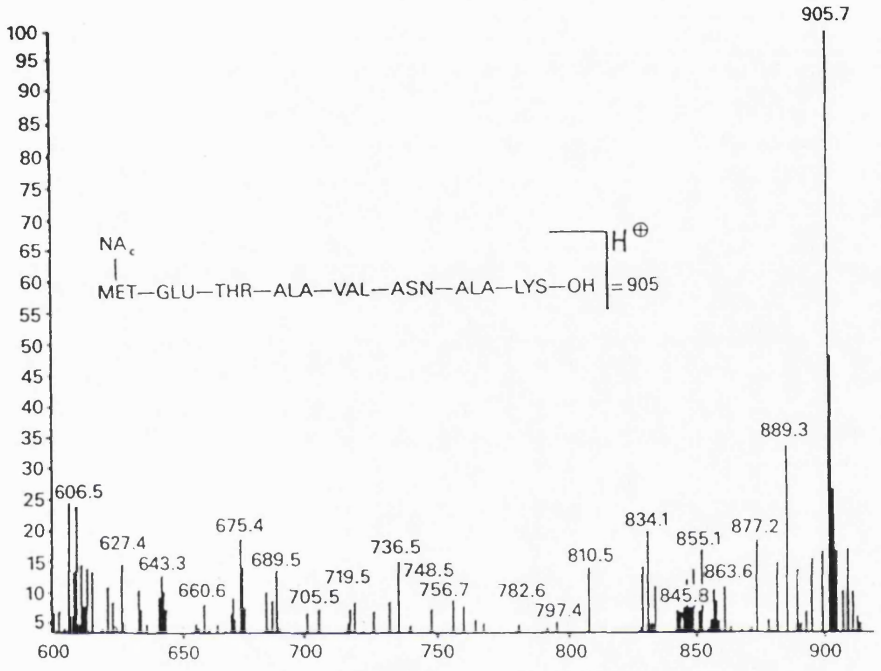


Figure 4.2 FAB spectrum of the N-terminal tryptic peptide of Thioesterase II

theoretical molecular weight of the unblocked N-terminal peptide at 862.4 Da. However for an N-terminally blocked peptide an α-hydrogen is replaced at the N-terminus, giving rise to a difference of 43.3 Da. The chemically feasible atomic compositions with this molecular weight are C_3H_7 and C_2H_3O. The only biologically feasible blocking group of these formulae is the acetyl moiety (molecular weight = 43.0 Da), propyl and isopropyl are unknown in this context whilst acetyl is well established as the major form of N-terminal blocking.

It has thus been shown conclusively that fatty acid synthetase-thioesterase II is N-terminally blocked with an acetyl group.

4.4.0 Determination of the primary structures of thymosins β_{11} and β_{12} in trout spleen

4.4.1 Introduction

ß-thymosins are 41-43 residue polypeptides of which two forms of ß-thymosins are co-expressed in some mammalian species ie thymosin $ß_4$ (Goodall *et al*, 1985) and either thymosin $ß_9$ in calf (Hannapel *et al*, 1982) and pig (Hannapel *et al*, 1989) or thymosin $ß_{10}$ in man, rat, mouse and cat (Goodall and Horecker, 1987). However only one form of ß-thymosin was identified in chicken, frog and trout (Erickson-Viitanen and Horecker, 1984).

In vertebrate tissues ß-thymosins are ubiquitous and occur in high concentrations, approaching 0.1-1% of total protein (Hannapel and van Kampen, 1987). They are expressed *in situ* with spleen being the richest source of thymosin $ß_4$ in mammals (Hannapel *et al*, 1982).

In spite of the ubiquity and high levels of ß-thymosins, their physiological role is not clear. They appear to be most closely associated with immune related cells and/or cell proliferation. Thymosin $ß_4$ has been shown to induce terminal deoxynucleotidyl transferase activity in bone marrow cells (Low *et al*, 1981) and to inhibit the migration of macrophages (Dalakas and Trapp, 1986). The endogenous N-terminal tetrapeptide of thymosin $ß_4$ has been found to inhibit bone marrow hematopoietic cell and hepatocyte proliferation (Grillon *et al*, 1990; Lenfant *et al*, 1989).

4.4.2 Experimental

Trout thymosins were provided By Dr A.A.Haritos (University of Athens) having been extracted and purified from excised trout spleen using a combination of centrifugation and gel filtration in a 95 x 2.5 cm Sephacryl S-200 column, using formic acid/pyridine buffer. Subsequent detection was achieved by radioimmunoassay using rabbit antibodies raised against synthetic segment 1-15 of rat thymosin $ß_4$. A reverse phase hplc step using a 250 x 4 mm Lichrosorb C18 column and a water / acetonitrile / 0.1% trifluoroacetic acid gradient was ultimately used to obtain two immunoreactive peaks as depicted in figure 4.3.

The two peptides had been sequenced by Dr B. Coles using an Applied Biosystems 470A gas phase sequencer with analysis of amino acids as their phenylthiohydantoin derivatives. Overlapping peptide fragments were generated by trypsin, thermolysin and S.aureus V8 digests (see section 2.3.0) with subsequent separation by reverse phase hplc using the same column and buffers as for the HPLC step in the purification of the intact peptides. Amino acid analysis was performed using a Waters Picotag system.

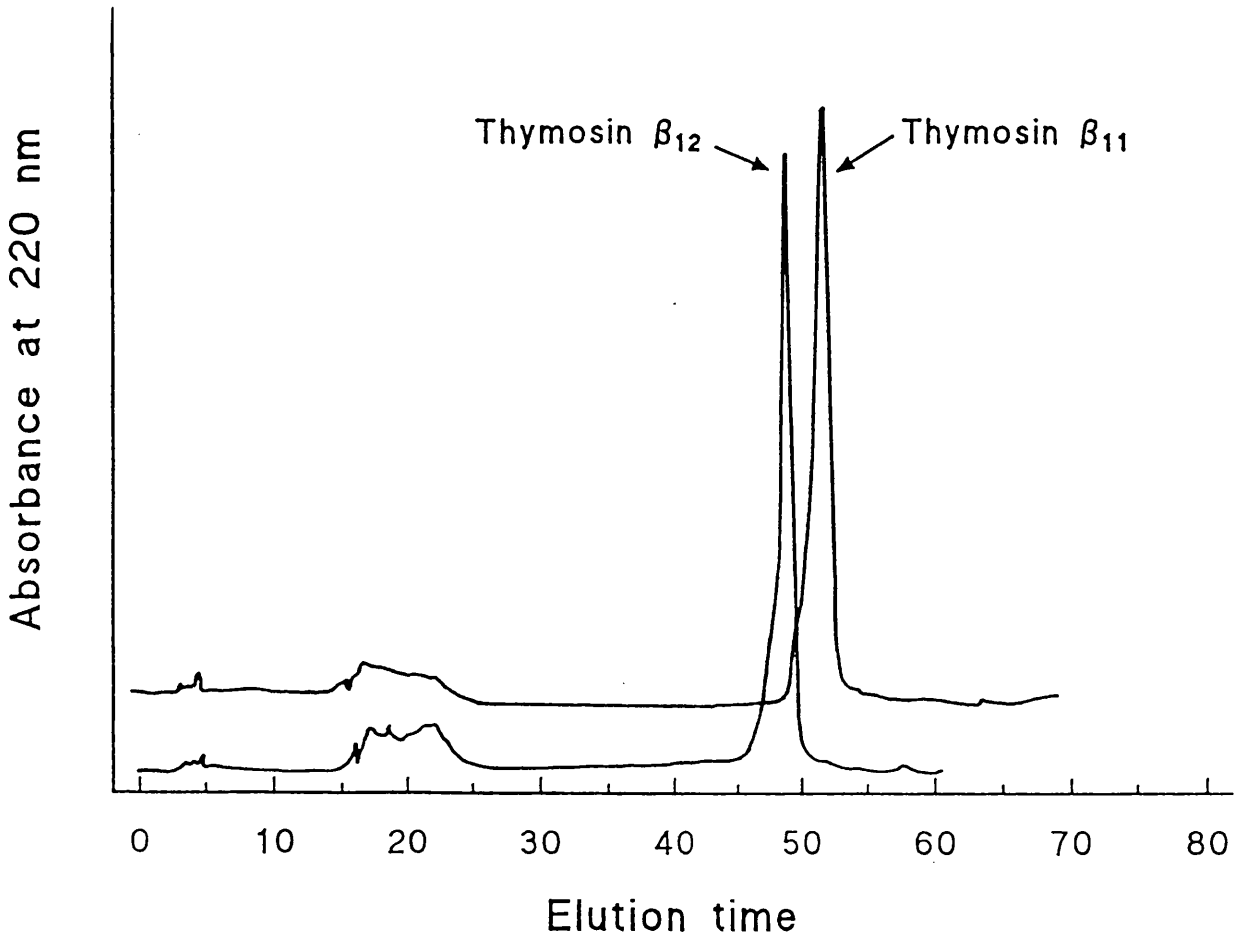


Figure 4.3 HPLC profile of thymosins β_{11} and β_{12}

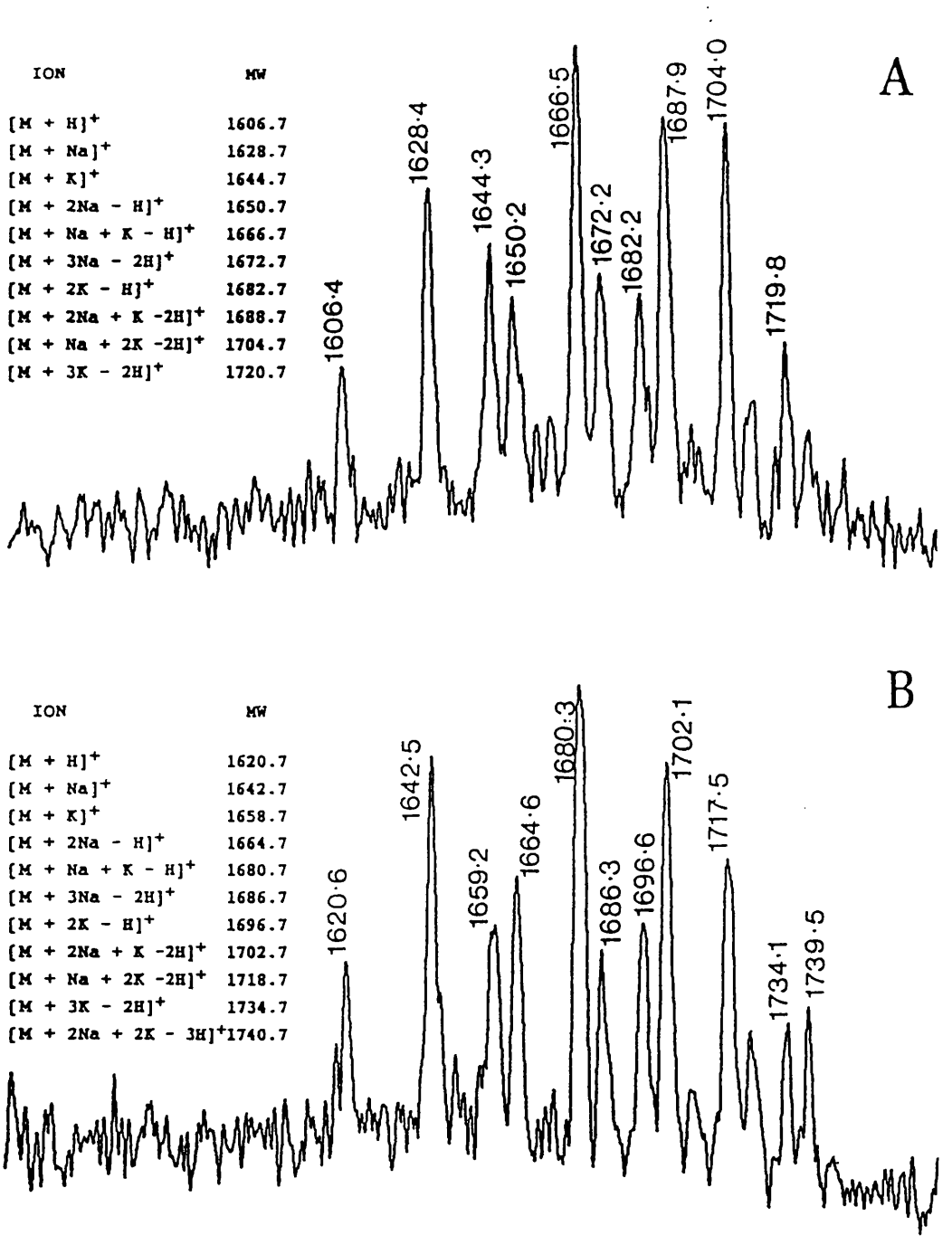


Figure 4.4 Liquid SIMS spectra of N-terminal tryptic peptides of (A) thymosin β_{12} and (B) thymosin β_{11}

Liquid SIMS spectra of the lyophilised N-terminal tryptic peptides of the two peptides were obtained on a VG 70-250SE mass spectrometer using a caesium ion gun operated at 20kV, with approximately 3 nmol of each peptide applied onto a thioglycerol/1% trifluoroacetic acid matrix. Scans were obtained in continuum mode and showed a consistent set of peaks in the range 1600-1750 as shown in figure 4.4.

4.4.3 Results and discussion

Trout spleen was selected for the isolation of ß-thymosins since this tissue has been reported to be the richest source of thymosin in other species (see introduction). Two ß-thymosins were isolated in contrast to the presence of a single form in trout liver (Erickson-Viitanen and Horecker, 1984).

Both ß-thymosins of trout had been shown to be N-terminally blocked when subjected to gas phase sequencing. However after treatment with trifluoroacetic acid at 48°C for 4-7 days (Wellner *et al*, 1990), to induce N→O acyl shift, several N-termini were evident and limited sequences had been obtained. When coupled with the data from the enzymatically digested peptides this enabled the sequences of the two peptides to be elucidated, as depicted in table 4.1. Identification of the N-terminal blocking group and further confirmation of the sequence of the N-terminal tryptic peptides was forthcoming from the liquid SIMS spectra (figure 4.4). Both thymosins displayed molecular weights in agreement with those predicted for the presence of an acetyl moiety at the N-terminus. (The calculated M_r of non-acetylated β_{11} and β_{12} are 1577.7 Da and 1563.7 Da respectively, being 41.7 Da and 41.9 Da less than the experimental values. As noted in section 4.3.3 this molecular weight difference must represent an acetyl moiety). The differences between thymosins $ß_{11}$ and $ß_{12}$ noted at

Source	Protease	Sequence
		1 10 20 30 40
Thymosin β_{11}	-	**AcSDKPNLEEVASFDKTKLKKTETQEKNPLPTKETIEQEKQAS**
	T	AcSDKPNLEEVASFDK
	S	VASFDKTKLKKTE
	Th	FDKTK
	T	LKK
	Th	LKKTETQEKNPLPTKETIEQEKQAS
	T	KTETQEK
	T	NPLPTK
	T	ETIEQEK
	T	QAS
		1 10 20 30 40
Thymosin β_{12}	-	**AcSDKPDLAEVSNFDKTKLKKTETQEKNPLPTKETIEQEKQATA**
	T	AcSDKPDLAEVSNFDK
	S	VASFDKTKLKKTE
	Th	FDKTK
	Th	LKKTETQEKNPLPTKETIEQEKQ
	T	TETQEK
	T	NPLPTK
	T	ETIEQEK
	T	QATA

Table 4.2 Alignment of proteolytic fragments of thymosins β_{11} and β_{12}. (T=Trypsin, S=*S.aureus V8*, Th=Thermolysin).

Source		Sequence
		1 10 20 30 40
Thymosin β_4	BRHC	AcSDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES
Thymosin β_4	Rab.	xADKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES
Thymosin β_4	Xen.	xSDKPDMAEIEKFDKAKLKKTETQEKNPLPSKETIEQEKQTSES
Thymosin β_9	Bov.	AcADKPDLGEINSFDKAKLKKTETQEKNTLPTKETIEQEKQAK
Thymosin β_9	Por.	xADKPDMGEINSFDKAKLKKTETQEKNTLPTKETIEQEKQAK
Thymosin β_{10}	Rat.	xADKPDMGEIASFDKAKLKKTETQEKNILPTKETIEQEKRSEIS
Thymosin β_{11}	Tro.	AcSDKPNLEEVASFDKTKLKKTETQEKNILPTKETIEQEKQAS
Thymosin β_{12}	Tro.	AcSDKPDLAEVSNFDKTKLKKTETQEKNILPTKETIEQEKQATA

Table 4.3 Comparison of thymosins β_{11} and β_{12} with known thymosins (BRHC=Bovine(Bov.),Rat,Human,Cat; Rab=Rabbit; Xen.=*Xenopus*; Por=Porcine; Tro=trout); x=unknown group. Boxed regions are invariant.

positions 5, 7, 10 and 11 were thus confirmed from the molecular weight determinations of the N-terminal fragments. Also evident in the spectra are a multitude of cationated species due to sodium and potassium which also confirm the molecular weight assignment. The presence of these cationated species has the presumed disadvantage of diminishing the ion current of the protonated species and hence sensitivity.

The sequence of one of the ß-thymosins corresponded to the previously reported thymosin $ß_{11}$ in the liver (Erickson-Viitanen and Horecker, 1984) with two differences at positions 5 and 7 (Asn replacing Asp and Glu replacing Gln). These differences appear to represent mistakes in the assumptions made in deducing the sequence of trout liver from amino acid composition analysis and comparison with the known sequence of thymosin $ß_4$. The sequencing of the novel second ß-thymosin, designated thymosin β_{12}, reveals a number of differences compared with the thymosin $ß_{11}$ at positions 5, 7, 10, 11 and 41 with the addition of an extra residue at position 42 also evident. Six differences in all between the two thymosins correspond to a much higher homology (86%) than the 11 differences (74% similarity) among the β-thymosin pairs β_4 and β_9 or β_{10} expressed in mammalian species.

A comparison of known structures of β-thymosins, including the two determined in this study, revealed extensive regions of conservation as noted in table 4.2. These boxed regions in table 4.2 may be essential to the physiological role of thymosins. In contrast the N- and C- terminal regions appear to be variable and therefore species dependent.

4.5.0 Myristoylation of N-termini

Myristic acid, a $C_{14:0}$ fatty acid, has been found in a number of proteins as an N-terminal blocking group, including cyclic AMP dependent protein kinase and protein phosphatase 2B (Carr and Biemann, 1984; Aitken *et al*, 1982). This fatty acyl derivative has been found thus far to occur exclusively on glycine residues. The role of the modification is not clear, although it is considered unlikely to be for membrane attachment and is thought to be involved in interactions between regulatory and catalytic subunits in the kinase and phosphatase cited earlier (Aitken, 1990). Myristoyl has also been discovered at the amino terminus of some viral oncogenic proteins including tyrosine specific kinases (Chow, 1987; Henderson, 1983). One such oncogenic tyrosine specific kinase is $pp60^{src}$ which has been shown to induce tumour formation in birds and rodents (Cantley *et al*, 1981). To facilitate studies on the properties of $pp60^{src}$ tyrosine kinase an N-terminal peptide was synthesised, purified by reverse phase hplc and characterised by mass spectrometry.

4.5.1 Experimental

The myristoylated (myr) peptide myr-Gly-Ser-Ser-Lys-Ser-Lys-Pro-Lys, being the N-terminal sequence of $pp60^{src}$, was a gift from Dr Torsen Saermark.

Liquid SIMS spectra of the peptide were obtained on approx. 10nmol of peptide using a VG 70-250SE mass spectrometer at an accelerating voltage of 8kV with the cesium gun operated at 20kV.

4.5.2 Results and discussion

The liquid SIMS spectrum obtained, shown in figure 4.5, is remarkable due

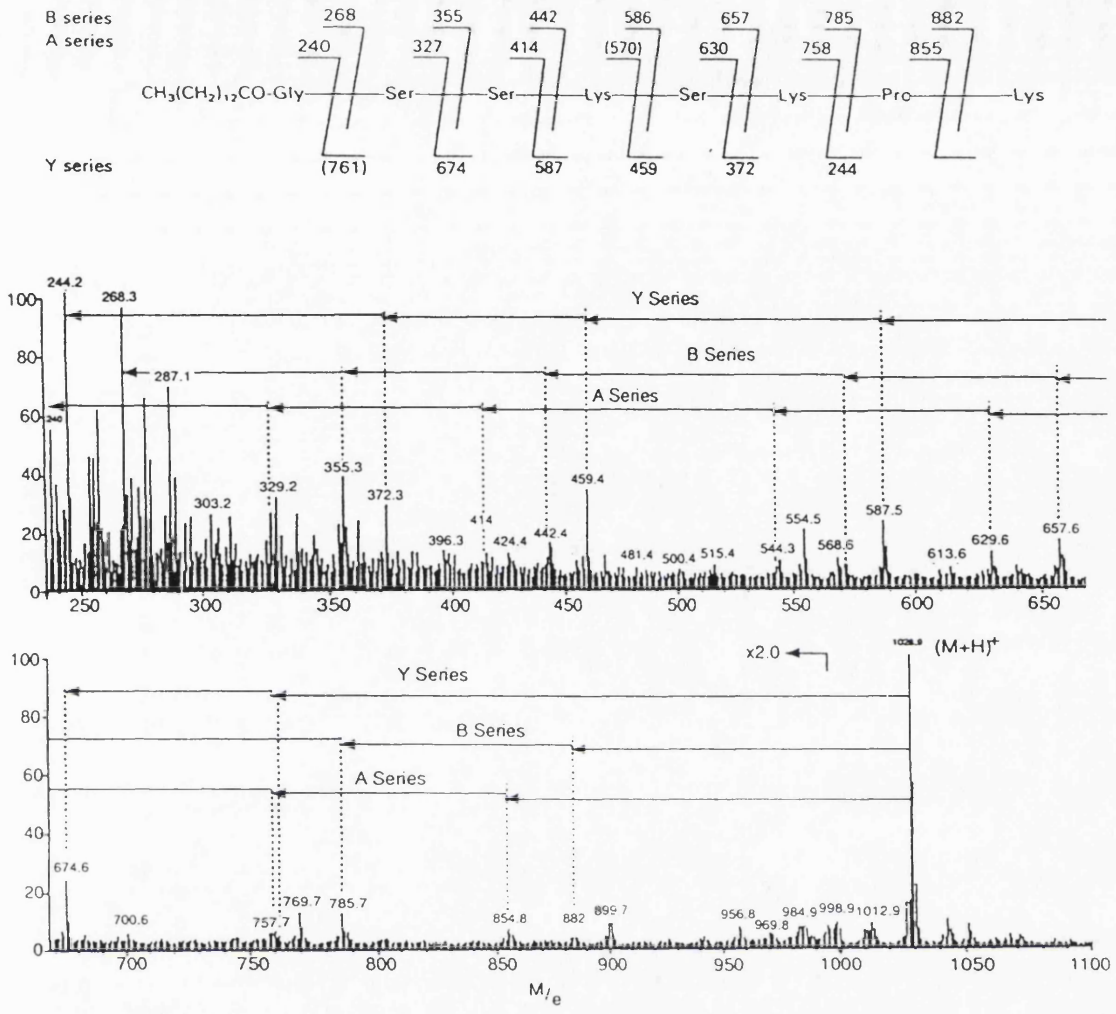


Figure 4.5 Liquid SIMS mass spectrum of myristoylated peptide from the N-terminus of $p60^{src}$.

to the large degree of interpretable fragmentation in evidence - a phenomenon rarely observed in normal (as opposed to tandem) peptide spectra. The main series of ions formed are the *y* series and the complementary *b* series, due to fragmentation at the amide functions in the peptide backbone, with the extent of fragmentation being sufficient to enable the sequencing of the majority of the peptide. For the poly-lysinated peptide PLYKKIIKKLLQS, Biemann (1990) noted that very little sequence data could be deduced from the liquid SIMS spectrum and interpreted this as being due to the presence of too many charge retaining basic sites. Acetylation of the lysines vindicated this hypothesis as an interpretable spectrum was then achieved, incidentally with the *b* and *y* series being predominantly formed. It is remarkable that the myristoylated N-terminal peptide under investigation does not exhibit the same lack of sequence related ions. However it is difficult to explain why the *d*, *v* or *w* ions often seen with peptides containing lysine or arginine are not seen. It is also worth noting that the normal FAB spectrum of the myristoylated N-terminal peptide of calcineurin (Aitken *et al*, 1984) also exhibited significant fragmentation, enabling sequence determination. The reason why myristoylated peptides seem to show significant fragmentation may be explained by the fact that hydrophobic peptides exhibit relatively strong ion currents under liquid SIMS conditions. The stronger the ion current is then the greater will be the absolute intensity of the sequence ions. As the myristoyl group has a relatively high molecular weight, then the intact peptide molecular ion and the majority of its sequence ions will be outside the range of strong interfering matrix peaks which can obscure sequence related data.

4.6.0 Pyroglutamyl N-terminal blocking

Cyclisation of N-terminal glutamic acid and glutamine (under mildly acidic or basic conditions) to form pyrrolidone rings or "pyroglutamyl" is a form of N-terminal blocking that occurs frequently (Orlowski, 1971). No clear role is evident for pyroglutamate formation, although it may have an effect analogous to that discussed earlier for acetylation, in the extension of the half life of a protein. The Edman degradation cannot proceed when this group is present (see section 2.2.1). It may also be noted that when glutamine is exposed as the newly formed N-terminus in the Edman sequencing of a peptide pyroglutamyl formation will occur to some degree, thus reducing yields. Mass spectrometry is an important means of detecting pyroglutamyl formation. For example, FAB-MS has been used to detect the presence of pyroglutamyl at the N-termini of lipotropin and pro-opiomelanocortin (Bateman *et al*, 1990)

4.6.1 Introduction

Thyrotrophin releasing hormone (TRH), pyroglutamyl-His-Pro-amide, is secreted by various mammalian tissues including those of the reproductive system (Rui *et al*, 1987). Other closely related peptides which cross react with TRH-antiserum have also been isolated, such as pyroGlu-Glu-Pro-amide which has been recently found in rabbit prostrate complex (Cockle *et al*, 1989a) and human semen (Cockle *et al*, 1989b), being characterised by Edman sequencing and mass spectrometry. No clear biological role was proposed for these peptides, however naturally occurring peptides with a C-terminal α-amide moiety almost always exhibit marked biological activity (Kahn *et al*, 1992) and it is expected that further studies

will reveal an important biological function(s) for these TRH-like species.

In the first of the studies in this section the structural elucidation of a thyrotrophin releasing hormone (TRH) like peptide, present in human seminal fluid, is discussed. The second study centres on a feasibility study on the use of B/E linked scanning to determine the C-terminal structure of a novel 40-50 amino acid polypeptide isolated from rabbit prostate complex. The tryptic C-terminal peptide of this polypeptide has been shown to be TRH-like, but to have a relatively poor cross reactivity of approx 1% with TRH-antiserum (Cockle, 1991). The peptide was intractable to Edman sequencing, as expected for the suspected pyroglutamyl N-terminus. However mild acid hydrolysis gave rise to only the N-terminal residue, glutamic acid, being determined. It is expected that the remainder of the peptide has undergone a facile internal cyclisation to form diketopiperazine, rendering Edman sequencing impossible.

4.6.2 Experimental

The human semen derived TRH-like peptide was provided by Dr Z. Kahn (National Institute of Medical research), having been isolated by a combination of gel exclusion, ion exchange and reverse phase chromatographies. Fractions containing TRH-like peptide were detected by radioimmunoassay using TRH antiserum. The TRH antiserum is known to have a high specificity for the pyroglutamyl moiety and also for the proline-amide function (Fraser and McNeilly, 1982). Further experimental details can be found in Kahn *et al* (1992).

Amino acid analysis was performed on an Applied Biosystems 420A system after hydrolysis of samples for 12 hrs in 6M HCl.

Edman sequencing was carried out on an Applied Biosystems 470 gas phase sequencer.

Mild hydrolysis of the semen peptide had been carried out by maintaining 20pmol of peptide at 100°C for 15 mins in 3M HCl. The acid was removed *in vacuo* prior to gas phase sequencing. The relevant synthetic peptides were also hydrolysed under the same conditions.

Prior to liquid SIMS analysis the endogenous peptide had been concentrated by microbore reverse phase HPLC using a Brownlee dual piston syringe microgradient system and a 1 x 100mm Spherisorb-5 RP-18 5μm C_{18} column. A 0-40% buffer B linear gradient was adopted with water as buffer A and methanol as buffer B, detection was by subsequent radioimmunoassay.

Liquid SIMS spectra were acquired on a VG 70-250 SE at an accelerating voltage of 8kV and a caesium gun operating voltage of 20kV. The lyophilised semen extracted peptide was dissolved in water and dried onto the probe to which 2μl of fresh thioglycerol + 1% TFA matrix had previously been applied. Other matrices employed were thioglycerol, glycerol and glycerol + 1% tfa. The synthetic peptides were analysed in a similar fashion. The control sample for the endogenous peptide had been obtained by lyophilising 0.5ml of buffer from a point in a blank gradient equivalent to where the endogenous peptide eluted. Spectra were acquired in the continuum mode over the mass range 500-40 Da. B/E linked scans were performed under computer control, using collision induced decomposition in the first field free region with helium gas at a pressure which gave a parent ion transmission of 80% for the ions of m/z 393 of a fresh sample of a mixture of LiF:CsI:NaI (0.5M: 1.0M: 0.5M).

Synthetic pyroGlu-Phe-Pro-NH_2 was provided by Dr Z.Kahn, whilst pyroGlu-Ser-Glu-NH_2 and pyroGlu-Glu-Ser-NH_2 were provided by Dr S.Cockle (University of Reading).

4.6.3.0 Results and discussion

4.6.3.1 Human semen derived peptide

Amino acid analysis of the endogenous human semen peptide had show a composition $Glx_{0.74}$, $Phe_{1.0}$, $Pro_{1.0}$ and a total of approx 2nmols of material. Prior to mild hydrolysis no sequence data could be obtained by the Edman method, indicating a blocked N-terminus. Subsequent to mild hydrolysis the sequence Glu-Phe-Pro was determined. The good cross reactivity with TRH-antiserum suggested the presence of pyroglutamyl and proline-amide (Fraser and McNeilly, 1982), the former post-translational modification is also consistent with the initial inability to Edman sequence the peptide. Further strengthening this hypothesis was the observation that synthetic pyroGlu-Phe-Pro-NH_2 co-eluted with the endogenous sample under the microbore HPLC conditions employed (Kahn *et al*, 1992). However conclusive proof could only be obtained by a mass analysis. Liquid SIMS analysis of synthetic pyroGlu-Phe-Pro-NH_2 using the matrices noted indicated that thioglycerol + 1% TFA gave the strongest pseudomolecular ion at m/z 373. The limit of detection was 800pmol using the aforementioned matrix. The reason for such a relatively large amount of material being required is ascribed to the rather hydrophilic nature of the peptide and also the presence of matrix related peaks. Due to the possibility that the structure of the endogenous peptide was not pyroGlu-Phe-Pro-NH_2 the mass

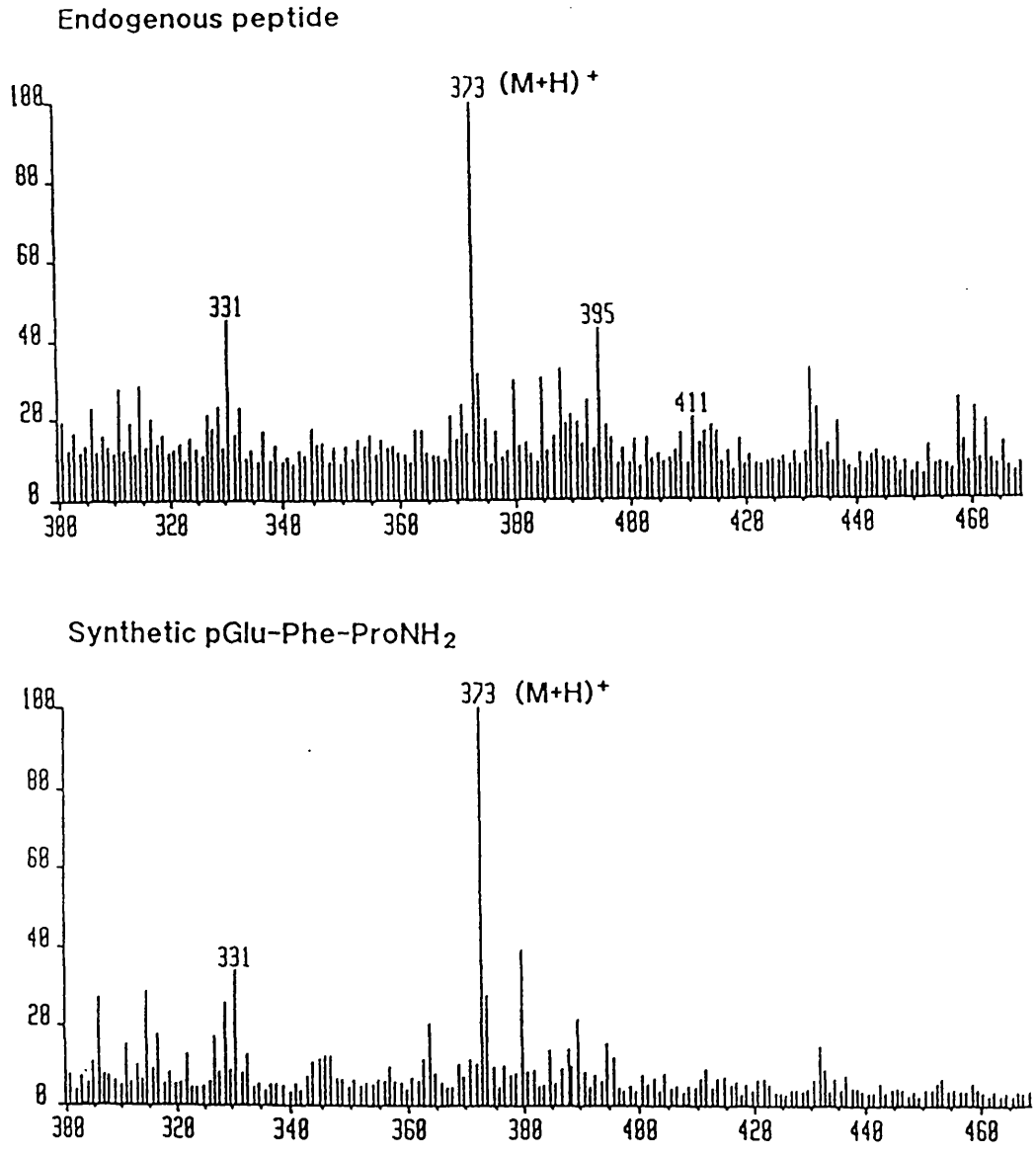


Figure 4.6 Liquid SIMS spectra of endogenous and synthetic pGlu-Phe-Pro-amide in identical buffer conditions

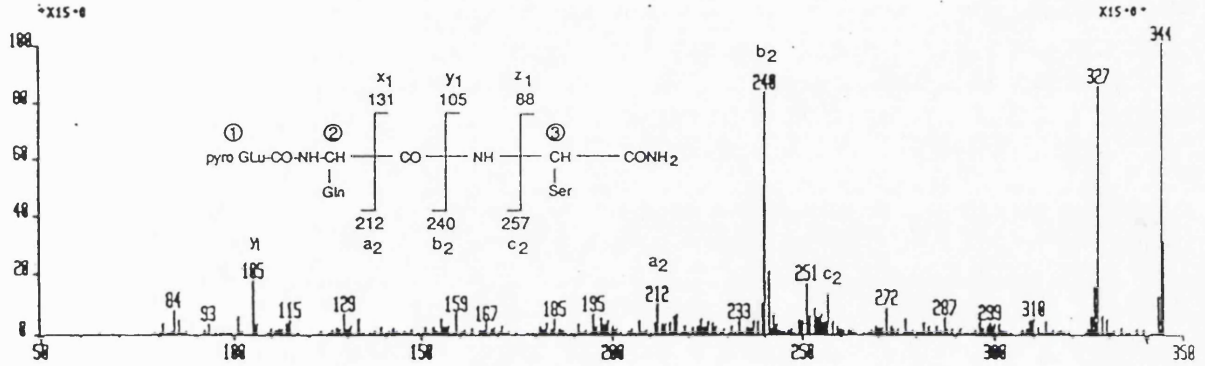
spectrometer was not operated in SIR mode, nor in the B/E linked scan mode due to the low amount of material (studies on the synthetic peptide had shown that at least 2nmol of this peptide were required for B/E sequencing to be performed). As the sequence was known, and the aim was to prove the presence of pyroglutamyl and C-terminal amide, a normal scan aimed at obtaining the molecular weight was thought to be the best methodology to adopt. The liquid SIMS spectra of the endogenous species and authentic pGlu-Phe-Pro-NH_2 are shown in figure 4.6. Ions are seen in both spectra at m/z 373 corresponding to the mass calculated for protonated pGlu-Phe-Pro-NH_2. Confirmation of the assignment of these ions as the protonated species is provided by the ions present at m/z 395, corresponding to the natriated pseudomolecular ions, and the ions present at m/z 412 which are due to potassium adduction. The mass spectral analysis proves the suspected post-translational modifications of pyroglutamyl and C-terminal amide conclusively. Thus the structure of a novel TRH-like peptide isolated from human semen is shown to be pyroGlu-Phe-Pro-amide.

4.6.3.2 Feasibility study of B/E linked scan sequencing of a novel TRH-like peptide in rabbit prostrate complex

Amino acid analysis of the C-terminal tryptic fragment of rabbit prostrate complex TRH-like peptide showed the composition Glx_2, Ser_1 (Cockle et al, 1991). It had also been shown to be a neutral peptide by ion exchange chromatographies and was thus expected to consist of a pyroglutamyl N-terminus (from the Edman sequencing and the TRH-antiserum cross reactivity), serine, glutamine and a C-terminal amide moiety - ie the peptide was thought to be either pyroglu-ser-gln-amide

or pyroGlu-Gln-Ser-amide. The latter synthetic peptide had been hydrolysed under the conditions noted in the experimental section, in an attempt to convert the pyroglutamyl moiety to glutamic acid, and subjected to Edman sequencing. As with the endogenous peptide only one cycle resulted, showing glutamic acid (ie hydrolysed pyroglutamyl). The diketopiperazine of the dipeptide Gln/Ser-amide was thought to have formed, this being an effect well characterised with, for example, his-pro-amide (Moss and Bungaard, 1990). One strategy for obtaining the sequence of this novel peptide, in the light of the previous results, would be to obtain B/E linked scan data. To prove that significant fragmentation occurs to enable their differentiation, and also to determine the amount of sample required, B/E linked scan data were obtained on the two theoretically possible structures of the rabbit prostrate complex peptide, pglu-Gln-Ser-amide and pglu-Ser-Gln-amide. The important fragment ions from the perspective of this analytical study are those which contain residue two or just residue three (counting from the N-terminal end), as shown in figure 4.7. The other fragment ions are common to both peptides. Glycerol + 1% TFA was used as the matrix for these analyses due to the problem of short lived spectra when using thioglycerol + 1% TFA. The B/E linked spectra obtained on 30nmol of each peptide are shown compared in figure 4.7. Unique diagnostically relevant ions are evident due to the a_2, b_2, c_2 and y_1 ions for both the ^{2}Gln-^{3}Ser isomer and the ^{2}Ser-^{3}Gln isomer. Unfortunately c_2 has the same mass as z_2 in the case of the ^{2}Ser-^{3}Gln isomer. The z_1 ion has the same mass in both isomers and thus cannot be used as diagnostic evidence. The facile formation of b_1 at m/z 240 in the case of the ^{2}Gln-^{3}Ser isomer would appear to offer the promise of low detection limits for this isomer. The spectra of varying amounts of pGlu-Gln-Ser-amide are shown in figure 4.7. In the

(a)



(b)

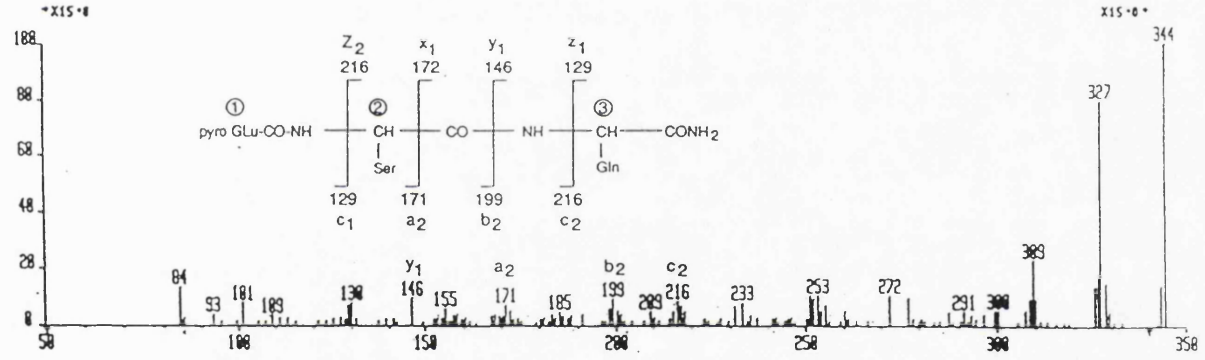


Figure 4.7 B/E linked scan spectra of (a) pGlu-Gln-Ser-amide (b) pGlu-Ser-Gln-amide

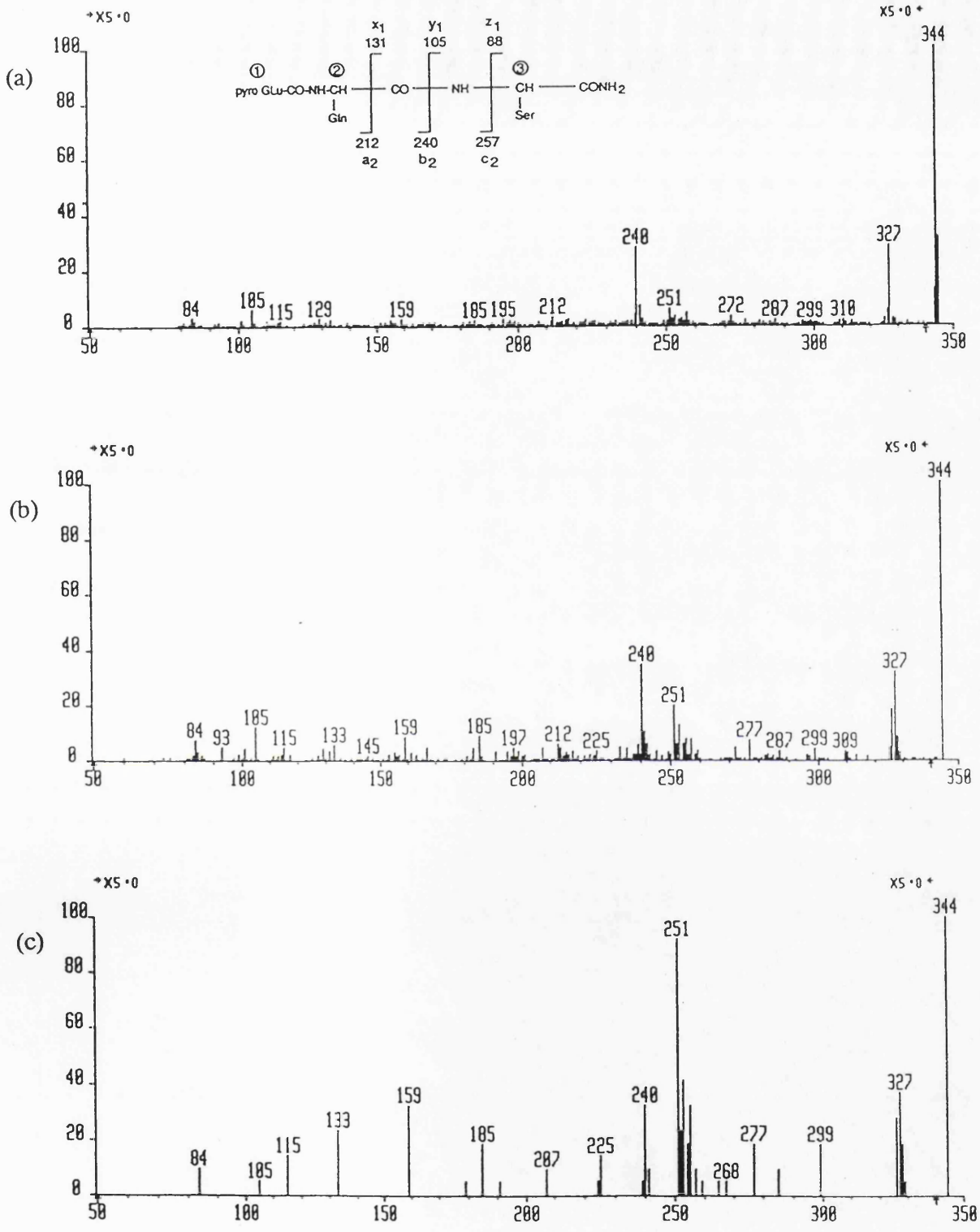


Figure 4.8 B/E linked scan spectra of pyroGlu-Gln-Ser-amide. (a) 36nmol (b) 4nmol (c) 1 nmol

spectrum at the 4 nmol level, the peptide is clearly distinguishable with the diagnostically relevant a_2, b_2 and y_1 ions in evidence. The limit of detection for this peptide appears to be approximately 1 nmol, due largely to the facile formation of the b_1 ions at m/z 240. Below this level the interference from matrix derived ions is too pronounced for confident interpretation. This may be a consequence of the large mass acceptance window inherent in the B/E technique. The limit for the ^{2}Ser-^{3}Gln isomer is higher, at approximately 2 nmol. It is evident that at least 1-2 nmol of endogenous rabbit prostrate complex peptide is required to elucidate its structure by B/E linked scanning. This may appear to be a small amount of material, but in this context it represents quite the opposite! Work is currently in progress to isolate and purify this amount of material, to enable the characterisation of this novel peptide.

CHAPTER FIVE

POST-TRANSLATIONAL MODIFICATIONS (II): PHOSPHORYLATION, PROSTHETIC GROUPS AND SYNTHETIC ARTEFACTS

5.0 Introduction

Apart from the N-terminal modifications noted in the previous chapter perhaps the most common modification is that of phosphorylation. Another common modification is the permanent covalent attachment of a co-factor as a "prosthetic group" essential for the function of a particular enzyme. Studies involving each of these modifications are discussed in this chapter, together with a discussion of the role of mass spectrometry in the analysis of synthetic peptides as a means of checking against undesired amino acid modifications.

5.1.0 Phosphorylation

Reversible phosporylation of proteins in vivo is a major mode of cell regulation, being mediated by a host of kinases and phosphatases (Hanks *et al*, 1988). Sites of phosphorylation are usually limited to serine and threonine, with tyrosine occasionally being phosphorylated. There are clearly defined consensus sequences for the substrates of most kinases however in the case of phosphatases there are no apparent consensus data despite their narrow range of action (Ingebritsen and Cohen, 1983).

The location of phosphorylated amino acids in proteins presents a difficult challenge. The most common approach is to use ^{32}P radiolabelled phosphate (usually in the form of ATP), either *in vitro* or with phosphate starved cells *in vivo*, and then to locate the radioactivity. Phosphoserine and phosphothreonine undergo elimination reactions under the acid conditions employed in the Edman degradation to form dehydro derivatives. Direct monitoring of released radiolabelled phosphate is an impractical method of determining sites of phosphorylation due to the low recovery

rates of 1-5%, however better results are obtained with the solid-phase methodology in which aqueous methanol and 2mM sodium phosphate are used in the wash cycle giving rise to good extraction of the labelled phosphate. Another commonly utilised method of detecting phosphopeptides is two dimensional paper chromatography (Amess *et al*, 1992), however identification of spots requires the availability of reference compounds.

Numerous methodologies have emerged for the conversion of phosphoamino acids to stable derivatives such as alanine (Richardson *et al*, 1978) and S-ethyl cysteine (Meyer *et al*, 1986) which can then be identified in the sequencer. However most of these derivatisation techniques only work well with the conversion of phosphoserine and are of limited use with threonine and tyrosine. Due to the limitations of the above techniques mass spectrometry is increasingly being utilised as a method of analysis.

In the following section a study of the phosphorylation of a model peptide based on an important substrate of protein kinase C is described.

5.2.0 Phosphorylation of "MARCKS" derived peptide

Protein kinase C (PKC) has been implicated as playing a key role in cell regulation in a variety of mediated effects such as DNA synthesis and cell division 0. A major substrate for PKC has been characterised as a glycoprotein of approximately 330 amino acids possessing a myristoylated N-terminus (Blackshear *et al*, 1986) and has been dubbed the "MARCKS" protein (Myristoylated Alanine Rich C Kinase Substrate) (Stumpo *et al*, 1989). As yet no role is apparent for this substrate, however it does serve as a useful marker of PKC activity *in vivo*

(Mahadevan *et al*, 1987). Characterisation of the phosphorylation site(s) is an essential step towards understanding the biological function of this enzyme. A highly conserved region of the protein having the sequence KKKKR FSFKK SFKLS GFSFKK has been identified as the domain where phosphorylation by PKC occurs (Patel and Kligman, 1987; Graff *et al*, 1989). However the stochiometry and sites of phosphorylation have not been clearly resolved due to the difficulty of unambiguously assigning digested fragments (Albert *et al*, 1987). A peptide having the sequence of the phosporylated domain was synthesised, with the aim of characterising its phosphorylation by PKC *in vitro*, from which the methodology could be extended to study the *in vivo* phosphorylation of the MARCKS protein in mouse fibroblast cells (Amess *et al*, 1992).

5.2.1 Experimental

A partially purified preparation of PKC was obtained from sheep brain (Toker *et al*, 1990) and assayed using the method of Parker *et al* (1984).

The peptide KKKKR FSFKK SFKLS GFSFKK, corresponding to residues 152-172 of the bovine brain MARCKS protein, according to the sequence determined by Stumpo *et al* (1989) and containing the phosphorylation site recognised by PKC, was synthesised using standard FMOC chemistry on an Applied Biosystems 430A peptide synthesiser.

Separate batches of 5 nmol of synthetic peptide were incubated at 10μM final concentration in 20mM Tris.HCl, pH 7.5; 10mM Mg^{2+}; 1.5mM Ca^{2+}; 10μM ATP containing 50 μCi [γ-^{32}P]ATP respectively); 50μg/ml phosphatidylserine; 20% v/v PKC (1.4 units /ml); for 1 hr at 30°C.

Estimation of the level of phosphorylation of the phosphopeptide was achieved by spotting 50 μl of the above onto P81 paper (4 x 2cm) after incubation. These were then washed, 3 x 10 ml with 75mM orthophosphoric acid, placed in 4ml scintillant (Beckman Ready Safe™) and radioactivity was measured by scintillation counting. The specific activity (cpm/pmol ATP) for each batch was estimated by spotting 50μl onto P81 paper (4 x 2cm) and measuring radioactivity by scintillation counting.

The phosphorylated peptide was equilibrated in 1% v/v TFA and fractionated by HPLC using a water/acetonitrile gradient (0-50% in 0.1% TFA) on an Aquapore Octyl reverse phase column (30 x 1 cm). Prior to LSIMS analysis the peptide was rechromatographed using the same gradient but on a Vydac C_{18} column (218TP54, 25 x 0.46 cm).

LSIMS data were acquired in the multichannel analysis mode of a VG 70-250SE mass spectrometer, at an accelerating voltage of 8kV and a Caesium gun voltage of 20kV, on 20pmol of dephosphorylated and 200pmol of phosphorylated peptide. The matrix used was *m*-nitrobenzyl alcohol containing 1% (v/v) TFA.

5.2.2 Results and discussion

The liquid SIMS spectra of the unphosphorylated and phosphorylated peptide (inset) are shown in figure 5.1. The triply phosphorylated peptide is seen to be the major phosphopeptide, with a pseudomolecular mass of 2822 Da, with the diphosphopeptide in evidence at 2763 Da. The presence of a cluster of metal cation derived species due to sodium and potassium is also evident. Whilst providing confirmation of the molecular weights of the phosphopeptides their formation has the

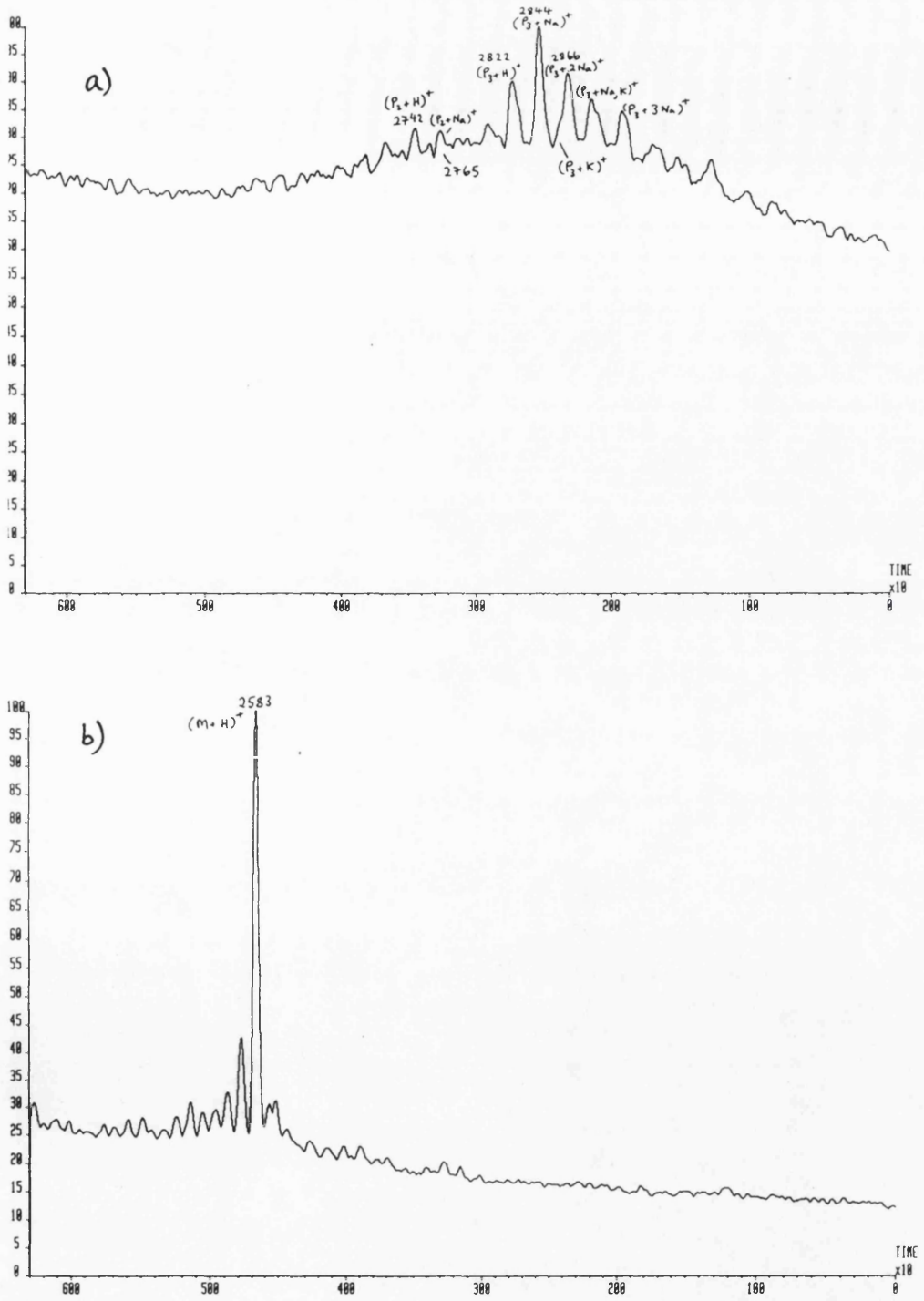


Figure 5.1 Liquid SIMS spectra of (a) phosphorylated and (b) unphosphorylated MARCKS peptide

presumed disadvantage of decreasing the ion current of the protonated species and hence sensitivity. This may be one reason why there is a dramatic difference in the sensitivity of the phosphopeptide spectrum, with 200pmol of sample, compared with that obtained with 20pmol of the unphosphorylated peptide. Another reason is that the presence of phosphate moieties will reduce the basicity of phosphopeptides compared with the unphosphorylated species. As sensitivity has been shown to increase with increasing gas phase basicity (Sunner *et al*, 1986), for positive ion analysis, then the phosphopeptides would be expected to show decreased sensitivity compared with the unphosphorylated peptide. The reason that negative ion spectra were not acquired for the phosphopeptide was due to fact that on the mass spectrometer used frequent arcing is experienced in this mode, which can effectively ruin continuum acquired spectra. With only 200pmol of sample the risk was not felt justified. By summing the height intensities of all the ions derived from the diphospho- and triphospho- peptides in the liquid SIMS spectrum a stochiometry of phosphorylation of 2.8 $\pm$ 0.1 mol phosphate / mol peptide is obtained (the error reflecting estimated error in measuring peak heights). This figure is perhaps open to criticism for a number of reasons :-

(i) It is assumed that both phosphorylated forms have similar ionisation efficiencies. In a recent article Gibson and Cohen (1990) argued that for (phospho)peptides of greater than 15 amino acids in length, this holds true if the surface activities of the peptides are similar. They used the data of Bull and Breese (1974) to calculate surface activities (Δf) and provide evidence that all the analytes must have calculated Δf values of either $\leq$ -200 cal/mol (all hydrophobic) or $\geq$ 200 cal/mol (all hydrophilic) for reliable quantification, using a Δf value of +650 cal/mol

for a phosphate moiety. For the diphospho- peptide Δf = -27 cal/mol, with a figure of +1 cal/mol calculated for the triphospho-peptide. These figures are not in the "safe" region noted. For an equimolar mixture of a phospho-peptide of Δf +81 cal/mol and the corresponding dephospho-peptide of Δf -11 cal/mol a ratio of 0.6:1 was obtained in the liquid SIMS spectrum using glycerol matrix. This suggests that the value obtained for the MARCKS peptide may be an underestimate of the degree of phosphorylation. However it should be pointed out that the conclusions of Gibson and Cohen were made on small peptides of less than ten residues length. Also the "Bull and Breese approach" does not adequately take chain length into account. A limit is predicted where alteration of a single amino acid does not significantly alter the surface activity (and hence the desorption characteristics) of a given species. In any case it is not unreasonable to deduce that two peptides which coelute using a hydrophobic column under reverse phase conditions, hence having a similar degree of hydrophobic character, would have similar surface activities.

(ii) It may also be argued that the peak at m/z 2763 is due to the fragmentation of the triphosphopeptide, losing a phosphate moiety. However the loss of phosphate was shown not to be a significant process in the analysis of the phospho and dephospho peptides of the earlier cited review (Gibson and Cohen, 1990).

(iii) The spectrum shown is due to one continuum scan only - the subsequent scans showed a large decrease in ion current in which sample related peaks were submerged in the chemical/electrical background. Lacey and Keough (1989) make the point that for pairs of compounds of differing hydrophobicities, accurate quantification requires that data be acquired over a number of scans to overcome suppression effects.

Bearing in mind points (i) to (iii) the value of 2.8 may represent an underestimate of the stochiometry of phosphorylation, but is unlikely to be an overestimate.

The stochiometry deduced from the LSIMS data may be compared with the figure of 2.2 $\pm$ 0.3 obtained from the scintillation counting approach. A major problem with the scintillation method is that hydrophilic peptides do not bind well to P81 paper and there will have been some loss of material in the washing steps. This would give rise to an underestimate in the phosphorylation stochiometry.

Comparing the values obtained by the mass spectrometric and paper based methods it is concluded that the former would appear to be more accurate.

The maximal level of phosphorylation of this peptide was determined to be 2.8 using a HPLC based method (Graff *et al*, 1989) and similar phosphorylation conditions. The Liquid SIMS result would appear to compare favourably with this value, despite the possible criticisms noted earlier, and strengthens the conclusion made as to the relative validity of Liquid SIMS compared with the paper based method. When it is remembered that the paper based approach is commonly adopted for determination of phosphorylation stochiometries and whilst more work needs to be done, it is apparent that there is a role for LSIMS analysis in this field.

5.3.0 Prosthetic groups

This class of modification embraces permanent covalent modifications of proteins which are at the active site of such enzymes. They are chemically changed during the course of a reaction and regenerated *in situ* without dissociation. An example of a prosthetic group is Flavin Adenine Dinucleotide (FAD), which is able

to gain two hydrogen atoms and participate in reduction and oxidation processes and is found in some membrane bound flavoprotein dehydrogenases and also some oxidases (Stryer, 1981). Another commonly occuring prosthetic group is biotin which is involved in metabolic processes such as fatty acid synthesis.

The detection of prosthetic groups presents the same difficulties as the detection of the modifications noted earlier. The use of mass spectrometry has a similar role to play in identifying such groups.

In this section the determination of the presence of the prosthetic group lipoic acid in the E2 subunit of bovine pyruvate dehydrogenase complex will be discussed.

5.4.0 Determination of the lipoate attachment site in pyruvate dehydrogenase complex

Pyruvate dehydrogenase complex (PDC) is one of three related multienzyme complexes present in mammalian mitochondria that are responsible for the oxidative decarboxylation of 2-oxo acids, such as pyruvic acid (Reed, 1974; Yeaman, 1986). Each complex consists of three component enzymes, termed E1,E2 and E3. E1 is a thiamin pyrophosphate-dependent 2-oxo acid dehydrogenase, and E3 is an FAD-dependent lipoamide dehydrogenase. E2, which forms the central core of the complexes, is an acyltransferase that utilises lipoic acid as an essential co-factor. The lipoic acid cofactor is covalently attached to a flexible domain of the E2 polypeptide, the mobility of the region allowing the functional end of the lipoic acid to interact with three different active sites on each complex (Wawrzynczak *et al*, 1981; Bleile *et al*, 1981).

Current evidence indicates that each E2 of mammalian PDC contains one lipoic acid residue, although a second group capable of undergoing slow acetylation is also present on the E2 polypeptide (White *et al*, 1980; Cate and Roche, 1979). In contrast, gene sequencing indicates that PDC from *E.coli* contains three lipoic acid residues per E2 polypeptide, present in three repeating highly conserved domains (Stephens *et al*, 1983). The primary structure around the lipoate residues of the *E.coli* enzyme is known, but there is no sequence information available concerning the mammalian enzyme (Hale and Perham, 1980). An additional polypeptide component of mammalian PDC has also been identified, termed protein X (De Marcucci and Lindsay, 1985; Jilka *et al*, 1986). Immunological data and peptide mapping studies indicate that that it is distinct from E2, although it has been demonstrated that protein X also contains a lipoic acid residue capable of undergoing acetylation (Hodgson *et al*, 1986). Protein X is present in only approx. 10% of the amount of E2 and its function is not yet clear.

5.4.1 Experimental

PDC from bovine heart and subsequently isolated lipoyl containing peptides were provided by Dr S.Yeaman (University of Newcastle) having been extracted as follows. PDC was purified from bovine heart by a standard procedure (Stanley and Perham, 1980) and the presence of lipoate assayed by the incorporation of radioactive acetyl from incubation with [3-^{14}C] pyruvic acid in the presence of N-ethyl maleimide as a reducing agent. These studies indicated an incorporation of 1 mol acetyl per 1 mol E2 polypeptide (allowing for the presence of protein X). The acetylated sample was then digested with pepsin and subjected to high voltage electrophoresis at pH 1.9.

Subsequent autoradiography showed the presence of two major acetylated peaks, termed A and B. Ion-exchange h.p.l.c. of peptides A and B, using a TSK DEAE 3SW column (LKB), provided a single major radioactive peak in each case, termed A1 and B1. Following reverse phase h.p.l.c. using a Vydac C_{18} column with trifluoracetate as the counter ion peptide A gave a single major peak, A1, whilst peptide B gave two major peaks, B1a and B1b. The final purification step was to repeat the reverse phase chromatography with ammonium acetate. This resulted in the splitting of peak A1 into two closely spaced peaks A1a and A1b. Likewise the peaks B1a and B1b also split into closely spaced pairs to give peptides termed B1a,B1a' and B1b,B1b'. Further experimental details are found in Bradford *et al* (1987). These six peptides were lyophilised and forwarded for structural analysis.

Sequence analysis was carried out using an Applied Biosystems 470 gas phase sequencer. Amino acid analysis was carried out on an Applied Biosystems 420 analyser after hydrolysis of peptides in 6M HCl for 24 hrs. One third of the amino acid phenylthiohydantoin derivative obtained at each cycle of Edman degradation was directly analysed, and the rest of the sample was recovered in the fraction collector of the sequencer, transferred to plastic vials containing Liquiscint scintillation cocktail (National Diagnostics) and ^{14}C radioactivity determined.

FAB mass spectra were obtained on a VG-ZAB-SE mass spectrometer operated at 8kV accelerating voltage with bombardment effected by an 8kV xenon beam from an IonTech gun. The peptides (approx 500 pmol) were resuspended in 10μl of methanol (Fisons hplc grade) and transferred to the probe tip. After partial evaporation, 1μl of glacial acetic acid was added, followed by 2μl of 1:1 (v/v) thioglycerol/glycerol matrix.

5.4.2 Results and discussion

The sequences of the isolated peptides as determined by gas phase sequencing are shown in table 5.1, with each peptide giving a single unambiquous sequence. Peptides A1a and A1b showed identical sequences with no residue being detected at position 5; however, a peak of radioactivity was released at this position in each case (see figure 5.2) consistent with the presence of an acetylated residue. Amino acid analysis of each peptide revealed the presence of a lysine residue, which was not detected during sequence analysis. The pairs of six residue peptides B1a,B1a' and B1b,B1b' were shown to differ in valine/isoleucine at position 4 but to be otherwise identical. No residue was detected at position 1 in peptides B1a, B1a', B1b and B1b', but again a peak of radioactivity was released at this position.

The FAB spectra for each pair of peptides were identical with pseudomolecular ions exhibited in the positive ion mode as follows; 1425 Da for peptides A1a and A1b; 994 Da for peptides B1a and B1a'; 980 Da for B1b and B1b' (see figures 5.3 and 5.4). The theoretical molecular weights of the expected (protonated) acetyl lipoyl-lysine derivatised peptides are 1299 Da, 868 Da and 854 Da respectively. Thus in each case a difference of 126 Da is noted. Reaction of the free thiol on the lipoyl moiety with N-ethyl maleimide accounts exactly for a difference of 126 Da. No evidence for differences in structure of the pairs of peptides was obtained that could explain their separation on reverse phase HPLC, their molecular masses being identical. One possibility is that the peaks may represent the 6- and 8-S-acetyl derivatives of the lipoyl-lysine residue (O'Connor *et al*, 1982; Yang and Frey, 1986).

The amino acid sequence surrounding the lipoic acid cofactor on E2 of

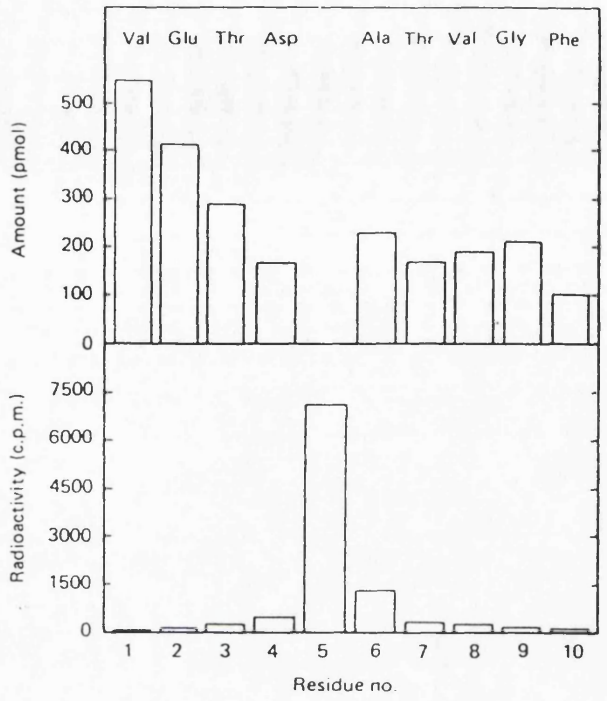


Figure 5.2 Sequence analysis of peptide A1b

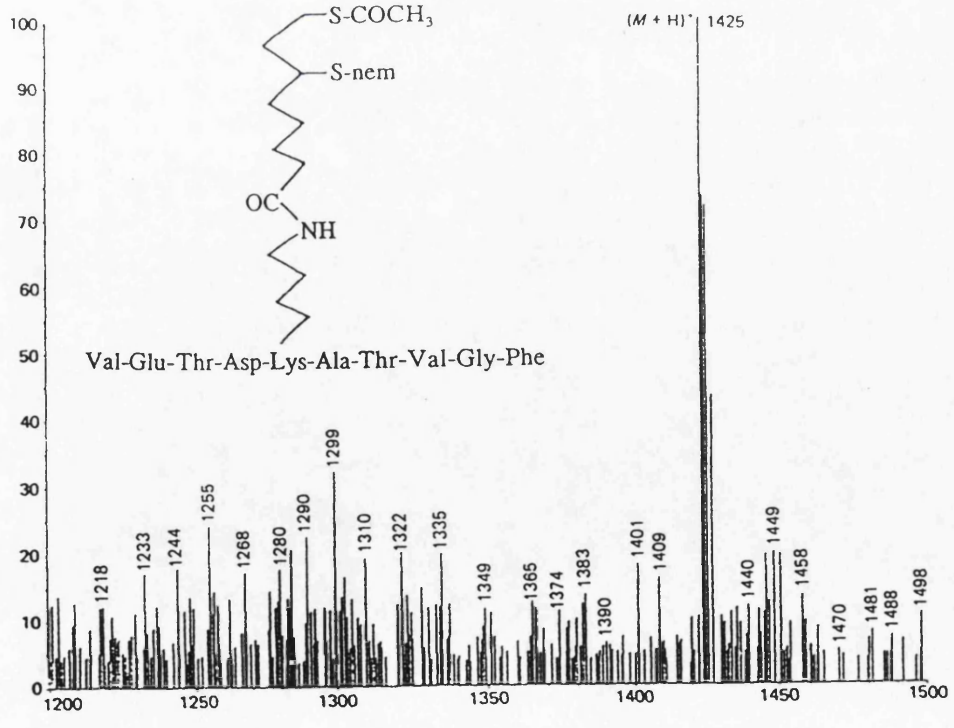


Figure 5.3 Liquid SIMS spectrum of peptide A1b

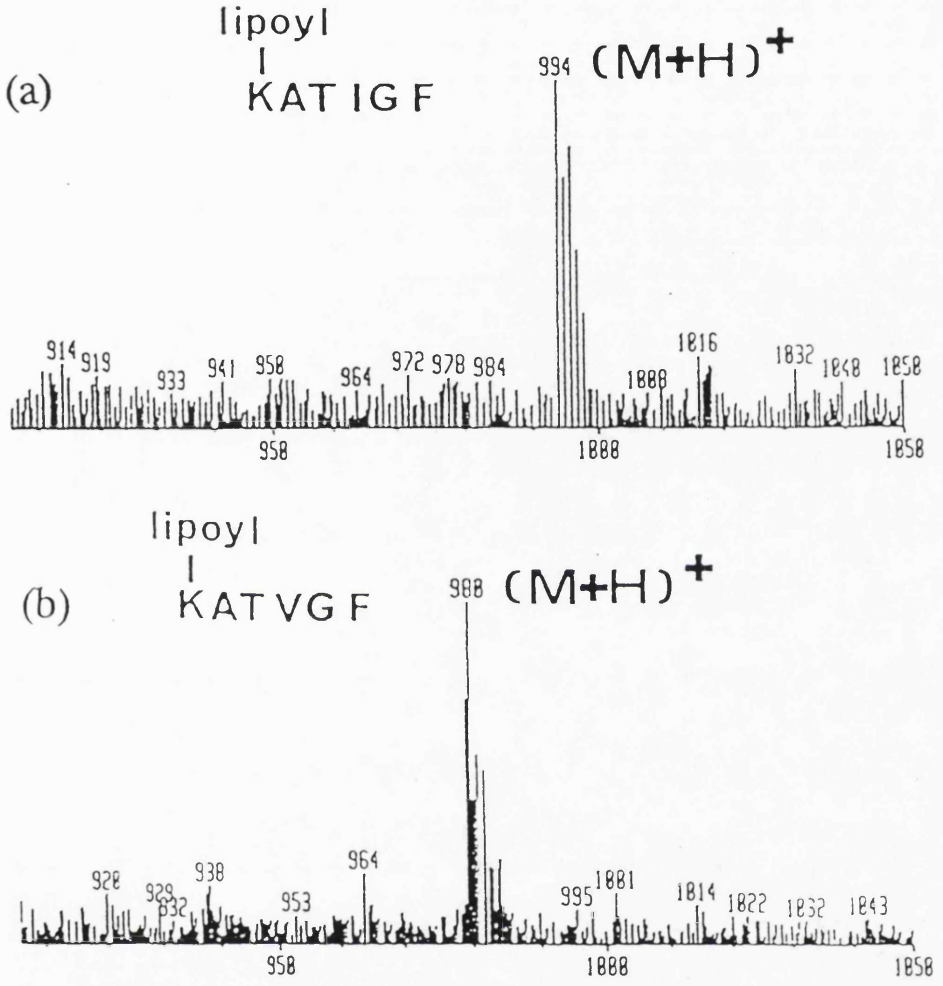


Figure 5.4 Liquid SIMS spectra of peptides B1a and B1b.

mammalian PDC has thus been determined. The sequences of the different lipoyl containing peptides are internally consistent, with the exception of the finding of an isoleucine residue instead of valine at position 4 in the peptides B1a and B1a'. No decapeptide containing an isoleucine residue was recovered, and it is possible that other differences may exist at the N-terminal side of the lipoyl-lysine. There are several possible explanations for the presence of the two different sequences. E2 may contain two or more lipoate residues, but the observed stochiometry of acetylation noted earlier argues against this possibility. Although three lipoate residues are thought to be present on each E2 polypeptide of *E.coli* PDC a maximum of two acetyl groups can be incorporated (Danson and Perham, 1976; Collins and Reed, 1977). Furthermore in E2 from *E.coli* PDC the sequence around the three lipoate residues is identical on each domain (Stephens *et al*, 1983). Although a second group on bovine kidney E2 can undergo slow acetylation, this does not occur when the reaction is carried out in the presence of N-ethyl maleimide (Cate *et al*, 1979; Bradford *et al*, 1987). Alternatively peptides B1a and B1a' may be derived from protein X and not from E2. However, this seems unlikely, as protein X is thought to be present at approx. 10% the amount of E2, whereas the isoleucine containing peptides B1b and B1b' constituted approx. 30% of the recovered acetylated peptides. A further possibility is that there is microheterogeneity in the gene for the E2 polypeptide present in the bovine population.

Using the European Molecular Biology Laboratory protein sequences database, the sequences determined demonstrate significant homology with the corresponding region of E2 of PDC (Hale and Perham, 1980), with 2-oxoglutarate dehydrogenase complex (OGDC) from *E.Coli* (Spencer *et al*, 1984) and with the lipoate-containing

Peptides	Sequences
A1a, A1b	Val-Glu-Thr-Asp-Lys*-Ala-Thr-Val-Gly-Phe
B1a, B1a	Lys*-Ala-Thr-Ile-Gly-Phe
B1b, B1b'	Lys*-Ala-Thr-Val-Gly-Phe

* Denotes lipoyl-lysine residue

Table 5.1 Amino acid sequences of lipoate containing peptides

Source	Enzyme	Sequence
Bovine heart (present work)	PDC E2	Val-Glu-Thr-Asp-Lys*-Ala-Thr-Val-Gly-Phe
		and Lys*-Ala-Thr-Ile-Gly-Phe
E. Coli	PDC E2	Val-Glu-Gly-Asp-Lys*-Ala-Ser-Met-Glu-Val
E. Coli	OGDC E2	Ile-Glu-Thr-Asp-Lys*-Val-Val-Leu-Glu-Val
Chicken liver	Glycine-cleavage H protein	Leu-Glu-Ser-Val-Lys*-Ala-Ala-Ser-Glu-Leu

* indicates lipoyl-lysine residue

Table 5.2 Amino acid sequences around the lipoyl-lysine residues of lipoate containing proteins.

region of the H protein of the chicken liver glycine cleavage system (Fujiwara *et al*, 1986), as shown in table 5.2.

5.5.0 Detection of "modifications" in synthetic peptides

The detection of amino acid modifications which may inadvertantly be present in synthetic peptides presents a similar analytical challenge to that experienced for the modifications discussed previously. Again, mass spectrometry has an important role to play in the detection of these mostly undesired modifications.

The use of synthetic peptides is widespread in biochemical research. Among the many applications are their use in the production of peptide antisera, as probes of enzyme activity, as models to study secondary/tertiary structural conformations and for their intrinsic biological activity (eg as hormones or neurotransmitters). Synthetic proteins have been manufactured containing as many as 99 amino acids (Ramage *et al*, 1990)

A major breakthrough in peptide synthesis came with the development of the "solid phase" methodology by Merrifield (1964). This technique has been refined over the years, one of the most popular current synthetic protocols utilises so-called "FMOC" (Fluorene MethOxy Carbonyl) chemistry (Atherton and Sheppard, 1989) and is outlined in figure 5.5. Thus it can be seen that the peptide is grown on the resin one residue at a time with side chain reaction prevented by the use of blocking groups. The yield for the addition of each amino acid is in the order of 99% with the FMOC chemistry. This sounds impressive, however for an average sized peptide of 25 residues this corresponds to an overall yield of 78% (assuming the

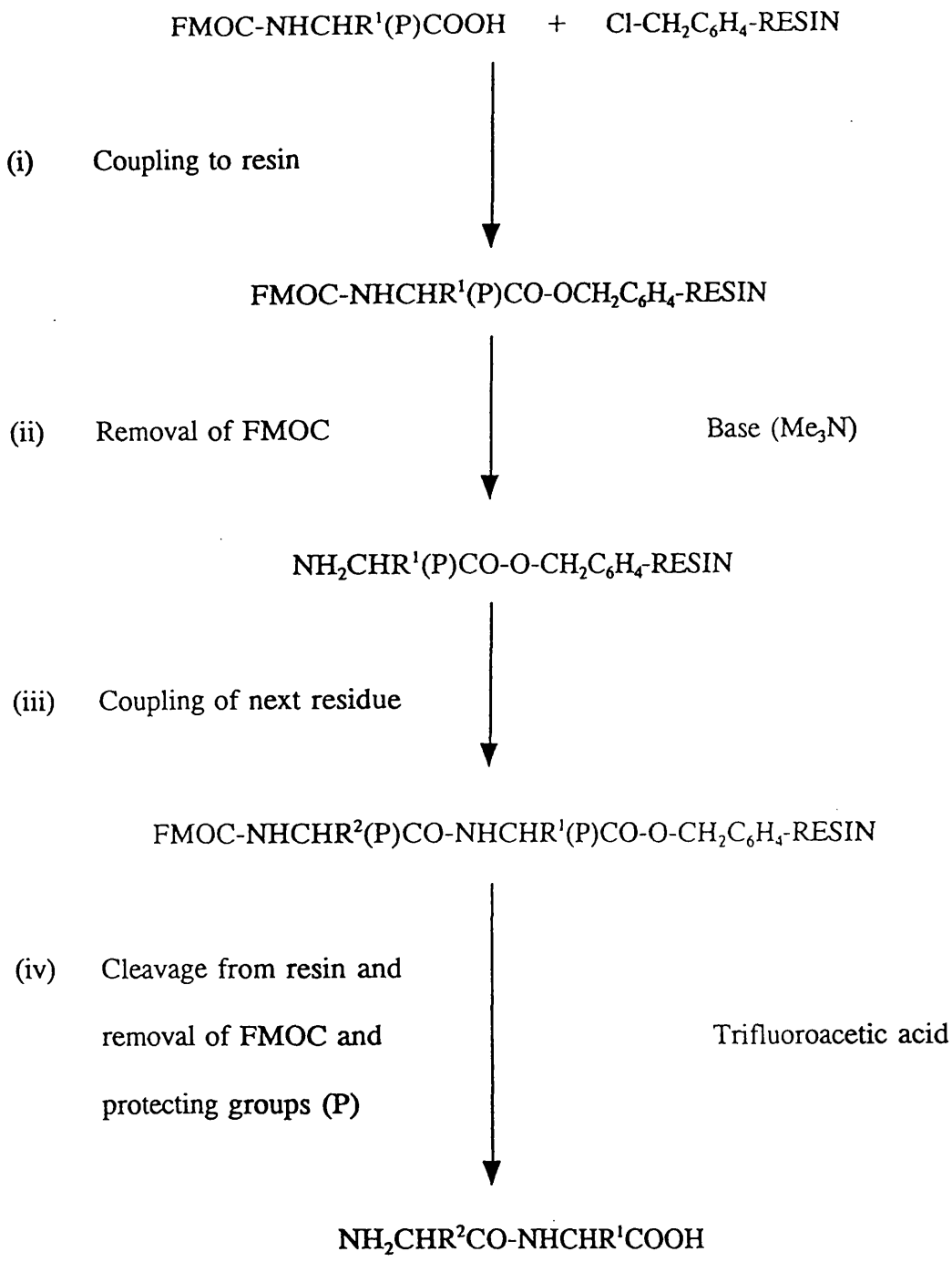


Figure 5.5 Schematic of solid phase peptide synthesis utilising Fluorene Methoxy carbonyl (FMOC) chemistry.

Amino acid	Protecting group	Mass difference if protecting group is present
Ala	-	-
Arg	4-methoxy-2,3,6-trimethylbenzene-sulphonyl (Mtr)	212.3
Asn	4,4 Dimethoxybenz-hydryl (MbH)	226.2
Asp	tertiary butyl (^{t}Bu)	56.1
Cys	^{t}Bu, trityl, acetamido-methyl (Acm)	56.1, 242.2, 71.1
Gln	MbH	226.2
Glu	^{t}Bu	56.1
Gly	-	-
His	Trityl, Boc	242.2, 100.1
Ile	-	-
Leu	-	-
Lys	tertiary butyl-oxy-carbonyl (Boc)	100.1
Met	-	-
Phe	-	-
Pro	-	-
Ser	^{t}Bu	56.1
Thr	^{t}Bu	56.1
Trp	-	-
Tyr	^{t}Bu	56.1
Val	-	-

Table 5.3 Commonly used side chain protecting groups in solid phase peptide synthesis.

deprotection/cleavage step is 100% efficient). The corollary of this is that the synthesised peptide will contain 22% of impurities, mainly in the form of deletion peptides, having very similar structures to the desired product. In practice the yield of desired product is less due to less than 99% coupling efficiency and also due to the incomplete removal of blocking groups. Overall yields of 50% are considered par for the course, further highlighting the corollary noted earlier.

A recent report of the Association of Biomolecular Resource Facilities cited by Geisow (1991) focussed on the analyses of a test peptide synthesised in 38 different facilities across the U.S.A. where tBOC and FMOC chemistries had been used about equally. The report noted that the peptides received were generally of poor purity with amino acid deletions, oxidised residues and a range of incompletely removed blocking groups present as impurities. In most of the experimental applications noted earlier the presence of impurities could cause erroneous results to be obtained.

It is evident that the quality control of synthetic peptides is of some importance. The most common methods of analysis are amino acid analysis and HPLC. Thus if the amino acid analysis indicates the required composition and HPLC gives rise to a single major peak, then the synthetic peptide is assumed to be correct and of high purity. However this level of quality control is open to serious criticism as follows:

(i) The presence of unremoved blocking groups cannot be detected by amino acid analysis due to the strongly acidic conditions used (typically 6M HCl for 20hrs). The technique also has a margin of error of approximately 5%, which is usually sufficient to obscure the detection of deletion peptides.

(ii) Modified peptides may have such similar structures to the desired peptide that they co-elute by HPLC.

Thus it is possible to have a quite badly contaminated product which appears to be pure when analysed using these techniques alone. Two examples are discussed which highlight the shortcomings of this approach, together with the analytical data provided by mass spectrometry.

5.5.1 Experimental

Peptide AGAEE LFARK FNA was synthesised by Mr Ralph Foulkes (NIMR) using FMOC chemistry on an Cambridge Research Biochemicals peptide synthesiser. HPLC was carried out using a 250mm x 4.5mm Vydac C18 column on a Gilson gradient system at a gradient of 1% / min starting with 0% buffer B (buffer A = 0.1% tfa; buffer B = 0.082% tfa in HPLC grade acetonitrile).

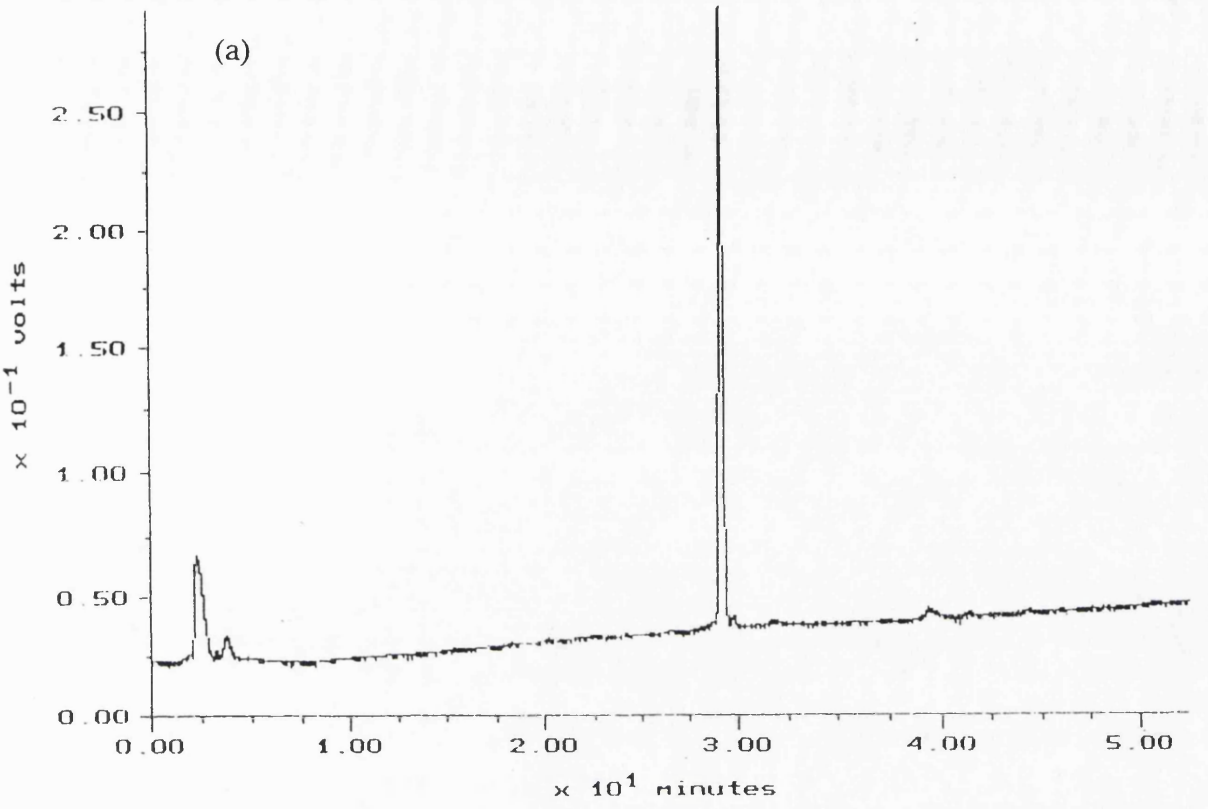
Amino acid analysis was performed on this peptide on a Beckman analyser utilising ninhydrin detection. Prior to analysis the sample was hydrolysed in 6M HCl for 20 hours at 100°C.

Acetyl-RRKWQ-KTGHA-VRHGR-L was purchased from a commercial supplier (University of Birmingham), having been tested for purity by amino acid analysis.

5.5.2 Results and Discussion

The synthetic peptide AGAEE LFARK FNA when analysed by HPLC gave rise to a single peak as depicted in figure 5.6(a). Amino acid analysis yielded the composition $Asx_{0.96}$, $Glx_{2.09}$, $Gly_{1.01}$, $Ala_{3.94}$, $Leu_{1.01}$, $Phe_{2.00}$, $Lys_{1.01}$, $Arg_{0.97}$ which

appears to confirm the expected composition. These two results give the strong impression that the synthetic product consists almost entirely of the required peptide. The mass spectrum depicted in figure 5.6(b) shows this to be an incorrect conclusion. Four major peaks are evident at m/z 1424 (that of the required peptide), 1239, 1091 and 963. The lower mass peaks are not due to fragmentation of m/z 1424. The mass differences between adjacent major peaks are 185 Da, 148 Da and 128 Da. The latter two differences correspond to the amino acids phenylalanine and lysine/glutamine respectively. Referring to the sequence of the peptide the amino acids lysine and phenylalanine are near the C-terminus. The end residues of asparagine and alanine have a combined mass of 185 Da and thus correspond to the loss noted between m/z 1424 and 1239. Thus the lower peaks represent three peptides with truncated C-termini. This truncation must have occured due to the rather harsh cleavage conditions employed in which the resin was incubated in trifluoroacetic acid for 20 hours. This has resulted in the limited hydrolysis of the peptide to produce the observed products. These have coeluted under the HPLC conditions employed, which is a common occurence with peptides of similar structure. The mass spectrum gives a qualitative picture of the impurities present, accurate quantification is not possible without the use of internal standards unless the Bull and Breese values are either $<$ -200 cal/mol or $>$ 200 cal/mol (Gibson and Cohen, 1990), which condition is not satisfied here. (AGAEE-LFARK-FNA Δf = 124 cal/mol, AGAEE-LFARK-F Δf = 11 cal/mol, AGAEE-LFARK Δf = 164 cal/mol, AGAEE-LFAR Δf = 131 cal/mol). However as the hydrophobicities of these peptides are so similar that they coelute under HPLC conditions, it may be that the surface activities are close enough for the peptides to have similar desorption properties.



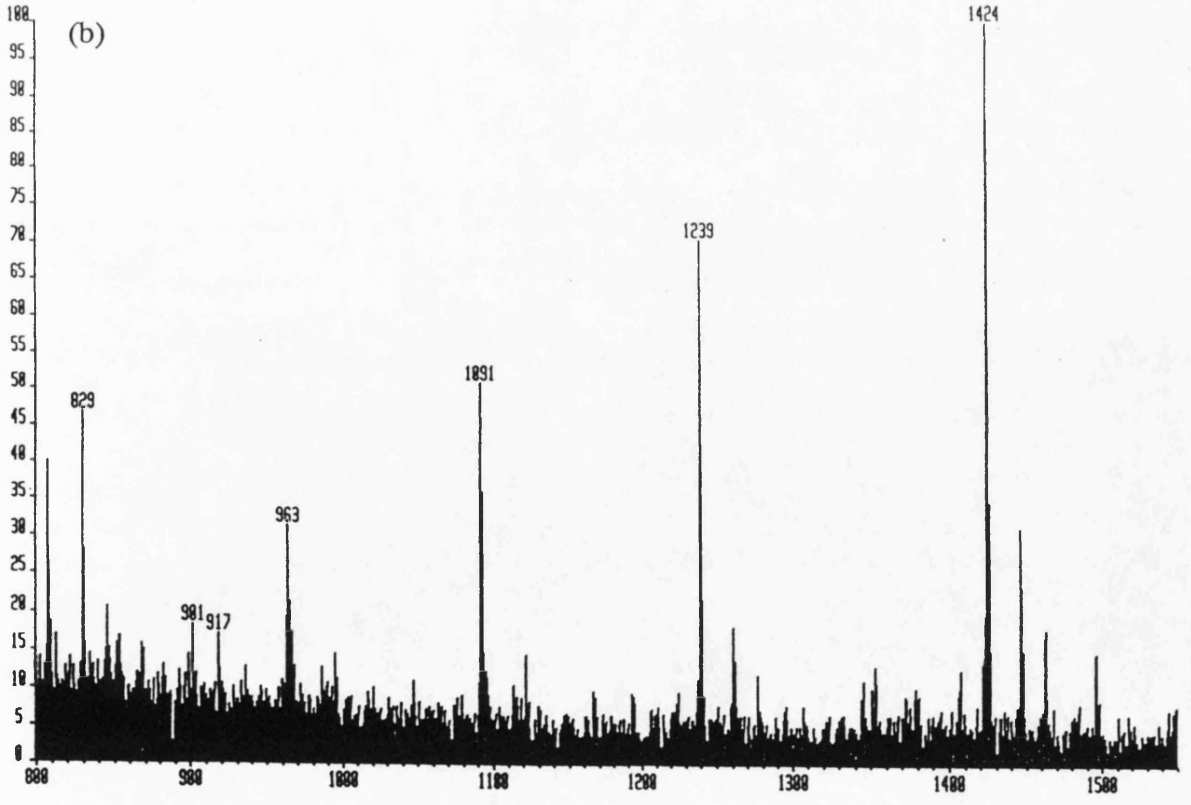


Figure 5.6 (a) HPLC profile (b) liquid SIMS spectrum of synthetic AGAEELFARKFNA (M, 1423)

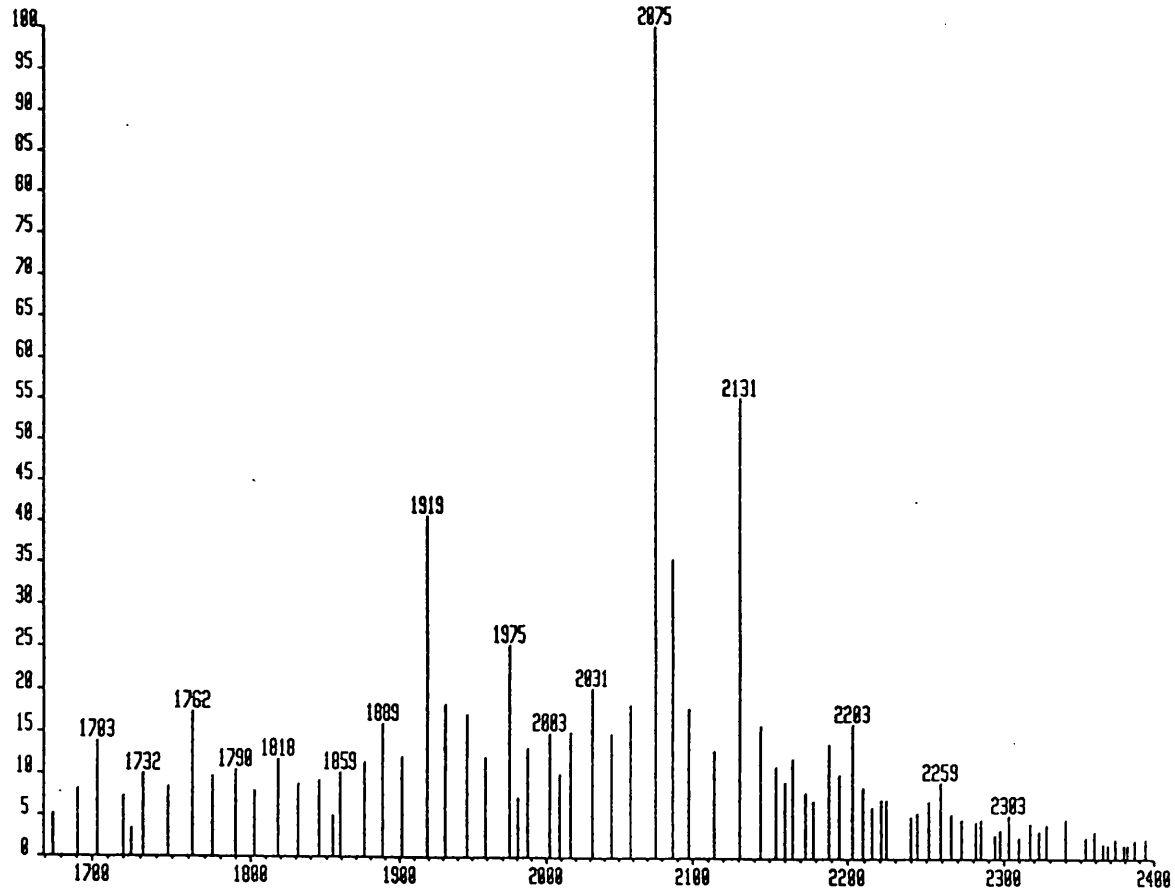


Figure 5.7 Liquid SIMS spectrum of synthetic AcRRKWQKTGHAVRHGRL (M_r 2074).

The spectrum of the peptide Ac-RRKWQ-KTGHA-VRHGR-L is shown in figure 5.7. As can be seen a number of peaks are evident in addition to the expected protonated molecular ions at m/z 2075, suggesting the presence of various impurities. The ions at m/z 1975, 2131 and 2259 represent successive additions of 56 Da to the ions of m/z 1919, 2075 and 2259 respectively. This mass difference corresponds to the presence of the tertiary butyl moiety which is used to block the side chain hydroxyl groups of Asp, Glu, Ser, Thr and Tyr residues in the FMOC protocol used (see table 5.3). As ^{7}Thr is the only such residue present then the tBu derivative must be here. The ion at m/z 1919 is 156 Da less than the expected pseudomolecular ion at m/z 2075 and thus represents an arginine deficient deletion peptide. The only analytical data supplied with this commercially obtained peptide was an amino acid composition which confirmed the expected sequence. The conditions employed in amino acid analysis to hydrolyse the peptide bonds result in the hydrolysis of the tertiary butyl groups as well, rendering their detection impossible.

The two examples discussed above represent common problems associated with peptide synthesis. The most common in the authors experience is the presence of incompletely removed blocking groups such as tertiary butyl. These cannot be detected by amino acid analysis or by Edman sequencing, due to the harsh chemistry required in these techniques which results in their removal. Other common modifications include the presence of deletion peptides and, more rarely, hydrolyis products. The detection of these is usually not possible by amino acid analysis or HPLC. The use of mass spectrometry has been shown to be essential in the detection of all of these modifications. Biemann and Scoble (1987) have also noted the use of MS in detecting the cyclisation of aspartic acid residues in apparently pure synthetic

peptides.

Without the use of mass spectrometry the purity of synthetic peptides cannot be guaranteed and, as noted in the introduction, it follows that conclusive experimental results based on the use of poorly characterised material cannot be guaranteed either.

CHAPTER SIX

THE DETECTION OF CHIRALITY BY FAB-MS

6.0 Introduction

Chirality is an important determinant of molecular recognition in biochemical systems, one enantiomer of a molecule showing biological activity whilst the other enantiomer may show little or none. For example, the naturally occurring amino acids are invariably the L enantiomers in human biochemistry, whereas D forms may be active in non-mammalian species.

Although mass spectrometry is an important technique for structural analysis, it has not played a major role in the differentiation of chiral isomers (Mandelbaum *et al*, 1983). The fragmentation of an ion is not affected by a single assymetric site, and thus D and L forms give identical spectra. However, the presence of two or more assymetric centres introduces more opportunity for differentiation as the differences in the three dimensional structures enhance the likelihood of one of the isomers adopting a more energetically favourable conformation due to its intramolecular bonding interactions. As the energy differences will be small these differences will be more readily observed in the low energy ions which are accessible for study using tandem mass spectrometric techniques. Thus Tabet *et al* (1985) were able to discriminate between enantiomers of a derivatised dipeptide, N-acetyl-Phe-Phe methyl ester, using B/E linked scanning and EI ionisation.

Apart from the use of polarimetry (Purdie and Swallows, 1989) and nmr spectroscopy (Dobashi *et al*, 1986), the main approach to the discrimination of chiral compounds is through interaction with a chiral substrate, this being the basis of chromatographic techniques employing a chiral stationary phase. Electron impact mass spectrometry involves unimolecular processes, and so offers no opportunity for interaction with a chiral substrate. However, softer ionisation techniques are either

bimolecular, as with chemical ionisation or occur in the condensed phase, as in fast atom bombardment. Thus they can be utilised to effect interactions between chiral species with a chiral substrate.

The first report of such a phenomenon was the non-statistical formation of protonated dimers between a racemic mixture of D- and L- diisopropyl tartrates (DIPT) using a CI source (Fales and wright, 1977). Deuterium labelling was used to distinguish the D and L forms, and it was demonstrated that there was no significant isotope effect, but there was a significant chiral effect with the formation of homochiral dimers being preferred to that of the heterochiral dimers. This CI work was extended by using a more quantitative and theoretical treatment, and with the differentiation of additional related compounds (Winkler *et al*, 1986). The differentiation of enantiomers of methionine and tryptophan by CI using l-amyl alcohol as the reagent gas has also been reported (Suming *et al*,1986). Recently, stereospecific reactions of secondary alcohols with diacetoxysuccinic anhydride have been utilised to determine chirality (Yang and Chen, 1992).

The chiral effects with DIPT have been reinvestigated, this time in the condensed phase using FAB mass spectrometry which offers the theoretical possibility of dimer discrimination. The discrimination of chiral hydroxylated amino acids has also been attempted, along with certain derivatives.

6.1 Experimental

FAB mass spectra were obtained using a VG ZAB-SE mass spectrometer operating at 8kV accelerating voltage, coupled with a VG 11/250 data system. An Ion Tech saddle field gun gave a beam of xenon atoms of 8keV energy, which was

used to bombard the samples.

The dialkyl tartrate mixtures of DIPT, DET, DMT and the serine isopropyl ester were gifts from Dr F.J.Winkler (University of Munich)

Derivatisation of the serine isopropyl esters to form tert-butoxycarbonyl amides was as follows. To a solution of 50mg (0.27 mol) of L-serine isopropyl ester hydrochloride in 1.5ml dimethylformamide (DMF), 25μl (0.27 mol) of N,N-diisopropylethylamine was added to release the N-terminal amine moiety (NaOH cannot be used because it would probably destroy any putative chiral effect, as discussed in section 6.2.4). An equimolar amount of di-tert-butyl carbonate (80mg, 0.27 mol) was then added and the reaction allowed to proceed at room temperature with constant stirring for two hours to give the required t-BOC derivative according to the following equation:

$((CH_3)O.CO)_2O + NH_2.CH(CH_2OH).CO.O.CH(CH_3)_2 \rightarrow$

$(CH_3)_3C.O.CO.NH.CH(CH_2OH).CO.O.CH(CH_3)_2 + (CH_3)_3C.O.COOH$

DMF was removed under reduced pressure using a rotary pump. 5ml of 1:1 water/ethyl acetate was added and the organic layer extracted, with the aqueous layer further extracted with 2 x 2.5ml ethyl acetate. The same procedure was adopted to derivatise the deuterated D isomer. After rotary evaporation of the combined organic extracts clear viscous liquids were formed for both isomers. The molecular weights of the two isomeric derivatives as determined by FAB-MS analysis ($[M+H]^+$ at m/z 247 (L) and 254 (D)) confirmed that derivatisation had occured.

To check that the derivatisation procedure had not induced racemisation the individual isomers were analysed by NMR using a Bruker 500 MHz spectrometer linked to an Aspect 3000 data system. The individual species were dissolved in

$CDCl_3$ and run at 20°C using tetramethylsilane as internal standard.

6.2.0 Results

The FAB spectrum of a 1:1 mixture of ($^2H_0$)D-DIPT and the deuterium labelled enantiomer ($^2H_{14}$)L-DIPT can be seen in figure 6.1. The protonated dimers can be seen clearly at m/z 469(DD), 483(DL) and 497(LL) indicating a clear preference for the formation of the heterochiral dimers. To establish that there was no significant isotope effect the spectrum of a 1:1 mixture of ($^2H_0$)L,($^2H_{14}$)L-DIPT was also obtained, and as can be seen in figure 6.2 no chiral discrimination is observed as the expected statistical ratio of 1:2:1 is evident for the protonated dimers. The chiral effect can be quantified using an equation due to Winkler:

$$Chiral\ effect = \frac{K_{Homochiral\ dimers}}{K_{Heterochiral\ dimers}} \qquad [6.1]$$

$$= \frac{\sqrt{[DDH^+]\ [LLH^+]}}{[DLH^+]} \qquad [6.2]$$

As there is a theoretical binomial distribution due to the two means of forming the heterochiral dimer (DLH or LDH) a factor of two is incorporated in the numerator to give:

$$Chiral\ effect = \frac{2\sqrt{[DDH^+]\ [LLH^+]}}{[DLH^+]} \qquad [6.3]$$

Thus if a species shows no chiral discrimination in the formation of its dimers and the mixture is a racemate [DDH] = [LLH] = 1/2[DL(LD)H] and from equation [6.3] the chiral effect equals unity.

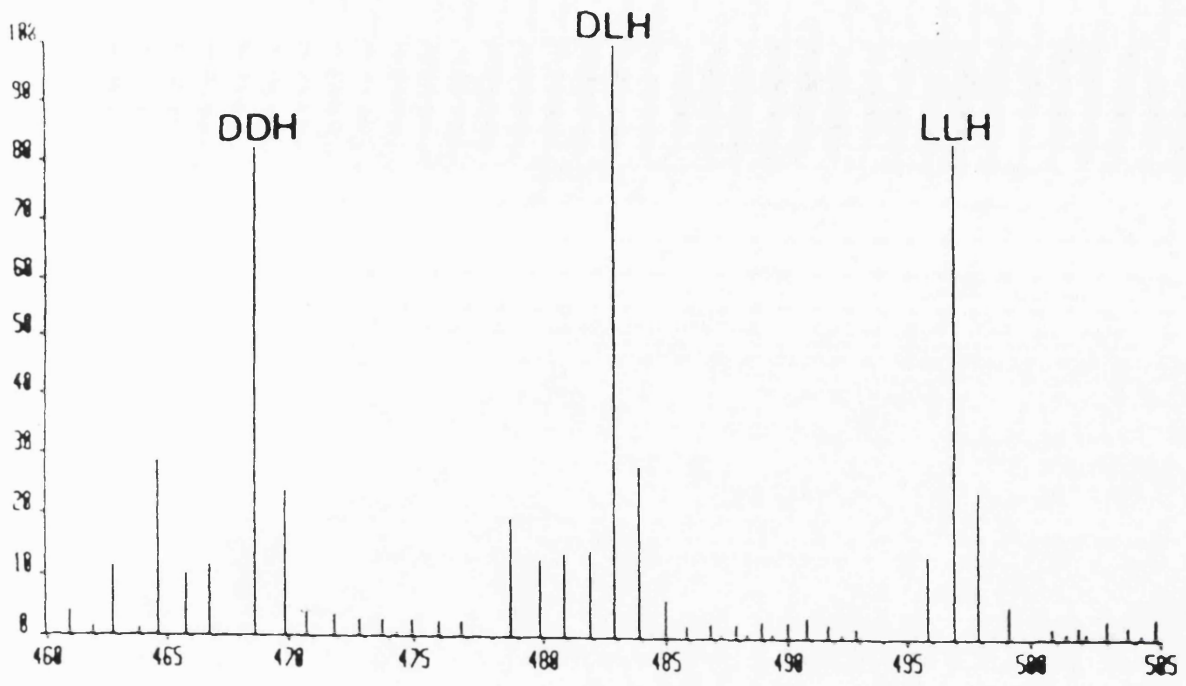


Figure 6.1 Partial FAB mass spectrum of ($^2H_0$)D,($^2H_{14}$)L-DIPT showing dimers

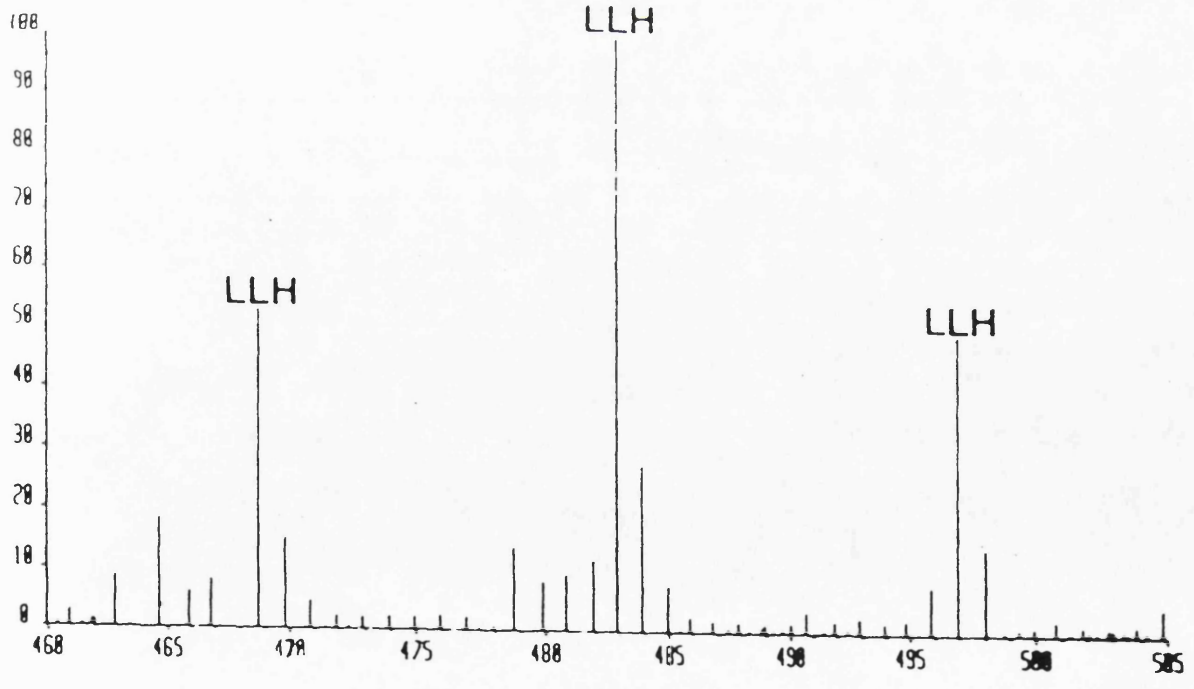


Figure 6.2 Partial FAB mass spectrum of ($^2H_0$)L,($^2H_{14}$)L-DIPT.

The chiral effects for the DIPT species noted earlier are tabulated in Table 6.1, together with the results from similar experiments carried out in duplicate with diethyl tartrate and dimethyl tartrate.

6.2.1 Unimolecular fragmentation

The stability of the protonated dimers was investigated by studying their unimolecular (metastable) fragmentation patterns in the first field free region of the mass spectrometer using B/E linked scanning. As shown in figures 6.1 and 6.2 the ($^2H_0$),($^2H_{14}$)-DIPT mixtures give three peaks at m/z 469, 483 and 497. Unimolecular decomposition of these gives protonated monomers, m/z 469 giving m*235, m/z 483 giving m*235 and m*249, and m/z 497 giving m*249. The relative intensities of the metastable peak heights normalised to their respective precursor ion intensities were then compared, summing the m*235 and m*249 in the case of the heterochiral prcursor ion. If the protonated dimers were of equal stability, and there were no isotope effects, equal degrees of unimolecular decomposition would be anticipated, i.e.

$$\frac{m*235}{469} : \frac{m*235 + m*249}{483} : \frac{m*249}{497} = 1 : 1 : 1$$

In experiments carried out in duplicate the ($^2H_0$)L,($^2H_{14}$)L-DIPT mixture gave a ratio close to 1.0 : 1.0 : 1.0, whereas the ($^2H_0$)D,($^2H_{14}$)L-DIPT gave a ratio of 1.0 : 1.5 : 1.0 (spectra of the latter are shown in figure 6.3). The errors on these duplicate experiments were about $\pm$ 0.1, but they clearly demonstrate the increased tendency for unimolecular breakdown of the heterochiral dimer as compared with the homochiral dimer.

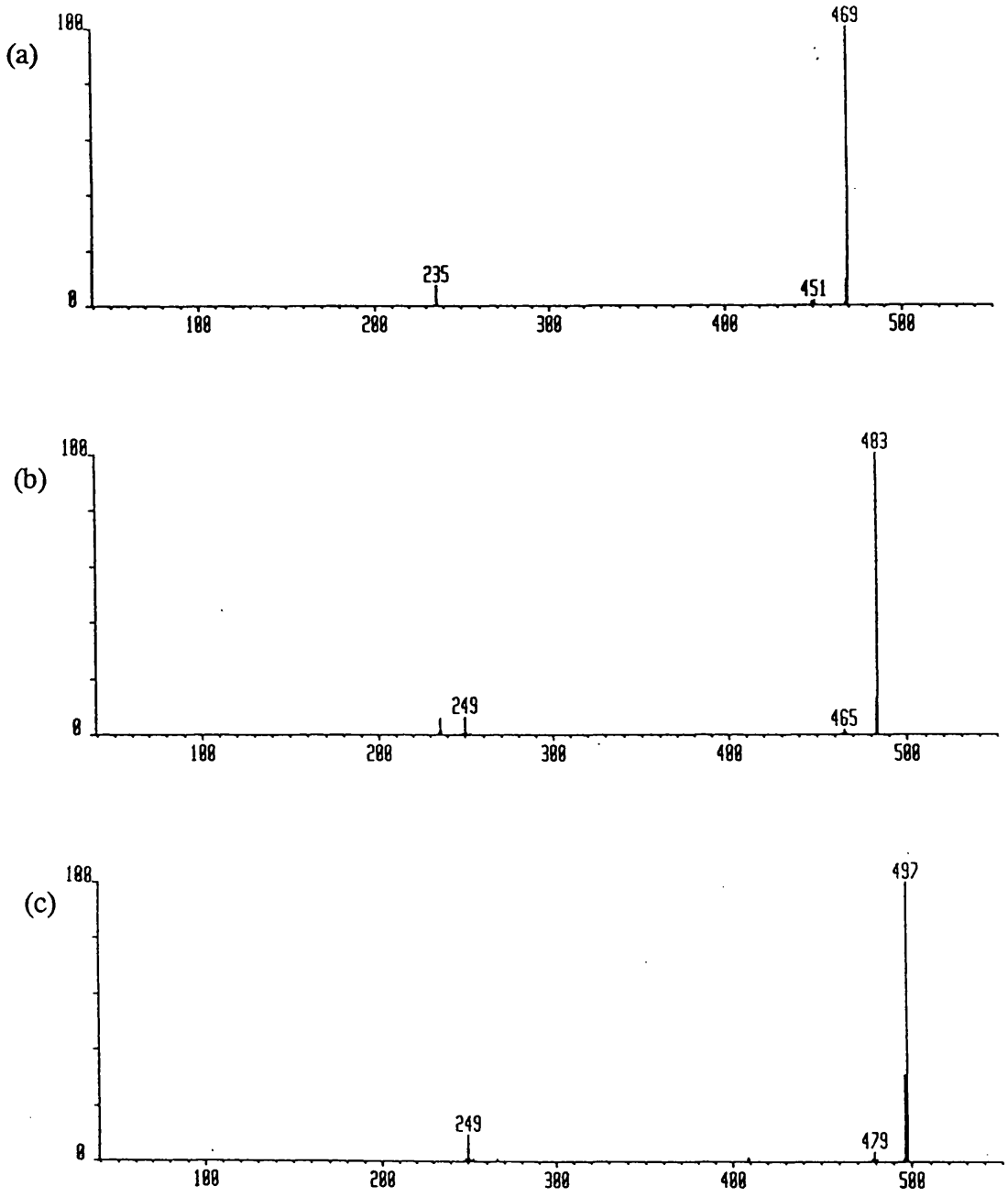


Figure 6.3 B/E linked scans showing unimolecular decomposition daughter ions of (a) DDH^+ (b) DLH^+ (c) LLH^+ dimers from ($^2H_0$)D,($^2H_{14}$)L-DIPT

A similar experiment using ($^2H_0$)D,($^2H_{10}$)L-DET, also duplicated gave a discrimination of 1.0 : 1.3 : 1.0, again $\pm$ 0.1. As with the chiral effect noted with the parent ions a decrease in discrimination is seen as the alkyl chain size is decreased.

6.2.2 Mechanistic Interpretation of DIPT chiral effect.

The experiments performed clearly indicate close agreement between the chiral effects observed using FAB ionisation and the earlier results by CI. The relative instability of the heterochiral protonated dimers is evidenced by the chiral effects noted in table 6.1. This instability was explained earlier in terms of a weak *gauche* interaction of the alkyl groups in the heterochiral dimer, as shown in Figure 6.4. The extent of the discrimination depends on the length of the alkyl chain as (see table 6.1). The non-linearity of the effect may be due to the fact that the isopropyl chain was used as opposed to the straight chain propyl isomer. However a clear progression is seen with an increasing chiral effect in the order methyl < ethyl < isopropyl.

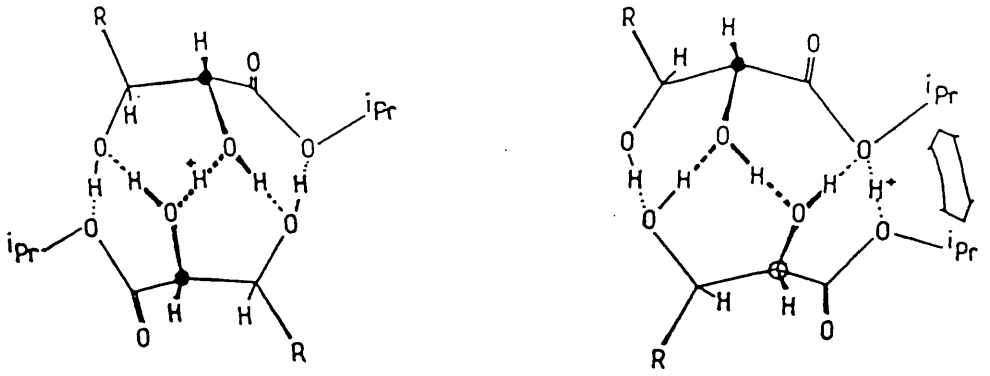
The difference in stability can be quantified in terms of a difference in the free energies of the dimers:

$$\Delta G = - R\,T\,lnK \qquad [6.4]$$

$$\therefore\ \Delta\Delta G = - R\,T \ln (chiral\ effect) \qquad [6.5]$$

From the FAB data the difference in free energy for the homo- and heterochiral dimers of DIPT is thus calculated to be 1.2 kJmol^{-1}, identical to the value calculated

Table 6.1 Internal chiral effects for the formation of homo- and heterochiral dimers of DMT, DET and DIPT, and external chiral effects for DET/DIPT adducts. Chemical ionisation data taken from Winkler *et al* (1986).			
		FAB	CI
Internal	$(^2H_0)$D,$(^2H_{14})$L-DIPT	1.63 ± 0.03	1.71 ± 0.04
	$(^2H_0)$L,$(^2H_{14})$L-DIPT	1.06 ± 0.10	1.01 ± 0.05
	$(^2H_0)$D,$(^2H_{10})$L-DET	1.57 ± 0.03	
	$(^2H_0)$D,$(^2H_6)$L-DMT	1.35 ± 0.04	
External	L-DET	1.50 ± 0.02	1.46 ± 0.10
	D-DET	1.45 ± 0.07	1.55 ± 0.10



homochiral dimer LLH^+ heterochiral dimer DDH^+

Figure 6.4 Theoretical structures of DIPT protonated dimers (Winkler *et al*, 1986)

for the earlier CI experiment (Winkler, 1986)

6.2.3 Ammonium / alkali metal cation bound dimers.

When ammonium acetate was added to the DIPT matrix, the chiral effect was annulled, as seen in figure 6.5 with the statistical 1 : 2 : 1 ratio manifested in the ammonium adducts. From a mechanistic perspective this may be surmised as being due to the rupturing of the 5-bond zip-type hydrogen bonding motif. Addition of potassium, sodium and lithium cations also cancelled the chiral effect.

6.2.4 External Chiral Effect.

When D-diethyl tartrate was added to the $(^2H_0)$D,$(^2H_{14})$L DIPT a chiral discrimination was observed in the adducts formed (D-DET,D-DIPT = "MDH"; D-DET,L-DIPT = "MLH") as seen in figure 6.6. This effect is termed an "external" chiral effect . A clear preference for the formation of the "homochiral" adduct is evident as may be expected from the homochiral preference in the generation of the "internal" dimers discussed earlier.

Quantification of this phenomenon can be achieved in like manner to the internal effect:-

$$\textit{External Chiral Effect} = \frac{K_{MDH^+}}{K_{MLH^+}} \qquad [6.6]$$

$$= \frac{[MDH^+]\,[LH^+]}{[DH^+]\,[MLH^+]} \qquad [6.7]$$

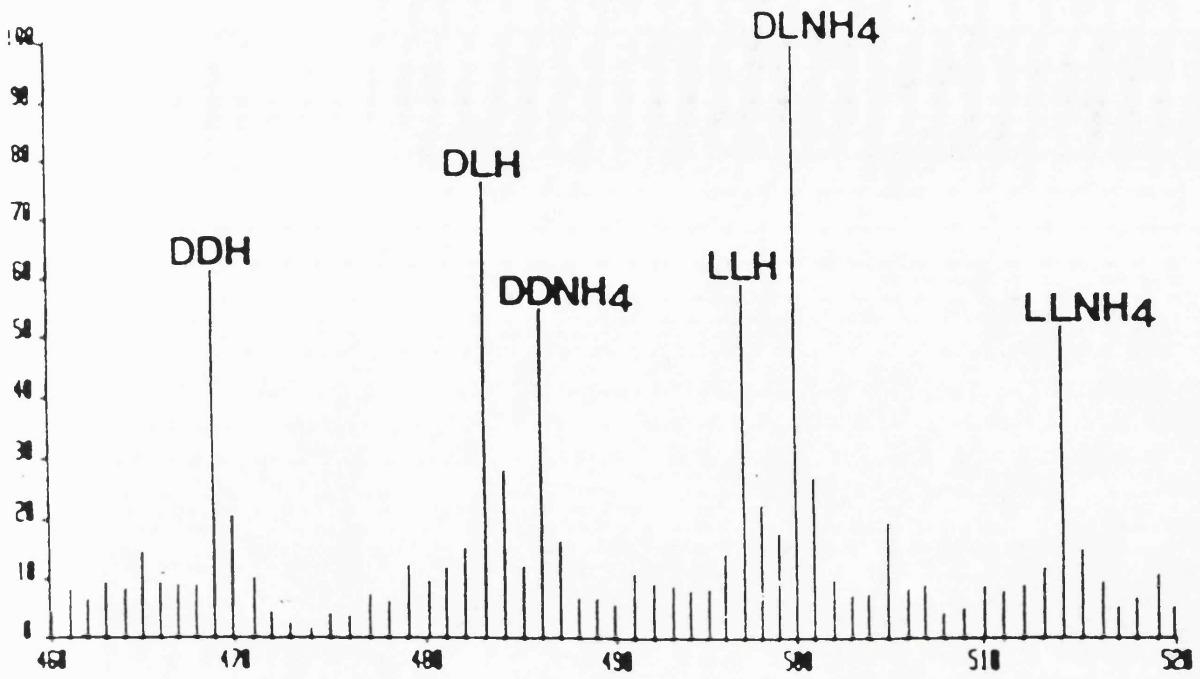


Figure 6.5 Partial FAB mass spectrum of ($^2H_0$)D,($^2H_{14}$)L-DIPT plus ammonia

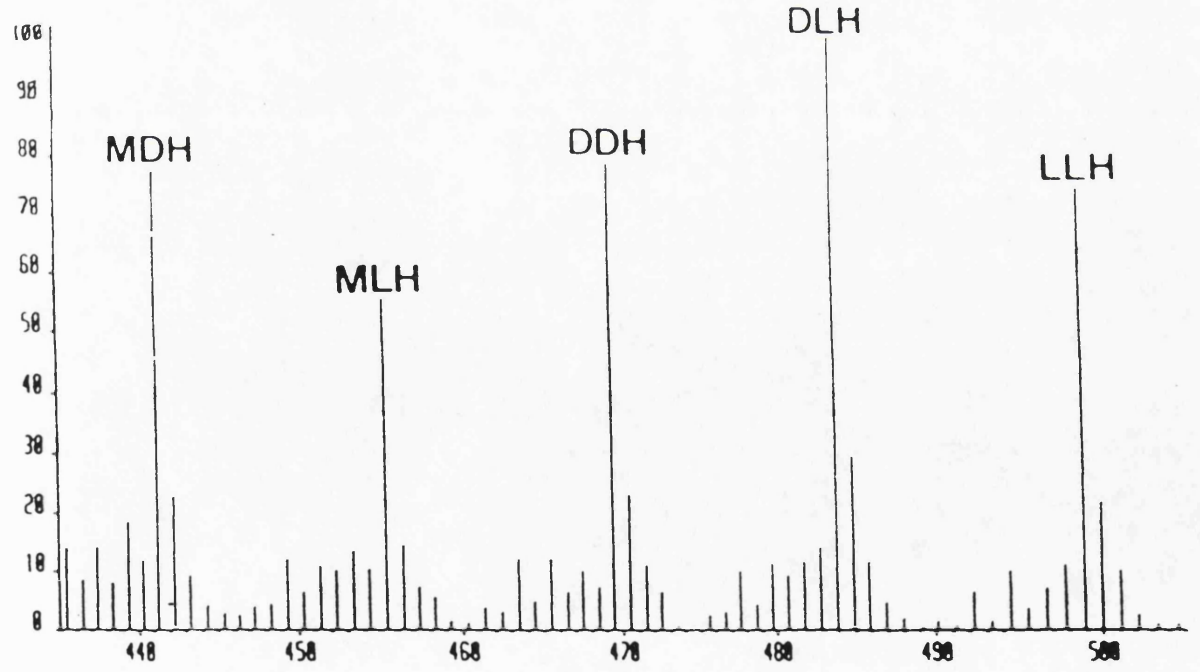


Figure 6.6 Partial FAB mass spectrum of D-DET plus ($^2H_0$)D,($^2H_{14}$)L-DIPT.

The external chiral effects due to the addition of D- and L- DET are noted in Table 6.2.

The use of the ($^2H_0$)D,($^2H_{14}$)L DIPT in this manner, to detect chirality in an external substrate, has implications for the development of "chiral matrices" which can be used for a given class of compounds.

6.2.5 Titration of mixtures.

The external effect was also used to generate a calibration curve for titration of mixtures of D-DET and L-DET in DIPT. This is presented in figure 6.7, in which the observed external chiral effect as calculated from Eqn. 6.7 is plotted against the percentage of D-DET in the mixture. Each point in the graph represents the average of determinations carried out in duplicate sets of 5-10 averaged scans. A linear relationship was obtained with a regression coefficient of 0.997.

6.2.6 Attempted discrimination of amino acids.

Having established a protocol for the analysis of one class of compounds, it was hoped to extend the methodology to differentiate between enantiomers of the more biologically significant class of the amino acids.

From the hydrogen bonding model described previously it seemed that the hydroxylated amino-acids, serine and threonine, offered the greatest potential for resolution. Using the concept of the external chiral effect, L-threonine was added to a 1:1 mixture of ($^2H_0$)D, ($^2H_{14}$)L DIPT chiral matrix with the aim of forming threonine-DET adducts which differed in the intensities of the L-L versus the L-D. The spectrum can be seen in figure 6.8, and unfortunately no differences are

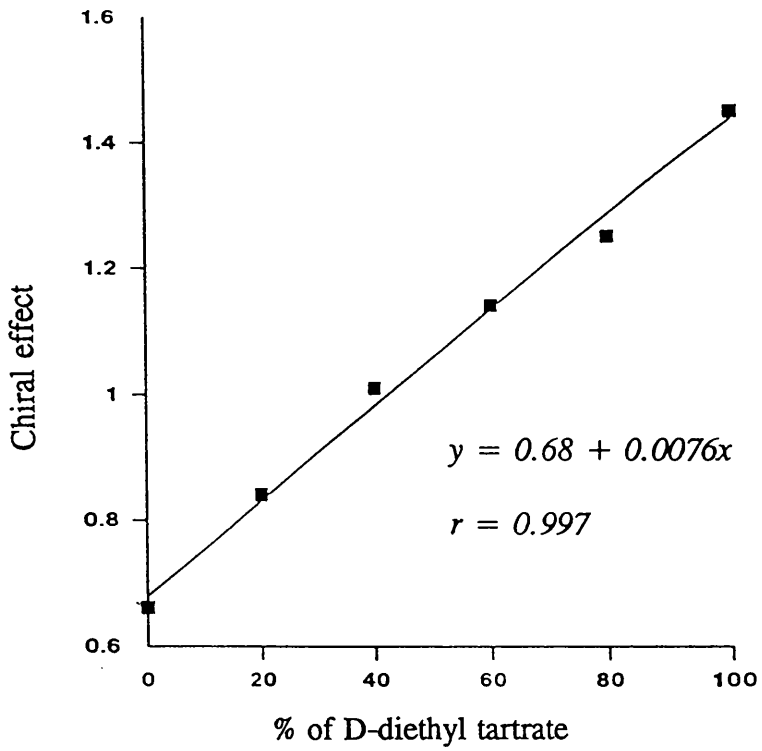


Figure 6.7 Linear regression plot for the quantification of mixtures of D- and L- DET in ($^2H_0$)D,($^2H_{14}$)L-DIPT matrix.

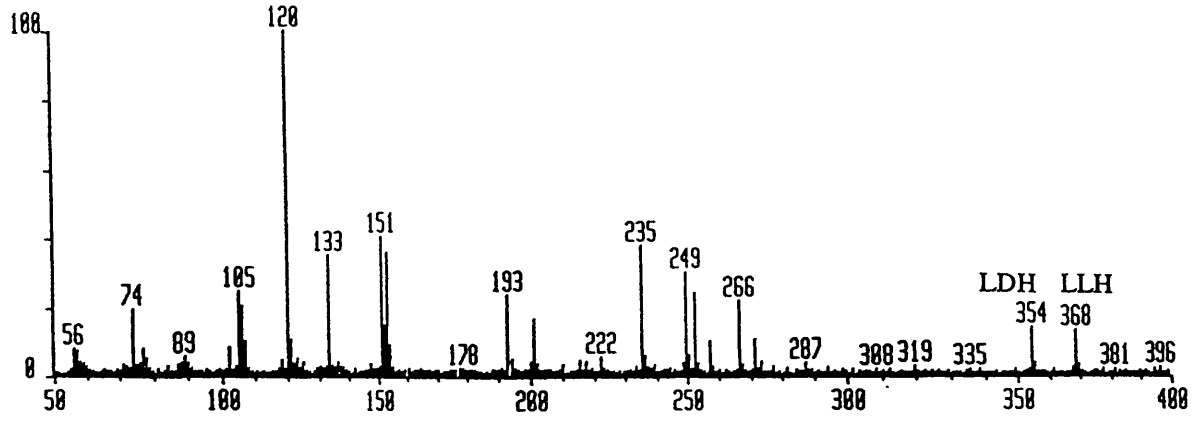


Figure 6.8 FAB spectrum of L-threonine in($^2H_0$)D,($^2H_{14}$)L-DIPT matrix.

observed.

6.2.7 *Isopropyl derivatisation of serine.*

The reason for the lack of a chiral discrimination may have been due to the absence of bulky groups which gave rise to the manifestation of the effect with the alkyl tartrate ester derivatives. Thus serine enantiomers were esterified with either unlabelled or deuterium labelled isopropyl groups at their carboxy termini, to produce species which would hopefully exhibit differentiation in their dimers due to steric hindrance.

A 1:1 mixture of the ($^2H_0$)L and ($^2H_7$)D isopropyl serine esters was analysed in glycerol matrix to study the possibility of an internal chiral effect being manifested. Homo and heterochiral dimers were apparent at m/z 295, 302 and 309 as depicted in figure 6.9. Unfortunately no internal chiral discrimination between the dimers was forthcoming with a ratio of 1:2:1 being observed.

6.2.8 *Tertiary butyl-oxy-carbonyl (t-BOC)- isopropyl derivatives of serine.*

From the clear formation of dimers evidenced in figure 6.9 it would appear that some type of hydrogen bonding mechanism is occuring. This may be similar to that described earlier but it may be that the steric hindrance is still not sufficiently great to enable differentiation. As it would be a simple step to synthesise t-BOC derivatives on the amine termini of the isopropyl derivatives, it was decided to adopt this strategy in the hope of producing the steric hindrance required. As can be seen in figure 6.10, no discrimination was forthcoming in the dimers observed. The integrity of the samples was evident from their ^{1}H-nmr spectra, the($^2H_0$)L isopropyl

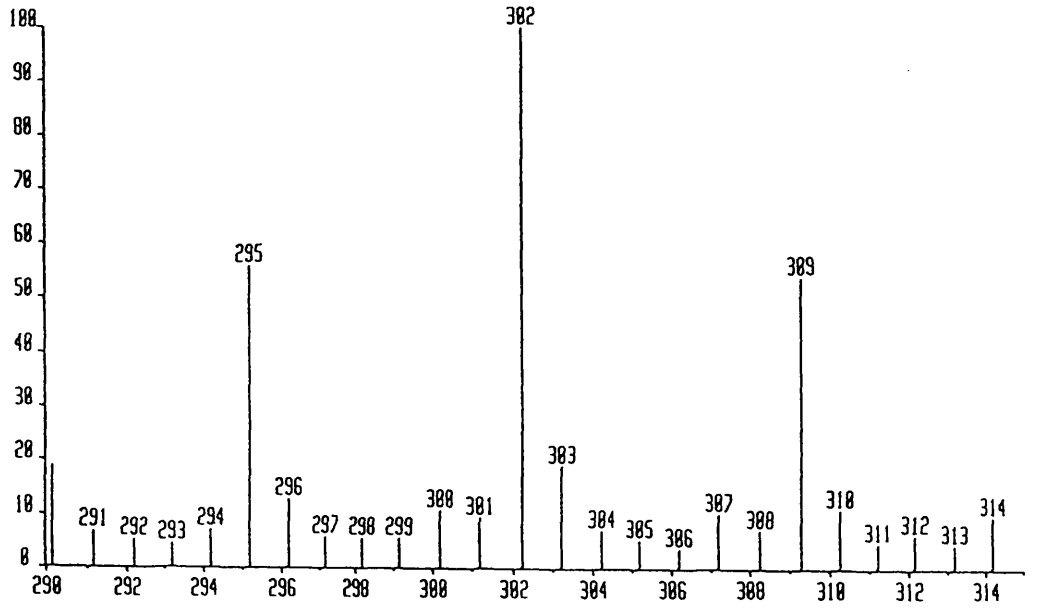


Figure 6.9 Partial FAB spectrum of ($^2H_0$)L,($^2H_7$)D serine isopropyl esters showing dimers

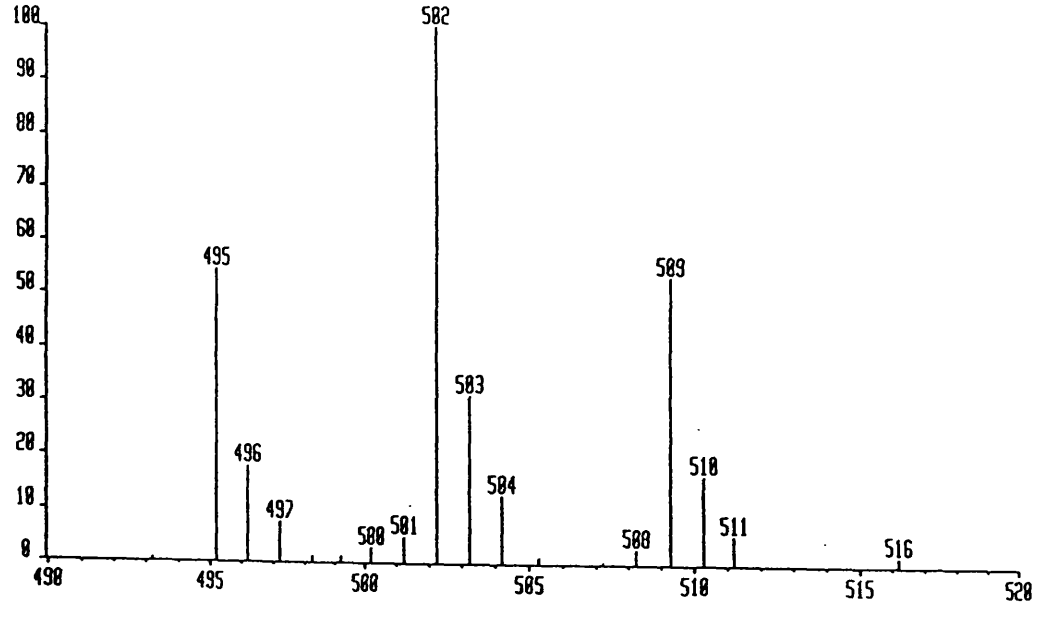


Figure 6.10 Partial FAB spectrum of ($^2H_0$)L,($^2H_7$)D N-tBOC serine isopropyl esters

serine ester gave a signal due to -NH- protons at δ=5.5 whilst the equivalent signal for protons of the ($^2H_7$)D isomer was at δ=5.7.

It remains a possibility that other blocking groups may induce discrimination in dimer formation of these species, however it is more probable that the hydroxylated amino acids do not form the cage like structures envisaged for the dialkyl tartrates.

6.3.0 Conclusions

It has been shown that discrimination can be effected between chiral isomers of various alkyl tartrates. This is explained in terms of the formation of tightly hydrogen bonded cage like structures. Addition of various cations results in annuling the chiral discrimination, explained by the destruction of the hydrogen bonded structures. The fact that chiral discrimination is observed with neither threonine nor serine ester derivatives is also assumed to be due to the inability to form the tightly bonded caged dimers. It is predicted that any class of compounds which are capable of forming dimeric structures similar to those depicted in figure 6.4 will exhibit a chiral effect.

CHAPTER SEVEN

CONCLUSIONS

The main aim of this thesis has been to utilise mass spectrometry, in an integrated analytical approach with more conventional techniques, to determine the structures of a variety of proteins and peptides. All of the species studied were intractable to conventional methods of protein structural analysis to some degree. A very high degree of success was achieved in fulfilling this aim and it is apparent that mass spectrometry has an important role to play in this field, particularly in elucidating post translational modifications. These cannot be detected by gene sequencing or Edman sequencing and whilst techniques for their analysis other than mass spectrometry exist, in general they lack the combined benefits of versatility, unambiguity, sensitivity and speed attainable by mass spectrometric methods for the examples quoted in this thesis. When it is considered that at least 80% of the approximately 100,000 proteins expressed in the cells of eukaryotic species are postulated to possess post translational modifications, the role of mass spectrometry in protein sequence determination may even be described as essential.

The sequencing and subsequent identification of isoforms of annexin V present in bovine brain was described and represents an example of efficient sequencing methodology exploiting advances over a wide spectrum of bioanalytical techniques. The integrated utilisation of HPLC, SDS-PAGE, western blotting, Edman sequencing, electrospray MS and FAB MS resulted in the determination of the primary structures of these isoforms. A suspected post translational modification, due to a mass difference of 4,000 Da between the two isoforms obtained by SDS-PAGE and the similarity of the isoforms by Edman sequencing, was shown to be due to an aberrant electrophoretic effect caused by a glu/lys sustitution in the isoforms. Without the use of electrospray MS this conclusion would have been reached with considerably more

difficulty.

The adoption of electrospray MS was described as a means of assessing the validity of published sequences of the electron transport protein plastocyanin from the blue green alga *Anabaena variabilis* and the cyanobacterium *Scenedesmus obliquus*. The sequence of plastocyanin from the former was confirmed, whilst the latter structure needed correction, due to the presence of an extra valine near the C-terminus and a typographical error in which lysine was substituted for a threonine residue. In general C-terminal analysis presents a problem area for conventional sequencing strategies and it is evident from this example that a role for the use of mass spectrometry exists.

N-terminal blocking is the most common class of post translational modification. Acetylation is by far the most widespread member of this class and was established as the mode of N-terminal blocking for the isoforms of bovine brain annexin V and thioesterase II from rat mammary gland. The primary structures of thymosins β_{11} and a novel isoform, designated β_{12}, from trout spleen have been determined and were also shown to possess N-terminal acetylation.

N-terminal blocking in the form of myristoylation was shown to be amenable to mass spectrometric analysis for a synthetic peptide derived from the oncogenic protein $pp60^{src}$. An unusual degree of fragmentation covering the entire sequence was noted in the normal liquid SIMS spectrum. This behaviour, whilst atypical of peptides in general, may be a characteristic of myristoylated peptides.

Cyclisation of N-terminal glutamine to form pyroglutamyl, as well as the processing of C-terminal glycine to produce C-terminal amidation, have been determined to occur in a novel tripeptide, pyroGlu-Phe-Pro-amide, present in human

seminal fluid. Again an integrated approach using mass spectrometry with Edman sequencing was adopted. A study was carried out to assess the feasibility of adopting liquid SIMS MS with B/E linked scanning to determine the sequence of a similar tripeptide present in rabbit prostate complex. Using synthetic peptides it was shown that at least 2 nmol of endogenous peptide are required to enable the structure to be deduced from two possible sequences. In the light of this study it has been decided that the protocol is feasible and work is underway to isolate the required amount of material.

From the study of the phosphorylation of a synthetic substrate of protein kinase C it is concluded that the stochiometry determined by liquid SIMS MS is more accurate than the value obtained by the commonly adopted approach of binding radioactive phosphopeptides to paper. The main disadvantage of the latter technique is that phosphopeptides tend to be removed from the paper in the sample preparation. As there is no alternative method of sample preparation at present, it would appear that liquid SIMS analysis affords an important alternative for determining phosphorylation stochiometries - however possible problems due to suppression effects must be born in mind.

Lipoylation is a relatively rare form of prosthetic post translational modification and is known to occur in the E2 component of pyruvate dehydrogenase complex. The sites of lipoylation have been determined on proteolytic fragments of E2 from *E.coli*, using FAB MS and Edman sequencing. It was concluded that heterogeneity exists in the E2 gene.

The characterisation of synthetic peptides was described and demonstrated to present difficulties which cannot be resolved using the commonly adopted protocol

of amino acid analysis and HPLC profiling. These difficulties, which are mainly caused by the presence of deletion, partially deprotected or hydrolysed products, were resolved by the utilisation of liquid SIMS MS. No other analytical technique affords the analytical power required and the adoption of mass spectrometric analysis is concluded to be essential in this field.

In summary it has been demonstrated in the various examples discussed above that mass spectrometry has an important and often essential role to play in the elucidation of protein and peptide primary structures. This role is at present complementary to the powerful techniques of gene sequencing and Edman sequencing, due to their inability to determine the modifications to the common twenty amino acids which are suspected to occur in 80% of endogenous proteins.

The remainder of the thesis focussed on the use of FAB-MS to discriminate chiral isomers. Clear resolution was obtained for D and L isomers of tartaric acid esters of varying chain length due to dimer effects. There is a fundamental requirement for deuterated analogues to study these effects. Application of the methodology to determine the chirality of threonine and derivatives of serine was not successful. The technique was seen to be rather limited due to the need to form tightly bound dimers in which sufficiently strong enantiomer dependent steric interactions between dimers are produced.

REFERENCES

Aberth W. and Burlingame A.L. (1984) *Anal. Chem.* 56 2915

Aberth W. and Burlingame A.L. (1990) in *Biological Mass Spectrometry* (Burlingame A.L. Ed) Elsevier (Amsterdam)

Aitken A., Geisow, M.J., Findlay, J.B.C., Holmes, C. and Yarwood, A. (1989) in *Protein Sequencing: A Practical Approach* (Geisow, M.J. and Findlay, J.B.C., eds) IRL Press Oxford, pp 43-68

Aitken A. (1990) in *Identification of Protein Consensus Sequences* Ellis-Horwood Ltd., Chichester; Simon and Schuster, New York, pp 1-167.

Aitken A. (1989) in *Focus on Laboratory Methodology in Biochemistry* (Fini, C. Floridi, A., Finelli, N.V., and Wittman-Liebold, B., eds.) vol. 1, p.p. 9-33. CRC Press Inc., Boca Raton, Florida.

Aitken A. (1988) *Meth. Enzymol.* 167 145

Aitken A., Cohen P. and Klee C.B. 1984) *Eur. J. Biochem.* 139 663

Albert K.A., Nairn, A.C. and Greengard P. (1987) *Proc. Natl. Acad. Sci. USA* 84 7046

Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J.D. (1989) *Molecular Biology of the Cell.* Garland (London)

Amess B., Manjarrez-Hernandez H.A., Howell S.A., Learmonth M. and Aitken A. (1992) *FEBS Lett.* 297 285

Amons R. (1987) *FEBS Lett.* 212 68

Arfin S.M. and Bradshaw R.A. (1988) *Biochemistry* 27 7979

Atherton E and Sheppard R.C. (1989) *Solid Phase Synthesis: A Practical Approach* IRL press (Oxford)

Bachmair A., Finley D. and Varshavsky A. (1986) *Science* 234 179

Bailey J.M. and Shively J.E. (1990) *Biochemistry* 29 3145

Baldwin M.A., Howell S.A., Welham K.J. and Winkler F.J. (1988) *Biol. Mass Spectrom.* 16 357

Barber M., Bordoli R.S., Sedgwick R.D. and Tyler A.N. (1981) *J. Chem. Soc. Chem. Commun.* 325

Barber M. and Green B.N. (1987) *Rapid Commun. Mass Spectrom.* 1, 80

Bateman A., Solomon S. and Bennett H.P.J. (1990) *J. Biol. Chem.* 265 22130

Bateman R.H., Burns P., Owen R. and Parr V.C. (1985) *Proc. 10th Int. Mass Spectrom. Conf.* 863

Bateman R.H. and Butt P.A. (1987) *Proc. 35th ASMS Conf. Mass. Spectrom. Allied Topics*, 253

Bazzi, M.D. and Neslestuen, G.L., (1991) *Biochemistry* 30 971.

Benne R., Van den Burg J., Brakenhoff J.P.J., Sloof P., Van Boom J.H. and Tromp M.C. (1986) *Cell* 46 819

Benninghoven A. and Sichtermann W.K. (1977) *Org. Mass Spectrom.* 12 595

Benninghoven A. and Sichtermann W.K. (1978) *Anal. Chem.* 50 1180

Biemann K. and Scoble H.A. (1987) *Science* 237 992

Biemann K. and Martin S.A. (1987) *Mass. Spectrom. Rev.* 6 1

Biemann K. (1990) *Meth. Enzymol.* 193 455

Blackshear P.J., Wen L., Glynn B.P. and Witters L.A. (1986) *J. Biol. Chem.* 261 1459

Bleile D.M. Hackert M.L. Pettit. F.H. and Reed L.J. (1981) *J. Biol. Chem.* 256 514

Boulter D., Haslett B.G., Peacock D., Ranshaw J.A. and Scawen M.D. (1977) *Plant Biochem.* 13 1

Boustead, C.M., Walker, J.H. and Geisow, M.J., (1988) *FEBS Lett.* 233 233.

Bradford A.P., Howell S.A., Aitken A., James L.A. and Yearman S.J. (1987) *Biochem. J.* 245 919

Budzikiewicz H., Djerassi C. and Williams D.H. (1967) *Mass Spectrometry of Organic Compounds* San Francisco: Holden Day

Bull H.B. and Breese K. (1974) *Arch. Biochem. Biophys.* 161 665

Burgoyne, R.D. and Geisow, M.J., (1989). *Cell Calcium* 10 1

Burlingame A.L., Millington D.S., Norwood D.L., Russell D.H. (1990) *Anal. Chem.* 62 268R

Cantley C.L. (1991) *Cell* 64 281

Carr C., Tyler A.N. and Cohen J.B. (1989) *FEBS Lett.* 243 65

Carr S.A. and Biemann K. (1984) *Meth. Enzymol.* 106 29

Carr S.A., Hemling M.E., Bean M.F. and Roberts G.D. (1991) *Anal. Chem.* 63 2802

Cate R.L. and Roche, T.E. (1979) *J. Biol. Chem.* 254 1659

Chait B.T. and Field F.H. (1986) *Biochem. Biophys. Res. Commun.* 134 420

Chang T.T., Lay J.O. and Francel R.J. (1984) *Anal. Chem.* 56 111

Chow M. Newman J.F.E., Filmar D., Hogle J.M. Rowlands D.J. and Brown F. (1987) *Nature* 327 482

Cockle S.M., Aitken A. and Howell S.A. (1991) *Biochem. Soc. Trans.* 19 3275

Cockle S.M., Aitken A., Beg. F., Morrell J.M. and Smyth D.G. (1989b) *FEBS Lett.*, 252 113

Cockle S.M., Aitken A., Beg. F. and Smyth D.G. (1989a) *J. Biol. Chem.* 264 7788

Cody R.B. and Freiser B.S. (1979) *Anal. Chem.* 51 547

Collins J.H. and Reed L.J. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74 4223

Cooks R.G. and Busch K.L. (1983) *Int. J. Mass Spectrom. Ion Phys.* 53 111

Cottrell J.S. and Greathead R.J. (1986) *Mass Spectrom. Rev.* 3 357

Cottrell J.S. and Evans S. (1987) *Anal. Chem.* 59 1990

Covello P.S. and Gray M.W. (1989) *Nature* 341 662

Cox K.A., Williams J.D., Cooks R.G. and Kaiser R.E. (1992) *Biol. Mass Spectrom.* 21 226

Craig A.G., Engstrom A., Bennich H. and Kamewsky I. (1987) *Proc. 35th ASMS Conference (Denver)* 528

Creighton T.E. (1983) *Protein structures and Molecular Principles*, W.H. Freeman and Co., New York.

Creutz, C.E., Zaks, W.J., Hamman, H.C., Crane, S., Martin, W.H., Gould, K.L., Oddie, K.M. and Parsons, S.J., (1987) *J. Biol. Chem.* 262 1860

Crompton, M.R., Moss, S.E. and Crumpton, M.J., (1988) *Cell* 55 1

Cross G.A.M. (1987) *Cell* 48 179

Crumpton, M.J. and Dedman, J.R., (1990). *Nature* 345 212.

Dalakas M. and Trapp B.D. (1986) *Ann. Neurol.* 19 349

Danson M.J. and Perham, R.N. (1976) *Biochem. J.* 159 677

Dass C. and Desiderio D.M. (1987) *Anal. Biochem.* 163 52

Dell A. (1990) *Meth. Enzymol.* 193 647

De Marcucci O. and Lindsay J.G. (1985) *Eur. J. Biochem.* 149 641

De Pauw E. (1986) *Mass Spectrom. Rev.* 5 191

De Pauw E. (1990) *Meth. Enzymol.* 193 201

Derrick P.J. and Sundqvist B. (1987) *Int. J. Mass Spectrom. Ion Proc.* 78 ix

Dobashi A., Saito N., Motoyama Y. and Hara S. (1986) *J. Am. Chem. Soc.* 108 307

Dole M., Mack L.L., Hines R.L., Mobley R.C., Ferguson L.D. and Alice M.B. (1968) *J. Chem. Phys.* 49 2240

Donato R., Giambanco I., Pula G. and Bianchi R. (1990) *FEBS Lett.* 262 72

Edman P. and Begg G. (1967) *Eur. J. Biochem.* 1 80

Erickson-Viitanen S. and Horecker B.L. (1984) *Arch. Biochem. Biophys.* 233, 815

Evans S. (1990) *Meth. Enzymol.* 193 61

Fales H.M. and Wright G.J., *J. Am. Chem. Soc.* 99 239 (1977)

Falick A.M. and Maltby D.A. (1989) *Anal. Biochem.* 182 165

Falick A.M., Walls F.C. and Laine R.A. (1986) *Anal. Biochem.* 159 132.

Farries T.C., Harris A., Auffret A.D. and Aitken A. (1991) *Eur. J. Biochem.* 196 679

Fasano, O., Aldrich, T., Tamanoi, F., Taparowsky, E., Furth, M. and Wigler, M. (1984) *Proc. Nat. Acad. Sci.* 81 4008.

Fenn J.B., Mann M., Meng C.K., Wong S.F. and Whitehouse C.M. (1989) *Science* 246 64

Fenselau C. and Cotter R.J. (1987) *Chem. Rev.* 87 501

Fraser H.M. and McNeilly A.S. (1982) *Endocrinology* 111 1964

Friedman M., Krull L.H. and Cavins J.F. (1970) *J. Biol. Chem.* 245 3868

Fujiwara K., Okamura-Ikeda K. and Motokawa Y. (1986) *J. Biol. Chem.* 261 8836

Funakoshi, T., Heimark, R.L., Hendrickson, L.E., McMullen, B.A. and Fujikawa, K., (1987) *Biochemistry* 26 5572

Funakoshi, T., Hendrickson, L.E., McMullen B.A. and Fujikawa, K. (1987) *Biochem. J.* 26 8087.

Gallagher R.T., Chapman J.R. and Mann M. (1990) *Rapid Commun. Mass Spectrom.* 4 369

Geisow, M.J. and Aitken, A. (1989) in *Protein Sequencing: A Practical Approach* (Geisow, M.J. and Findlay, J.B.C., eds) IRL Press Oxford, pp 85-98.

Geisow M.J. (1991) *Trends Biotech.* 9 294

Gibson B.W. and Cohen P. (1990) *Meth. Enzymol.* 193 480

Gohl W., Kutscher R., Lave H.J. and Wollnik H. (1983) *Int. J. Mass. Spectrom. Ion Proc.* 48 411

Gomez-Marquez J., Dosil M., Segade F., Bustelo X.R., Pichel J.G., Dominguez F. and Freire M. (1989) *J. Immunol.* 143 2740

Goodall G.J. and Horecker B.L. (1987) *Arch. Biochem. Biophys.* 256 402

Goodall G.J. Richardson M., Furuichi Y., Wodnar-Filipowicz A. and Horecker B.L. (1985) *Arch. Biochem. Biophys.* 236 445

Gower J.L. (1985) *Biol. Mass Spectrom.* 12 191

Graff J.M., Stumpo D.J. and Blackshear, P.J. (1989) *J. Biol. Chem.* 264, 11912

Graff J.M., Rajan R.R., Randall R.R., Nairn A.C. and Blackshear P.J. (1991) *J. Biol. Chem.* 266 14390

Grand R.J.A. (1989) *Biochem. J.* 258 625

Green B.N., Oliver R.W.A., Falick A.M., Staackleton C.H.L. Roitman E. and Witkowska H.E. (1990) in *Biological Mass Spectrometry* (Burlingame A.L. and McCloskey J.A. Eds) Elsevier (Amsterdam)

Green B.N. (1988) *Mass spectrometry of small proteins* VG Analytical publication

Grillon C., Rieger K., Bakala J., Schott D., Morgat J. L., Hannapel E., Voelter W. and Lenfart M. (1990) *FEBS Lett.* 274 30

Gross M.L. (1989) *Mass Spectrom. Rev.* 8 165

Gualberto J.M., Lamattina L., Bonnard G., Weil J.H. and Grienenberger J.M. (1989) *Nature* 341 660

Hale G. and Perham R.N. (1980) *Biochem. J.* 187 905

Hale G. and Perham R.N. (1979) *Biochem. J.* 177 129

Hanks S.K., Quinn A.M. and Hunter T. (1988) *Science* 241 42

Hannapel E., Davonst S. and Horecker B.L. (1982) *Proc. Natl. Acad. Sci. USA* 79 1708

Hannapel E., Wartenberg F. and Bustelo X.R. (1989) *Arch. Biochem. Biophys.* 273 396

Hannapel E. and Van Kampen M. (1987) *J. Chromatogr.* 397 279

Harrison A.G. (1983) *Chemical Ionisation Mass Spectrometry* Boca Raton FL : CRC Press

Harrison A.G. and Tsang C.W. (1972) In *Biochemical Applications of Mass Spectrometry* ed. G.R.Wallner pp135-156 New York: Wiley-Interscience

Hayes R.N. and Gross M.L. (1990) *Meth. Enzymol.* 193 237

Hedin A., Hakansson P. and Sundqvist B.U.R. (1987) *Int. J. Mass Spectrom. Ion Proc.* 75 275

Heerma W., Kamerling J.P., Slotboom A.J., Scharrenberg G.J.M. Green B.N. and Lewis I.A.S. (1983) *Biol. Mass. Spectrom.* 10, 13

Heine C.E.C., Holland J.F. and Watson J.T. (1989) *Anal. Chem.* 61 2674

Henderson L.E., Krutzsch H.C. and Oroszlan S. (1983) *Proc. Natl. Acad. Sci. USA* 80 339

Henry K.D., Quinn J.P. and McLafferty F.W. (1991) *J. Am. Chem. Soc.* 113 5447

Hewick T.M., Hunkapiller M.W., Hood L.E. and Dreyer W.H. (1981) *J. Biol. Chem.* 256 7990

Hillenkamp F. and Karas M. (1990) *Meth. Enzymol.* 193 280

Hillenkamp F., Karas M., Beavis R.C. and Chait B.T. (1991) *Anal. Chem.* 63 1193A

Hodgson J.A., De Marcucci, O.G. and Lindsay, J.G. (1986) *Eur. J. Biochem.* 158 595

Hopper S., Johnson R.S., vath J.E. and Biemann K. (1989) *J. Biol. Chem.* 264 20438

Hortin G., Folz R. Gordon J. F. and Strauss A.W. (1986) *Biochem. Biophys. Res. Commun.* 141 326

Howell S.A., Welham K.J., Winkler F.J. and Baldwin M.A. (1987) *Proceedings of 16th meeting of the British Mass Spectrometry Society* 45

Howell S.A., Welham K.J. and Baldwin M.A. (1989) *Proceedings of 17th meeting of the British Mass Spectrometry Society* 121

Howell S.A., Ambler R., Aitken A., Welham K.J., and Baldwin M.A. (1990) *Proceedings of 18th meeting of the British Mass Spectrometry Society* 158

Huber, R., Romisch, J. and Paques, E.P. 1990 *EMBO J.* 9, 3867.

Hunkapiller M.W. Strickler J.E. and Wilson K.J. (1984) *Science* 226 304

Hunt D.F., Stafford G.C., Crow F.W. and Russel J.W. (1976) *Anal. Chem.* 48 2098

Hunt D.F., Shabanowitz J., Yates J.R., McIver R.T., Hunter R.L., Syka J.E.P and Amy J. (1985) *Anal. Chem.* 57 2728

Huttner, W.B. (1987) *Trends Biochem. Sci.* 12 361

Ingebritsen T.S. and Cohen P. (1983) *Eur. J. Biochem.* 132 255

Jennings K.R. and Dolnikowski G.G. (1990) *Meth. Enzymol.* 193 37

Jilka J.M. Rahmatullah, M., Kazema, M and Roche, T.E. (1986) *J. Biol. Chem.* 261 1858

Johnson R.S. and Biemann K. (1987) *Biochemistry* 26 1209

Johnson R.S., Matthews W.R., Biemann K. and Hopper S. (1988). *J. Biol. Chem.* 263 4182

Johnson R.S., Martin S.R. and Biemann K. (1988) *Int. J. Mass Spectrom. Ion Proc.* 86 137

Johnson R.S., Martin S.A., Biemann K., Stults J. T. and Watson J.T. (1987) *Anal. Chem.* 59 2621

Jonsonn G., Hedin A., Hakansson P., Sundqvist B.U.M., Bennich H. and Roepstorff P. (1989) *Rapid Commun. Mass Spectrom.* 3 190

Karas M. and Hillenkamp F. (1988) *Anal. Chem.* 60 2301

Karas M., Bahr U. and Hillenkamp F. (1989) *Int. J. Mass Spectrom. Ion Phys.* 92 231

Karp D.R. (1983) *J. Biol. Chem.* 258 12745

Katoh S. (1960) *Nature* 186 533

Kelly J. (1975) Phd thesis, University of Edinburgh

Kelley J.A., Nau H., Forster H.J. and Biemann K. (1975) *Biol. Mass Spectrom.* 2 313

Khan Z., Aitken A., Garcia J. and Smyth D.G. (1992) *J. Biol Chem.* 267 7464

Kiryushkin A.A., Fales H.M., Axenrod T., Gilbert E.J. and Milne G.W.A. (1971) *Org. Mass Spectrom.* 5 19

Klee, C.B., (1988) *Biochem J.* 27 6645.

Kligman, D. and Hilt D.C., (1988) *Trends Biochem. Sci.* 13 437

Knapp D. (1990) *Meth. Enzymol.* 193 314

Knudsen J., Clark, S. and Dils, R. (1976) *Biochem. J.* 160 683

Kornfeld R. and Kornfeld S. (1985) *Ann. Rev. Biochem.* 54 631

Lacey M.P. and Keough T. (1989) *Rapid Commun. Mass Spectrom.* 3 46

Laemmli U.K. (1970) *Nature* 227 680

Lawson G. and Todd J.F.J. (1972) *Chem. Brit.* 8 373

Lenfant M., Wdzieczak-Bakala J., Guittet E., Prome J.C., Sotty D. and Frindel E. (1989) *Proc. Natl. Acad. Sci. USA* 86 779

Libertini L.J. and Smith S. (1978) *J. Biol. Chem.* 253 1393

Lightbody J.J. and Krogmann D.W., (1967) *Biochim. Biophys. Acta* 131 508

Low T.L.K., Hu S.K. and Goldstein A.L. (1981) *Proc. Natl. Acad. Sci. USA* 78 1162

Mahadevan L.C., Aitken A., Heath J. and Foulkes, J.G. (1987) *EMBO J.* 6 921

Mains R.E., Eipper B.A., Glembotski C.C. and Dores R.M. (1983) *Trends Neurosci.* 6 229

Mandelbaum A., *Mass Spectrom. Rev.* 2 223 (1983)

Mallet A.I. (1990) *Proceedings of 18th meeting of the British mass spectrometry society* 75

Mann M. (1990) *Org. Mass Spectrom.* 25 575

Marshall A.G. and Grosshans P.B. (1991) *Anal. Chem.* 63 215A

Matsuidara, P. (1987) *J. Biol. Chem.* 262 10035.

Matsuo T., Matsuda H., Katakuse I., Shimonishi Y., Maruyama Y., Higuchi T. and Kubota E (1981) 53 416

Maxam A.M. and Gilbert W. (1980) *Meth. Enzymol.* 65 499

McCrery D.A., Ledford E.B. and Gross M.L. (1982) *Anal. Chem.* 54 1435

McEwen C.N. and Larsen B.S. (Eds) (1990) *Mass Spectrometry of Biological Materials.* Marcel Dekker (New York)

McFarlane R.D. and Torgerson D.F. (1976) *Science* 191 920

McFarlane R.D. (1990) *Meth. Enzymol.* 193 263

McLafferty F.W. (1980a) *Acc. Chem. Res.* 13 33

McLafferty F.W. (1980b) *Interpretation of Mass Spectra* California: Univ. Sci. Mill Valley

Meili J. and Seibl J. (1984) *Org. Mass Spectrom.* 19 581

Merrifield R.B. (1963) *J. Amer. Chem. Soc.* 85 2149

Merrifield R.B. (1964) *Biochemistry* 3 1385

Meyer H.E., Hoffmann-Posorske E., Korte H. and Heilmeyer Jr. L.M.G. (1986) *FEBS Lett.* 204 61

Millington D.S. and Smith J.A. (1977) *Org. Mass Spectrom.* 12 264

Morrice, N., Geary, P., Cammack, R., Harris, A., Beg, F. and Aitken, A. (1988) *FEBS Lett.* 231 336.

Morris H.R. (1980) *Nature* 286 447

Morris H. R., Panico M. and Taylor G.W. (1983) *Biochem. Biophys. Res. Commun.* 117 299.

Morris H.R. and Puuci P. (1985) *Biochem. Biophys. Res. Commun.* 126 1122

Morris H.R. (1972) *FEBS Lett.* 22 257

Moss J. and Bundgaard H. (1990) *J. Pharm. Pharmacol.* 42 7

Munson M.S.B. and Field F.H. (1966) *J. Am. Chem. Soc.* 88 2621

Munster H., Theobald F., Budzikiewicz H. and Schröder E. (1987) *Int. J. Mass Spectrom. Ion Phys.* 79 73

Naylor S., Findeis A.F., Gibson B.W. and Williams D.H. (1986) *J. Am. Chem. Soc.* 108 6359

Neumann G.M. and Derrick P.J. (1984) *Org. Mass Spectrom.* 19 165.

Nourse B.D. and Cooks R.G. (1990) *Anal. Chim. Acta.* 228 1

O'Connor T.P. Roche T.E. and Paukstelis J.V. (1982) *J. Biol. Chem.* 257 3110

O'Farrell P.H. (1975) *J. Biol. Chem.* 250 4007

Opsal R.B., Owens K.G. and Reilley J.P. (1985) *Anal. Chem.* 57 1884

Orlando R. (1992) *Anal. Chem.* 64 332

Orlowski M. and Meister A. (1971) in *Enzymes IV* (Boyer P.A., Ed) Academic Press (London)

Parker P.J.M, Stabek, S. and Waterfield, M.D. (1984) *EMBO J.* 3 953

Patel, J. and Kligman D. (1987) *J. Biol. Chem.* 262, 16686

Pepinsky, R.B., Tizard, R., Mattaliano, R.J., Sinclair, L.K., Miller, G.T., Browning, J.L., Chow, E.P., Burne, C., Huang, K-S, Pratt, D., Wachter, L., Hession, C., Frey, A.Z. and Wallner, B.P., (1988) *J. Biol. Chem.* 263 10799

Pesch R., Jung G., Rost K. and Tietje K.H. (1989) *Proc. 37th ASMS Conf. Mass Spectrom. Allied Topics* 1079

Phillips G.T., Nixon J.E., Dorsey J.A., Butterworth P.H.W., Chesterton C.J. and Porter J.W. (1970) *Arch. Biochem. Biophys.* 138 380

Posthumus M.A., Kistemaker P.G., Meuzelaar H.L. and Ten Noever de Brauw, M.C. (1978) *Anal. Chem.* 50 985

Poulter L. and Taylor L.C.E. (1989) *Int. J. Mass Spectrom. Ion Proc.* 91 183

Purdie N. and Swallows K.A. (1989) *Anal. Chem.* 61 77A

Rademacher T.W., Parekh R.B. and Dwek R.A. (1988) *Ann. Rev. Biochem.* 57 785

Ramage R., Green J. and Oqunjobi O.Y. (1990) in *Innovation and Perspectives in Solid Phase Synthesis* (Epton R. Ed.) SPCC (Birmingham)

Randhawa Z.I. and Smith S., (1987), *Biochemistry* 26 1365

Reddy S.J., Röllgen F.W., Maas A. and Beckey H.D. (1977) *Int. J. Mass Spectrom. Ion Phys.* 25 147

Reed L.J. (1974) *Acc. Chem. Res.* 7 40

Richardson W.S., Munksgaard E.C. and Butler W.T. (1978) *J. Biol. Chem.* 253 8042

Roepstorff P. and Fohlmann J. (1984) *Biol. Mass Spectrom.* 11 601

Roepstorff P. (1990) *Meth. Enzymol.* 193 432

Rose M.E. and Johnstone R.A.W. (1982) *Mass Spectrometry for Chemists and Biochemists* Cambridge University Press

Rui H., Welinder B.S., Purvis K. and Dolva O. (1987) *J. Endocrinol.* 114 329

Russell D.H. (1990) *Mass Spectrom. Rev.* 9 405

Safford R., de Silva J., Lucas C., Windust J.H.C., Shedden J., James C.M., Sidebottom C.M., Slabas A.R., Tombs M.P. and Hughes S.G. (1987) *Biochemistry* 26 1358

Schlaepfer D.D., Mehlman T., Burgess W.H. and Haigler H.T., (1987) *Proc Natl Acad Sci USA* 84, 6078.

Schobitz B., Netzker R., Hannapel E. and Brand K. (1990) *J. Biol. Chem.* 265 15387

Schwartz B.L. and Bursey M.M. (1992) *Biol. Mass Spectrom.* 21 92

Shiea J. and Sunner J. (1991) *Org. Mass Spectrom.* 26 38

Simpson R.J., Begg G.S., Reid G.E., Rubira M.R., Ward L.D. and Moritz R.L. (1988) in *Methods in protein sequence analysis* Wittman-Leibold B (ed) Springer-Verlag

Slabas A.R., Aitken A., Howell S.A., Welham K. and Sidebottom C.M. (1989) *Biochem. Soc. Trans.* 17 886

Slabas A.R., Ormesher, J. Roberts, P. A. Sidebottom, C. M. Tombs, M. P. Jeffcoat R. and James A.T. (1983) *Eur. J. Biochem.* 134 27

Smith D.L. and Zhan Z. (1990) *Meth. Enzymol.* 193 374

Smith L.M., Sanders J., Kaiser T., Hughes P., Dodd C. and Hood L.E. (1986) *Nature* 321 674.

Smith R.D., Loo J.A., Edmonds C.G., Barinaga C.J. and Udseth H.R. (1990) *Anal. Chem.* 62 882

Spencer M.E. Darlison M.G., Stephens P.E. Duckenfield, I.K. and Guest, J.R. (1984) *Eur. J. Biochem.* 141 361

Stanley C.J. and Perham R.N. (1980) *Biochem. J.* 191 147

Stephens P.E. Darlison, M.G., Lewis, H.M. and Guest, J.R. (1983) *Eur. J. Biochem.* 133 481

Stoll R. and Röllgen F.W. (1979) *Org. Mass Spectrom.* 14 642

Stryer L. (1981) *Biochemistry* (2nd Edition) Freeman (San Francisco)

Stults J.T. (1990) *Methods of Biochemical Analysis* 34 145

Stumpo D.J., Graff J.M., Albert K.A., Greengard P. and Blackshear P.J. (1989) *Proc. Natl. Acad. Sci. USA* 86 4012

Suming H., Yaozu C., Longfrei J. and Shuman X., *Org. Mass Spectrom.* 21 7 (1986)

Sundqvist B., Roepstorff P., Fohlman J., Hedin A., Hakansson P., Kamensky I., Lindberg M., Salepour M. and Save G. (1984) *Science* 226 696

Sunner J.A., Kulatunga R. and Kebarle P. (1986) *Anal. Chem.* 58 1312

Sunner J.A., Morales A. and Kebarle P. (1987) *Anal. Chem.* 59 1378

Sunner J.A., Morales A. and Kebarle P. (1988) *Anal. Chem.* 60 98

Sykes A.G. (1985) *Chem. Soc. Rev.* 14 283

Tabet J.C., Kagan H.B., Poulin J.C., Meyer D. and Fraise D., *Spectros. Int. J.* 4 81 (1985)

Thannhauser T.W., Konishi Y. and Scheraga H.A. (1985) *Meth. Enzymol.* 143 115

Thorne G.C. and Gaskell S.J. (1989) *Rapid Commun. Mass. Spectrom.* 3 217

Toker A., Ellis C.A., Sellers L.A. and Aitken A. (1990) *Eur. J. Biochem.* 191 421

Tomer K.B. (1989) *Mass. Spectrom. Rev.* 8 445

Tomer K.B., Crow F.W., Gross M.L. and Kopple K.D. (1984) *Anal. Chem.* 56 880

Uggerud E. and Derrick P.J. (1991) *J. Phys. Chem.* 95 1430

Van Dorrselaer A., Bitsch F., Green B.N., Jarvis S., Lepage P., Bischoff R., Kolbe H.V.J. and Roitsch C. (1990) *Biol. Mass Spectrom.* 19 692

Van Berkel G.J., McLuckey S.A. and Glish G.L. (1991) *Anal. Chem.* 63 1098

Wagner D.S., Salari A., Gage D.A., Leykam J., Felter J., Hollingsworth R., Watson J.T. (1991) *Biol. Mass Spectrom.* 20 419

Wawrzynczak, E.J. Perham, R.N. and Roberts G.C.K. (1981) *FEBS Lett.* 131 151

Weber, K., Johnsson, N., Plessmann, U., Van, P.N., Soling, H-D, Ampe, C. and Vanderkerckhove, J., (1987) *EMBO J.* 6 1599.

Wellner D., Panneerselvan C. and Horecker B.L. (1990) *Proc. Natl. Acad. Sci. USA* 87 1947

White R.H. Bleile, D.M. and Reed L.J. (1980) *Biochem. Biophys. Res. Commun.* 94 78

Wilkins C.L., Weil D.A., Yang C.L.C. and Ijames C.F. (1985) *Anal. Chem.* 57 520

Williams D.H., Findeis A.F., Naylor S. and Gibson B.W. (1987) *J. Am. Chem. Soc.* 109 1980

Winkler F.J., Stahl D. and Maquin F., *Tetrahedron Lett.* 27 355 (1986)

Witkowska H.E., Green B.N. and Smith S. (1990). *J. Biol. Chem.* 265 5662

Wold F. (1981) *Ann. Rev. Biochem.* 50 783

Wold F. (1984) *Trends Biochem. Sci.* 9 256

Wood P.M. and Bendall D.S., (1975). *Biochim. Biophys. Acta* 387 115

Yan C.B., Grinnell B.W. and Wold F. (1989) *Trends Biochem. Sci.* 14 264

Yang Y.S. and Frey P.A. (1986) *Biochemistry* 25 8173

Yang H-J. and Chen Y-Z. (1992) *Org. Mass Spectrom.* 27 736

Yeaman S.J. (1986) *Trends Biochem. Sci.* 11 293

Yost R.A. and Boyd R.K. (1990) *Meth. Enzymol.* 193