

eros ✓
17/4/14

FOR REFERENCE ONLY

ProQuest Number: 10290219

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10290219

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

**Proteomic Strategies for Protein and Biomarker
Identification by Matrix-Assisted Laser
Desorption/Ionization Mass Spectrometry
(MALDI-MS)**

Lucy Vivien Ratcliffe

**A thesis submitted in partial fulfilment of the
requirements of Nottingham Trent University
for the degree of Doctor of Philosophy**

April 2006

19 JUL 2007

41 0661968 8



371690

PKB/06 RAT

Abstract

This thesis describes the development of novel strategies for the analysis of peptides by MALDI mass spectrometry. The developed techniques are applied to the identification of protein and proteomic biomarkers for melanoma. A commercial atmospheric pressure (AP-MALDI) source (MassTechnologies, Burtonsville, MD, USA) was modified to allow operation with a high powered nitrogen laser and independent PC control of the sample stage. A software interface was developed using LabVIEW 6.1 that allows full control of the target position with respect to the laser fibre optic interface, allowing the target to be adjusted within any point within a particular sample spot to enhance signal quality. The modified AP-MALDI-QIT interface was evaluated for the analysis of standard peptide mixtures and tryptic digests of proteins.

AP-MALDI-QIT analysis of tryptic peptides following capillary liquid chromatographic (LC) separation and direct analysis of a protein digest is reported. Peptide fragments were identified by peptide mass fingerprinting from mass spectrometric data and sequence analysis obtained by tandem mass spectrometry of the principal mass spectral peaks using a data-dependent scanning protocol. These data were compared with those from mass spectrometric analysis using capillary LC/MALDI-time-of-flight (TOF) and capillary LC/electrospray ionisation (ESI)-quadrupole TOF. For all three configurations the resulting data were searched against the MSDB database, using MASCOT and the sequence coverage compared for each technique. Complementary data were obtained using the three techniques.

A bottom-up proteomic methodology for the peptide profiling of human serum samples using MALDI mass spectrometry was developed. Reproducibility studies were carried out to define the MALDI measurement precision. Pre-analytical sample handling factors, such as room temperature incubation and freeze thaw cycles have also been investigated. The methodology developed was applied to the analysis of serum peptides from stage IV melanoma patients and healthy control subjects. Prediction of human melanoma metastatic cancer from peptide profiling using artificial neural networks (ANNs) model classified 98 % of samples correctly. The identification of three out of six ions predicted by the ANNs model to be indicative biomarkers that have good predictive performance were identified using MALDI PSD, AP-MALDI MS/MS and LC-ESI-MS/MS. Two of the ions were shown to belong to the same identified peptide, α -1-acid glycoprotein precursor (1, 2) which correctly predicted 95 % (i.e. 45/50) of metastatic melanoma patients.

Acknowledgments

I would like to thank my supervisors, Professor Colin Creaser, Professor Robert Rees and Dr Balwir Matharoo-Ball for their advice during the course of this study.

Also I would like to thank Philip Green, Peter Kilby and Susan Crosland at Syngenta, for their help and guidance during the collaboration on the LC-MALDI work.

I gratefully acknowledge the Analytical Chemistry Trust Fund and the Engineering and Physical Research Council (EPSRC) for an analytical science studentship.

I am eternally grateful to Pete Ryan at KR Analytical, for the loan of equipment often at short notice. A special thanks to Rob, for his assistance with the tedious task of ZipTip clean-up for the melanoma study. I wish to thank Lee for carrying out the Artificial Neural Networks analysis.

I would like to thank the good friends I have made over the years at Nottingham Trent, especially Jo and also the entire inorganic research group for making me a pseudo lab member.

To Paul, you have always been there for me, thank you for your never-ending support and help through the trials and triumphs of my research.

Finally and most importantly, I would like to dedicate this thesis to my parents and my brother, for all their love, support and encouragement especially throughout the past three years. Without them none of this would have been possible.

Contents

Chapter 1: Introduction	1
1. Mass Spectrometry	2
1.1 Mass Analysers	3
1.1.1 Time-of-Flight	3
1.1.2 Hybrid time-of-Flight	6
1.1.3 Linear Quadrupole	7
1.1.4 Quadrupole Ion Trap	10
1.1.5 Tandem Mass Spectrometry	14
1.1.5.1 Scan Modes for Tandem Mass Spectrometry	15
1.1.5.2 Collisionally Induced Dissociation (CID)	16
1.2 Ionization	18
1.2.1 Fast Atom Bombardment	18
1.2.2 Electrospray Ionization	18
1.2.3 Matrix-Assisted Laser Desorption/Ionization (MALDI)	19
1.2.3.1 Mechanisms of MALDI	20
1.2.3.2 MALDI Matrices	23
1.2.4 Atmospheric Pressure MALDI	24
1.2.4.1 AP-MALDI Combined with Time-of-Flight Mass Spectrometry	25
1.2.4.2 AP-MALDI Combined with Ion Trap and Fourier Transform Mass Spectrometry	26
1.2.4.3 Applications of AP-MALDI	31
1.2.4.4 Comparison of AP-MALDI and Vacuum MALDI	32
1.2.4.5 IR AP-MALDI	35
1.3 Analysis of Peptides by Mass Spectrometry	37
1.4 Proteomics and Strategies for Proteome Analysis	40
1.4.1 Proteomic Strategies Using Mass Spectrometry	41
1.4.2 Protein Separation	43
1.4.3 Protein Identification Using Database Searching	46
1.4.4 Biomarker Identification – Comparative Proteomics	47
1.4.4.1 Biomarker Identification by Mass Spectrometry	48
1.5 Research Aims and Outcomes	53
1.6 References	54
Chapter 2: Development and Modification of a Commercial Ionization Source for AP-MALDI	67
2.1 Introduction	68
2.2 Hardware Configuration	69
2.3 LabVIEW	71

2.4 Software Development	72
2.4.1 AP-MALDI Plate Coordinates	72
2.4.2 Development of Basic Motion	73
2.4.3 Motion Control Development	75
2.4.4 Joystick Control	79
2.5 Evaluation of AP-MALDI Ion Source and Control Software	80
2.5.1 Materials	80
2.5.2 AP-MALDI-QIT	80
2.5.3 Effects of Sample Preparation Technique and Application to Target	80
2.5.4 Analysis of gp70 and Neurotensin	81
2.5.5 Proteolytic Digestion of Ubiquitin	81
2.6 Results and Discussion	82
2.6.1 Sample Preparation Techniques	82
2.6.2 AP-MALDI Analysis of gp70 and Neurotensin	85
2.6.3 AP-MALDI-QIT Analysis of Tryptic Peptides	86
2.7 Conclusion	88
2.8 References	89
Chapter 3: Capillary LC interfaced with AP-MALDI/QIT for the Identification of Tryptic Peptide	90
3.1 Introduction	91
3.2 Experimental	93
3.2.1 Materials	93
3.2.2 Sample Preparation and Capillary LC Separation	94
3.2.3 AP-MALDI-QIT	95
3.2.4 MALDI-TOF	95
3.2.5 ESI-qTOF	95
3.2.6 Database Searching	97
3.3 Results and Discussion	98
3.3.1 LC/MALDI Peptide Mass Fingerprinting	98
3.3.2 LC/AP-MALDI and LC/ESI Tandem Mass Spectrometry	101
3.3.3 Direct MALDI Peptide Mass Fingerprinting	103
3.3.4 AP-MALDI-QIT Tandem Mass Spectrometry Without LC Pre-Separation	109
3.4 Conclusions	113
3.5 References	114

Chapter 4: Development and Evaluation of Methodology for the Analysis of Serum Peptides Using MALDI-TOF and AP-MALDI-QIT Mass Spectrometry	116
4.1 Introduction	117
4.2 Experimental	122
4.2.1 Materials	122
4.2.2 BCA Assay	122
4.2.3 Method Development for the Analysis of Native Serum Peptides	123
4.2.3.1 Dilutions	123
4.2.3.2 ZipTip Clean-up	123
4.2.3.3 Ultrafiltration	123
4.2.3.4 Acetonitrile Precipitation	124
4.2.4 Method Development for the Analysis of Tryptic Serum Peptides	125
4.2.4.1 Proteolytic Digestion	125
4.2.4.2 Acetonitrile Precipitation Prior to Digestion	125
4.2.4.3 ZipTip Clean-up Prior to Digestion	126
4.2.5 Optimised Methodology for the Analysis of Native Peptides	126
4.2.6 Optimised Methodology for the Analysis of Tryptic Serum Peptides	127
4.2.7 Serum Aging Study	128
4.2.8 Mass Spectrometry	129
4.2.8.1 MALDI-TOF	129
4.2.8.2 AP-MALDI	130
4.2.8.3 Target Plates	130
4.2.8.4 Target Plate Cleaning Method	131
4.3 Results and Discussion	132
4.3.1 BCA Assay	132
4.3.2 Analysis of Native Serum Peptides	133
4.3.2.1 Dilutions and ZipTip Clean-up	133
4.3.2.2 Ultrafiltration	134
4.3.2.3 Acetonitrile Precipitation	137
4.3.3 Analysis of Tryptic Serum Peptides	141
4.3.3.1 Proteolytic Digestion	141
4.3.3.2 Acetonitrile Precipitation Prior to Digestion	142
4.3.3.3 ZipTip Clean-up Prior to Digestion	146
4.3.4 Optimisation of Mass Spectrometry for the Analysis of Serum Peptides	149
4.3.5 Mass Spectral Reproducibility	152
4.3.6 Serum Aging Study	155
4.3.6.1 Native Serum Peptides	155
4.3.6.2 Sensitivity Analysis of Native Serum Peptide Data	159
4.3.6.3 Tryptic Serum Peptides	163
4.3.6.4 Sensitivity Analysis of Tryptic Serum Peptide Data	166
4.4 Conclusions	169
4.5 References	170

Chapter 5: Identification of Diagnostic Biomarkers Associated with Stage IV Melanoma Using MALDI-TOF Mass Spectrometry and Bioinformatics	174
5.1 Melanoma	175
5.1.1 Stages of Melanoma	176
5.1.2 Development and Progression of Melanoma	177
5.1.3 Melanoma Biomarkers	178
5.2 Methods and Materials	181
5.2.1 Serum Samples	181
5.2.2 Sample Preparation	181
5.2.3 Mass Spectrometric Analysis	182
5.2.3.1 MALDI-TOF MS	182
5.2.3.2 AP-MALDI MS/MS	182
5.2.3.3 LC-ESI MS/MS	182
5.2.4 Database Searching	184
5.2.5 ANNs Analysis	185
5.3 Results	
5.3.1 Identification of Melanoma Stage IV Biomarkers: Native Serum Peptides	187
5.3.2 Identification of Melanoma Stage IV Biomarkers: Tryptic Serum Peptides	193
5.4 Discussion	201
5.5 Conclusion	207
5.6 References	208
Chapter 6: Thesis Conclusions and Further Work	213
6.1 References	218
Appendix: Publications and Presentations	219

This thesis is to the best of my knowledge, original except where due reference is made.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Lucy Vivien Ratcliffe

April 2006

Chapter 1

Introduction

1. Mass Spectrometry

Mass spectrometry was first reported in 1913 by J.J. Thompson when it was used to demonstrate the existence of stable isotopes in elements such as neon.¹ Since then it has been developed into a major technique for the analysis of a wide variety of chemical and biological materials. A mass spectrometer is an instrument which ionises a sample, separates the ions according to their mass to charge (m/z) ratio and records the relative abundance of the ions to produce a mass spectrum. The main components of a mass spectrometer are shown in Figure 1.1 and outlined below.

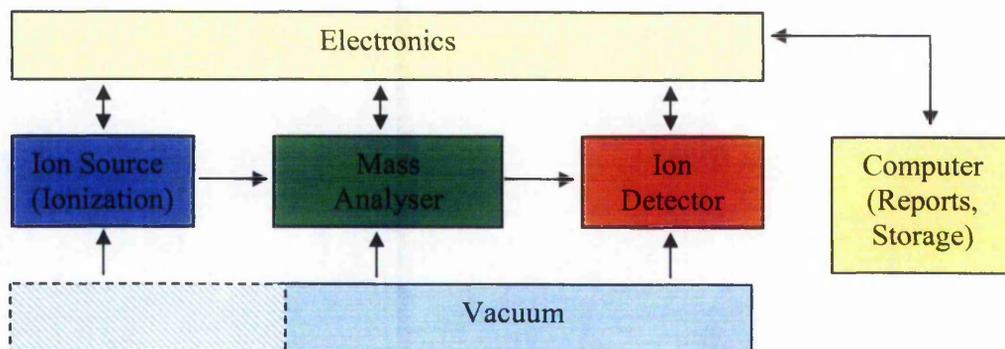


Figure 1.1: Schematic of a Mass Spectrometer

1.1 Mass Analysers

Once ions have been formed in the source region they are focussed into the mass analyser which separates the ions based on their m/z ratio. A quadrupole ion trap and time-of-flight mass spectrometer were used for the experimentation discussed in this thesis and these are detailed below.

1.1.1 Time-of-Flight

Time-of-flight mass spectrometry (TOF-MS) was first reported by Wiley and McLaren in 1955.² The mode of operation is relatively straightforward with ions separated on the basis of their velocity difference following ion acceleration from the ion-source region into a field free drift region (flight tube). Ions are accelerated to have the same nominal kinetic energy (equation 1) and each mass travels at a different velocity. The higher the mass the lower the velocity and hence the high mass ions arrive at the detector plate later than low mass ions.

$$KE = zeV_{acc} = \frac{mv^2}{2} \quad \text{Eq. 1}$$

Where z = charge upon the ion, e = fundamental unit of charge (1.602×10^{-19} C) and V_{acc} = accelerating voltage, v = velocity and m = mass of the ion

The ions travel towards the detector with a velocity, v given by equation 2:

$$v = \left(\frac{2zeV_{acc}}{m} \right)^{1/2} \quad \text{Eq. 2}$$

The time it takes for an ion to travel down the length of the flight tube (d) is known as the flight time (t):

$$t = \frac{d}{v} \quad \text{Eq.3}$$

Hence, velocity may be expressed as:

$$v = \frac{d}{t} \quad \text{Eq. 4}$$

The m/z ratio may be calculated by substituting v into equation 2 and rearranging the equation:

$$m/z = 2eV_{acc} \frac{t^2}{d^2} \quad \text{Eq. 5}$$

The main advantages of TOF analysers are that they have theoretically unlimited mass range, high transmission efficiencies and are compatible with pulsed ionization techniques. For example, molecules with molecular masses above 300 kDa have been detected using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF)^{3, 4} (Section 1.2) and detection limits of 100-200 amol amounts of proteins have been reported.⁵

The main limitation of linear TOF analysers is poor resolution, resulting from variations in the velocities of ions of the same m/z . Two innovations have been developed to overcome this problem, the inclusion of an electrostatic reflector (reflectron) and delayed pulsed extraction. The reflectron focuses ions of the same m/z but with different kinetic energies, at the detector thus dramatically improving resolution, Figure 1.2. This is achieved because ions of higher kinetic energy arrive at the reflectron before ions of the same m/z with lower kinetic energy. However, they penetrate further into the electrostatic field and spend longer in the reflectron than lower kinetic energy ions do.

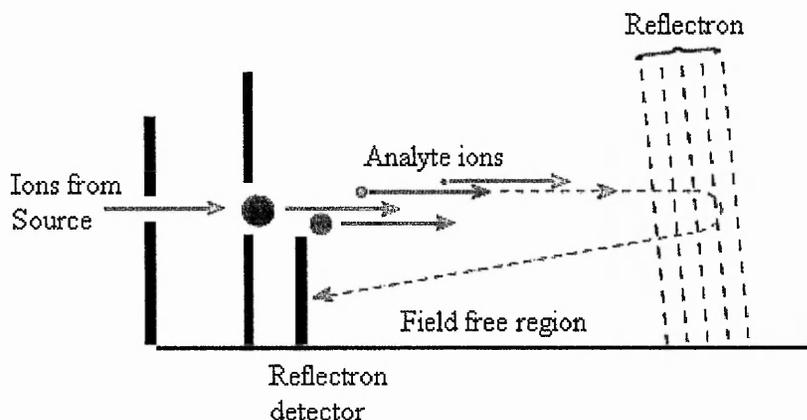


Figure 1.2: Schematic of a reflectron TOF instrument ⁶

MALDI-TOF allows two different types of fragmentation reactions to be distinguished, post-source decay (PSD) and prompt fragmentation also referred to as in-source decay (ISD). Fragmentation either occurs in the source region within a few hundred ns following the laser pulse for ISD or on a μs time scale in the field free region of the mass spectrometer for PSD.⁷

⁸ Although incorporating a reflectron lowers sensitivity and limits the mass range, it enables the process of PSD to be observed. Metastable fragmentation may occur in the flight tube prior to the ions entering the reflectron. Ions lose kinetic energy with the loss of mass because the energy is partitioned between fragments and therefore the time-of-flight will differ for the precursor and fragments. As previously described the reflectron is capable of focusing these low kinetic energy ions allowing them to be observed. Time-of-flight mass spectrometers may be fitted with a deflection electrode allowing selection of a precursor ion in order to observe its fragmentation by PSD.

A technique known as delayed pulsed extraction, originally known as time-lag focusing, is another method used to improve resolution of TOF analysers.⁹ The acceleration voltage is pulsed rather than continually applied, allowing ions to remain in the source for a period of time prior to extraction into the flight tube. The voltage transferred to the ions compensates for any differences in initial kinetic energies ensuring ions of the same m/z reach the detector at

the same time. Delayed extraction also enables fragment ions to be detected by ISD using a linear TOF instrument. The incorporation of a reflectron / delayed extraction complicates the calculation of m/z , therefore the following equation is used to calculate the m/z from its flight time:

$$m/z = at^2 + b \quad \text{Eq. 6}$$

where a and b represent constants obtained from calibration compounds of known molecular weight.

1.1.2 Hybrid Time-of-Flight

The most common hybrid TOF instrument is the quadrupole TOF (qTOF), which consists of a mass filtering quadrupole and a collision cell with a reflectron TOF, Figure 1.3.

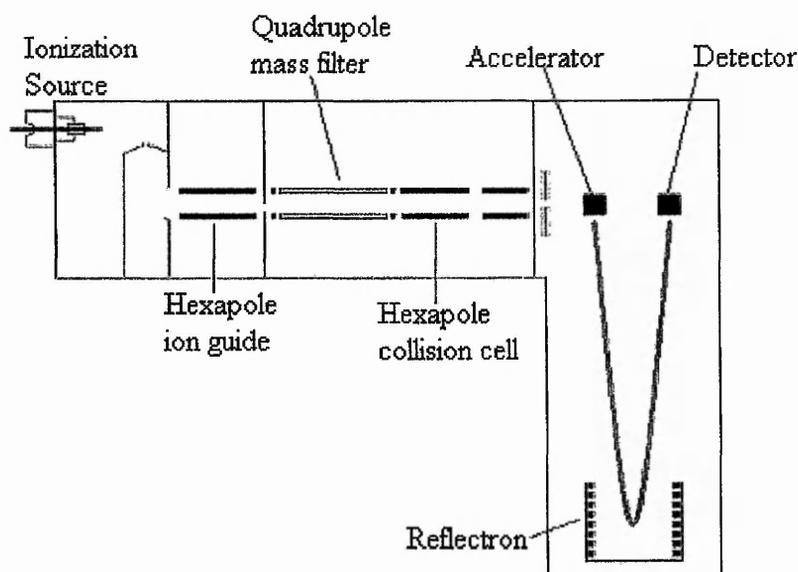


Figure 1.3: Diagram of a qTOF instrument

The quadrupole can be operated either by permitting m/z values of all ions to be recorded (RF only mode) resulting in a full scan or by acting as a mass filter allowing selection of precursor ions. The selected ions undergo collisionally induced dissociation (CID), described later in this

Chapter (Section 1.1.5.2), in the hexapole collision cell, the fragment ions then enter the TOF to be analysed. qTOF instruments combine the high resolution, sensitivity and mass accuracy of the TOF analyser with the benefits of tandem mass spectrometry resulting in a powerful proteomics tool. HPLC is often interfaced with qTOF instruments.^{10, 11} MALDI has been interfaced with a qTOF for the analysis of a wide variety of compounds.¹²⁻¹⁴ Another recently introduced tandem TOF instrument is the TOF/TOF, this configuration consists of two TOF analysers separated by a collision cell. The first TOF selectively isolates the precursor ion and the second TOF resolves the product ions before detection. Multi-stage tandem mass spectrometry (MS^n) is possible using an ion trap TOF hybrid.^{15, 16}

1.1.3 Linear Quadrupole

The linear quadrupole mass filter was invented by Paul and Steinwedel at the University of Bonn in 1956.¹⁷ A wide range of compounds have been analysed by this device^{18, 19} which consists of four parallel rods, Figure 1.4.

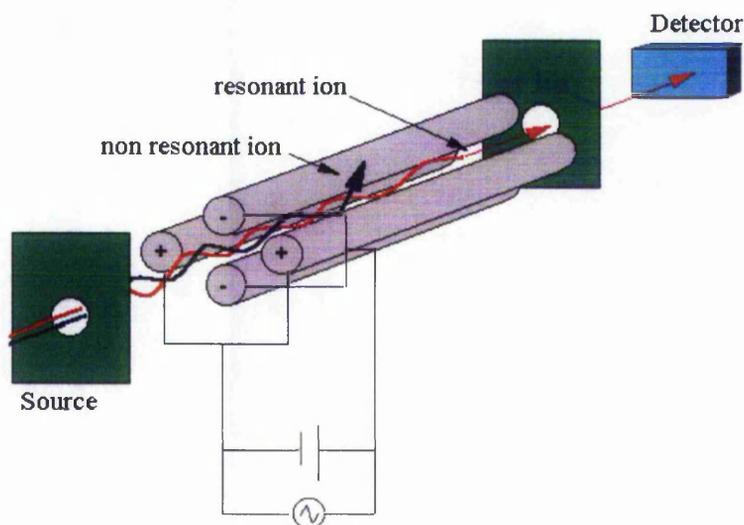


Figure 1.4: Schematic of a linear quadrupole mass filter²⁰

DC and superimposed RF voltages are applied to these hyperbolic or cylindrical rods, the voltages applied to one pair of rods has the opposite polarity to that applied to the other set and the RF voltages are out of phase by 180 °. Ions are accelerated from the ion source into the quadrupole analyser by a small accelerating voltage, usually around 5 V. Inside the quadrupole the electric fields help to either stabilise or destabilise the travelling ion. Only ions with stable trajectories will pass along the quadrupole to the detector, unstable ions will collide with rods or pass between the rods, and not reach the detector. Whether an ion is stable can be determined by Equations 7 and 8, which are derived from Newton's second law (Force = mass x acceleration). These equations are similar to the Mathieu equation, which was originally used to describe the vibrations of a drum skin.²¹

$$a_x = -a_y = \frac{8eU}{(mr_0^2\Omega^2)} \quad \text{Eq. 7}$$

$$q_x = -q_y = \frac{4eV}{(mr_0^2\Omega^2)} \quad \text{Eq. 8}$$

Where e is charge, m is mass, r_0 is the quadrupole radius, U and V are the applied DC voltage and RF voltage respectively and Ω is the angular frequency of the RF ($\Omega = 2\pi f$ where f is the frequency in Hz). An ion is stable and will pass along the analyser if its values of a and q lie within specific boundaries, these may be represented on a stability diagram. A segment of a stability diagram for a linear quadrupole is shown in Figure 1.5. An ion must be stable in both x and y trajectories in order to travel through the analyser to the detector, the working point of these ions lie within the boundaries of the stability diagram. Ions whose working points are outside the stability regions are unstable in one or both axis and will not be detected.

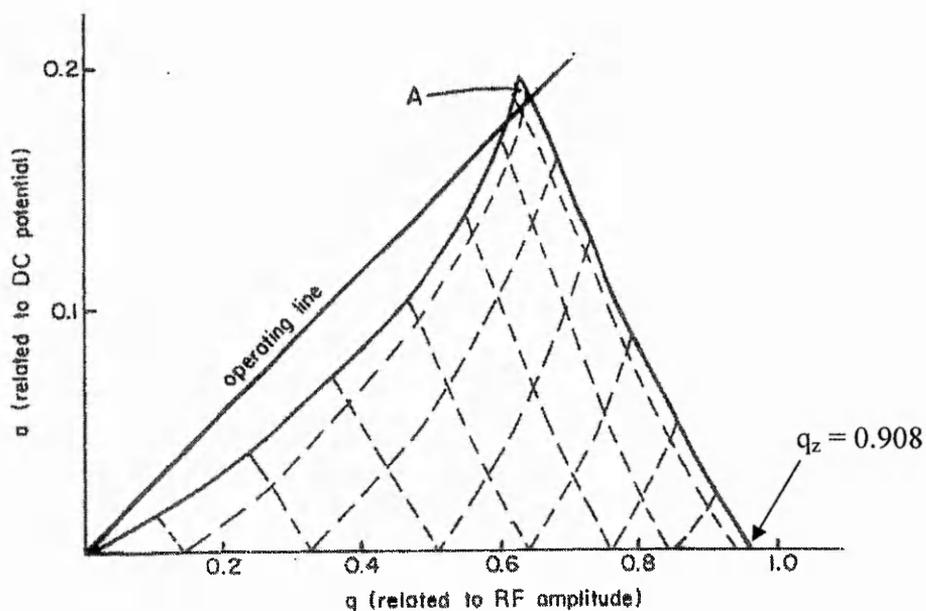


Figure 1.5 – Stability diagram for a linear quadrupole²²

Detection of ions of increasing m/z is possible by scanning the quadrupole voltages at a constant DC/RF voltage ratio in a mass selective stability mode. As the voltage increases the working point of the ions moves along the operating line shown in Figure 1.5. One mass at a time will have a and q values within the top apex boundary, have stable trajectories and therefore pass through the quadrupole to the detector. Ions with adjacent m/z ratios will have unstable trajectories and will not be detected. If the device is operated without an applied DC voltage (RF-only), then the a term will be zero and all ions which have $0 < q < 0.908$ will have a stable trajectory.

The development of instruments with several quadrupoles in sequence, for example the triple quadrupole, enables tandem mass spectrometric analysis to be carried out. In a triple quadrupole the first and third quadrupoles act as the mass filters and the central quadrupole operates as an rf-only collision cell. The collision cell is filled with an inert gas, such as

helium or argon, to induce dissociation (Section 1.1.5.2). The third quadrupole may be replaced with a TOF mass analyser, as described in the hybrid TOF section.

1.1.4 Quadrupole Ion Trap

The quadrupole ion trap (QIT) is in essence a three dimensional analogue of the quadrupole mass filter.²³ A quadrupole ion trap consists of three hyperbolic electrodes, a ring electrode and two end-caps, which generate a three-dimensional quadrupole field, Figure 1.6. The ring electrode is positioned symmetrically between the end cap electrodes.

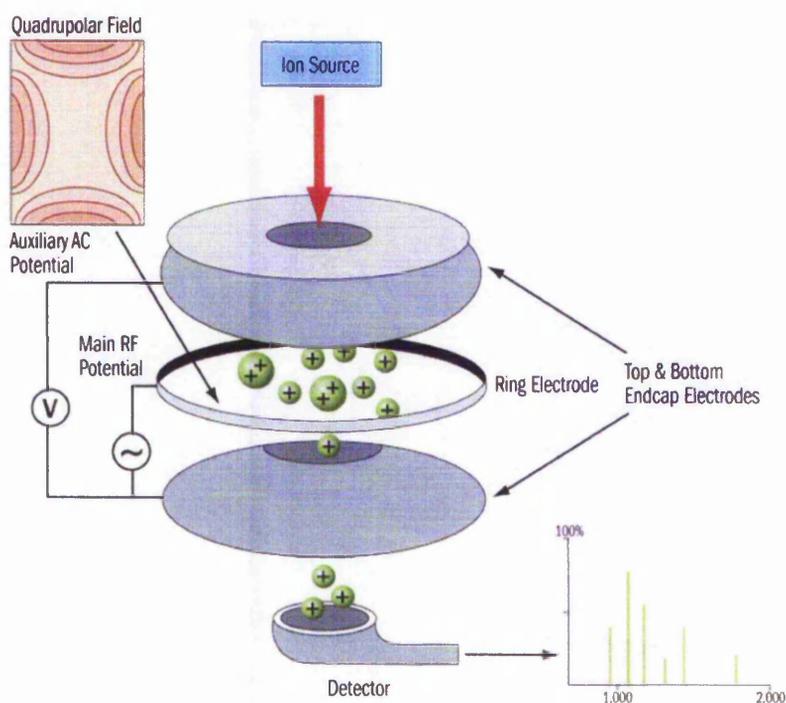


Figure 1.6: Diagram of a quadrupole ion trap, (source: Michael Willet, Bruker Daltonics)

The geometries of the electrodes are defined so that the device produces a parabolic potential well for the entrapment of ions. A potential well is created through the application of an oscillating frequency (RF) and/or DC voltages applied to the ring electrode and in the simplest arrangement the end-caps are grounded. The force experienced by the ions within this field is proportional to the distance of the ion from the centre of the trap, with ions directly in the

centre having no applied force. Ions are focused to the centre of the trap through the RF and the incorporation of a buffer gas, usually helium at a pressure of 1 mtorr. Collisions with the buffer gas reduce the kinetic energy of the ions allowing them to be held more efficiently in the centre.²⁴

The motion of ions within a pure quadrupole field can be described mathematically by the Mathieu equation with regions of stability described by the parameters a_z and q_z :

$$a_z = \frac{-8eU}{(mr_0^2\Omega^2)} \quad \text{Eq. 9}$$

$$q_z = \frac{4eV}{(mr_0^2\Omega^2)} \quad \text{Eq.10}$$

Where e is the charge of an electron, U is the applied DC voltage, V is the applied RF potential, Ω is the angular frequency of the fundamental RF, r_0 is the internal radius of the ring electrode and m is the mass. A plot of these parameters gives rise to an ion trap stability diagram, Figure 1.7. Ions within the stable region have secular frequencies, Ω_0 , where:

$$\Omega_0 = \beta\Omega/2 \quad \text{Eq. 11}$$

The β values of a trapped ion ranges from 0 to 1 and describes the axial (β_z) and radial (β_r) components of the secular frequency of motion of the ion in relation to the applied RF field. Ions along the same iso- β lines, marked on the stability diagram, possess the same secular frequency.

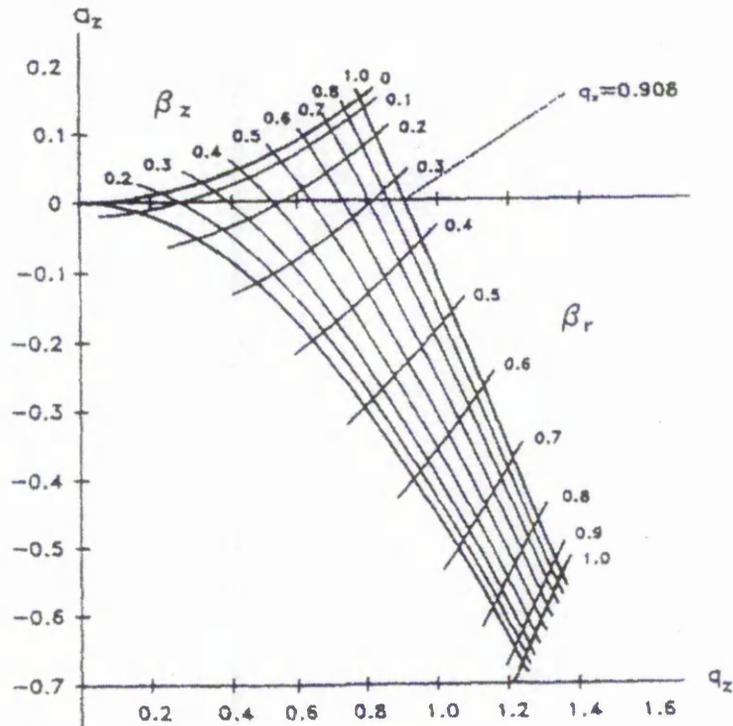


Figure 1.7: Stability diagram for a quadrupole ion trap²⁵

The geometry of an ideal quadrupole field can be described by the relationship $r_0^2 = 2z_0^2$, where r_0 is the internal radius of the ring electrode and $2z_0$ is the separation between the endcap electrodes. In many commercial ion traps the distance between the end caps has been increased to enhance performance. A pure quadrupole field does not exist and so $r_0^2 \neq 2z_0^2$ and the trapping parameters are calculated slightly differently to reflect this:

$$a_z = \frac{-16eU}{m(r_0^2 + 2z_0^2)\Omega^2} \quad \text{Eq.12}$$

$$q_z = \frac{8eV}{m(r_0^2 + 2z_0^2)\Omega^2} \quad \text{Eq.13}$$

The stability of an ion is dependent upon mass, charge and the dimensions and operating conditions of the trap. Ions are stable and can be stored within the trap if both their radial and axial (r and z) trajectories are stable. Ions must have unstable trajectories to be ejected from

the trap to the detector. The working point of an ion is defined by the a_z and q_z values and can be altered by varying the amplitude of the RF and DC voltages. Values of a_z and q_z are selected to ensure stability or instability of an ion trajectory of the ion of interest.

There are two modes of operation for an ion trap mass analyser. The first is the mass selective stability mode, where the amplitude of RF and DC voltages is increased at a constant RF/DC ratio to allow stability and therefore selective storage of ions of increasing m/z values in the trap.²⁶ Potentials may be chosen so that ions of a selected m/z will be stable and therefore stored in the trap whilst all other ions are unstable and ejected from the trap. The trapped ions can then be subjected to tandem mass spectrometry as discussed in Section 1.1.5.

The second mode of operation is the mass selective instability mode, which is the most common mode of operation, where no DC voltage is applied (i.e. $a_z = 0$) and the RF voltage is increased at a controlled rate.²⁷ Under these conditions, the working point of the ion is determined only by the q_z value, Figure 1.8. Stable trajectories are obtained for ions with working points in the range $0 < q_z < 0.908$, but ion trajectories become unstable in the axial direction (between the endcaps) at q_z values above 0.908 and ions are ejected through the endcap electrode to the detector. Therefore as the RF voltage applied to the ring electrode is ramped, the q_z value of the ions increases until the working point reaches >0.908 (where $\beta_z = 1$) at which point the ions are sequentially ejected to the detector allowing a mass spectrum to be obtained.

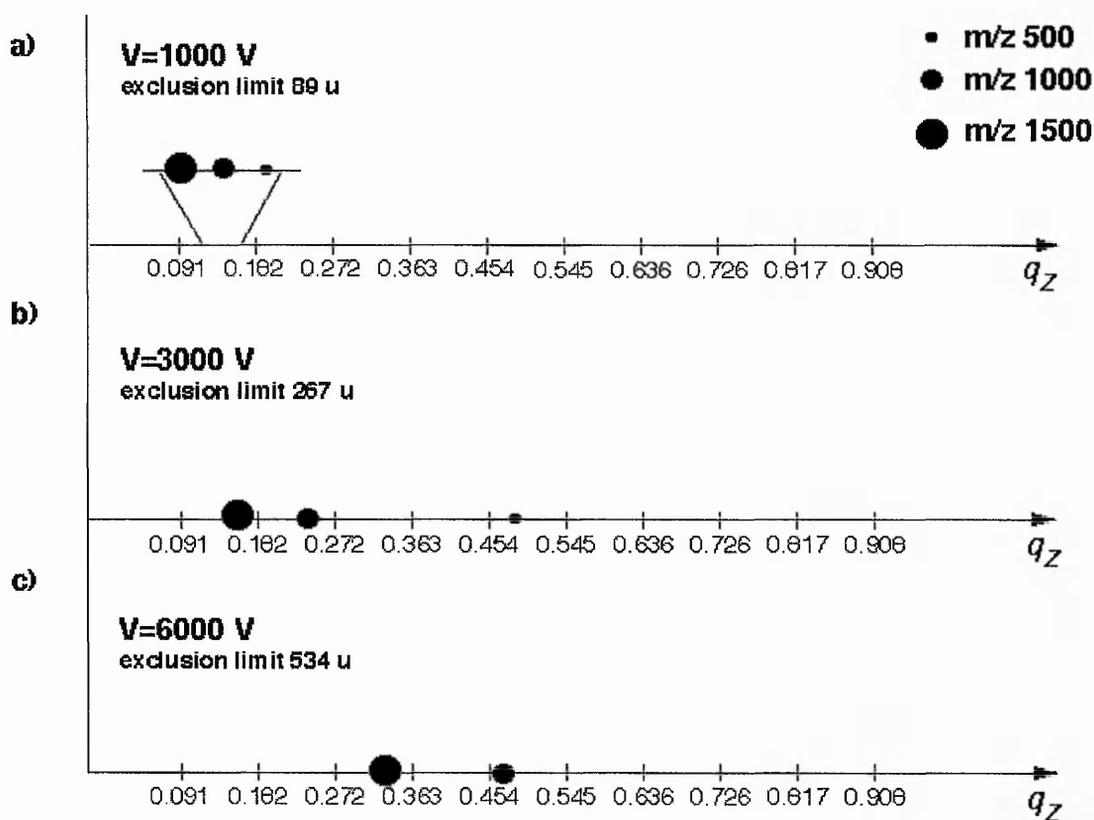


Figure 1.8: Simplified stability diagram for the ion trap ($a_z = 0$)²⁴

A main limitation of the QIT is poor resolution caused by space-charging effects, which can occur when the trap is overloaded with ions. The high density build up of stored ions leads to repulsion of like charges causing the ion cloud to disperse. Ions of the same m/z will have different working points of stability and will be ejected from the trap at slightly different times causing peak broadening to be observed. A function known as automatic gain control (AGC) was introduced to overcome this problem.²⁸ AGC performs a rapid pre-scan measuring the number of ions being allowed into the trap. The ionization time is calculated and set for the analytical scan to maximise the signal whilst minimising the space-charging effect.

1.1.5 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) is a powerful technique used to gain structural information on a particular ionic species. There are two main categories of instruments

capable of tandem mass spectrometry, firstly those with analysers joined in series (tandem in space), for example triple quadrupole and Q-TOF spectrometers, and those which trap the ions, such as QIT and FT-ICR (tandem in time). In-space instruments perform precursor ion selection, fragmentation and product ion separation in different regions of the mass spectrometer. The in-time instruments carry out all processes sequentially in the same region of the mass spectrometer.

1.1.5.1 Scan Modes for Tandem Mass Spectrometry

There are three main scan types which can be used for MS/MS, the most common being the product ion scan. This is when a precursor ion is isolated, fragmentation carried out by CID and the resulting product ions detected, shown in Figure 1.9.

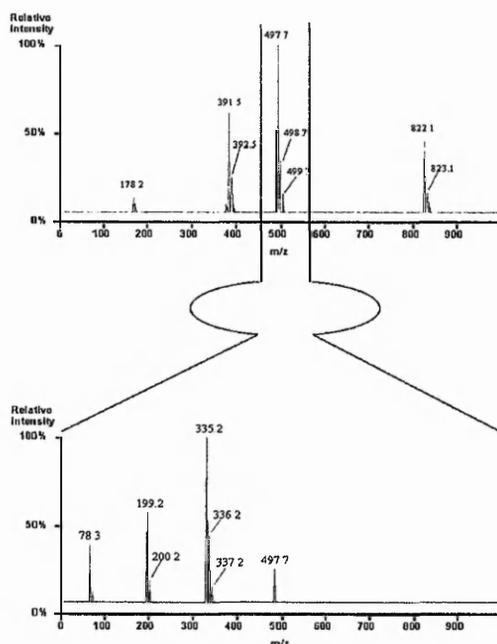


Figure 1.9: Schematic of tandem mass spectrometry product ion scan

Other modes of operation include the precursor ion scan, which results in the identification only of precursor ions that dissociate to give product ions of a specific m/z value. The neutral loss scan is used for monitoring analytes that have common neutral losses. These latter two

scan types cannot be performed with tandem in time instruments. However, the main advantage of tandem in time instruments is that the product ion scan can be repeated allowing multi-stage tandem mass spectrometry (MS^n) to be performed resulting in greater structure elucidation.

1.1.5.2 Collisionally Induced Dissociation (CID)

In order for a precursor ion to fragment, energy must be transferred to the ion. There are several activation mechanisms available, namely post-source decay (PSD) and in-source decay (ISD) which are used in TOF instruments. Collisionally induced dissociation (CID) used for inducing fragmentation in ion traps is the most widely used ion dissociation method for tandem mass spectrometry.²⁹ In a QIT, the ions of interest are selected by m/z and are isolated in the device by the ejection of all other ions from the trap. CID is carried out by the application of an auxiliary rf voltage applied to the end caps in resonance with the secular frequency, Ω_0 , (Eq. 11) of the precursor ion. Ions are accelerated, gain kinetic energy and collide more frequently with the helium buffer gas. Upon collision the kinetic energy is converted into internal energy and fragmentation will ensue if sufficient energy is deposited in the precursor ion. The product ions formed may be ejected to the detector and mass analysed. Alternatively, a product ion may be selected and retained in the trap for further stages of tandem mass spectrometry MS^n ($n \geq 2$).

CID may be categorised as either high energy or low energy. The latter is observed in a quadrupole ion trap since many collisions with the buffer gas are necessary to impart the energy required for fragmentation to occur. High energy CID is observed in magnetic sector or TOF/TOF instruments where fragmentation may occur through a single collision with the buffer gas. The maximum amount of energy converted into internal energy from the collision in the centre of mass frame of reference (E_{cm}) is given by equation 14.

$$E_{cm} = E_{lab} \frac{M_t}{M_p + M_t} \quad \text{Eq.14}$$

Where E_{lab} is the ion kinetic energy in the laboratory frame of reference, M_t is the mass of the buffer gas and M_p is the precursor ion mass. For low energy CID the E_{lab} value is within the eV range (typically 1-100 eV) and for high energy CID the value is in the keV range. Different spectra are obtained from the two collision energies, generally low energy CID favours facile bond cleavage resulting in a limited variety and quantity of product ions, whilst high energy CID produced significant fragmentation leading to more complex spectra. The fragmentation of peptide ions under CID is discussed in Section 1.3.

1.2 Ionization

Ionization occurs in the ion source region of the mass spectrometer, this may be held under vacuum or at atmospheric pressure. A wide variety of ionization techniques have been developed, although only a few are applicable to the analysis of biomolecules and these are outlined below.

1.2.1 Fast Atom Bombardment

Fast atom bombardment (FAB) was introduced in 1981 by Barber *et al.*³⁰ This technique consists of mixing the analyte with a low volatility liquid matrix, such as glycerol on a target. A beam of neutral atoms is produced by the ionization of an inert gas (argon or xenon) and focused onto the target.

Alternatively, a stream of atomic atoms (cesium) may be used. This related technique is termed liquid secondary-ion mass spectrometry (LSIMS). On collision with the sample a shock wave ejects the ions and molecules into the gas phase. The resultant ions are accelerated and directed towards the mass analyser. FAB is a relatively soft ionization technique that has now been largely replaced with ESI and MALDI.

1.2.2 Electrospray Ionization

The first report of electrospray ionization (ESI) was in 1968 by Dole³¹ and it has since been used primarily for the analysis of large biomolecules,^{32, 33} such as proteins,³⁴ peptides^{35, 36} and polysaccharides.^{37, 38} The principle of ESI is relatively straightforward, Figure 1.10. A solution of analyte flows to the tip of a narrow metal or metal coated glass capillary which is held at a high voltage (1-5 KV). A separation of charge occurs at the liquid surface emerging from the capillary as a result of the applied voltage, creating a fine aerosol of charged droplets. Evaporation of the solvent occurs which leads to a reduction in droplet size until the Rayleigh limit is reached, where the repelling coulombic forces between the ions in the droplet is equal

to the liquid surface tension. Due to the electrostatic forces between the ions the droplet breaks up into several smaller droplets. Protonated or cationised ions are eventually formed through the repetition of this process. There are two conflicting theories on how gas phase ions are produced from the charged droplets. The charge residue model proposes that some of the droplet's charge is retained as the last of the solvent evaporates. In contrast to this, the ion evaporation model suggests that the surface field on a droplet becomes strong enough to emit an ion into the gas phase.³⁹

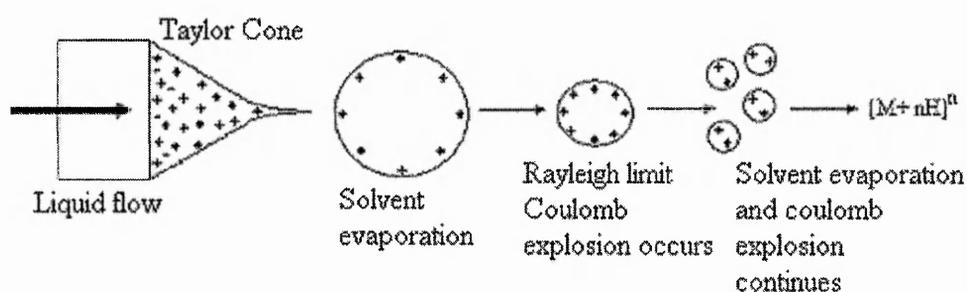


Figure 1.10: Schematic of gas-phase ion formation in the electrospray ion source

ESI generates multiply charged ions of the type $[M+nH]^{n+}$ as well as the singly charged $[M+H]^+$ species. This is particularly beneficial for the analysis of large biomolecules as it effectively reduces the mass to charge ratio of the analyte of interest. Since proteins can become multiply charged the reduced m/z usually values fall within the mass range of simple mass analysers, such as the quadrupole ion trap which generally have a mass range of $m/z < 2000$.

1.2.3 Matrix-Assisted Laser Desorption/Ionization (MALDI)

In 1978 Posthumos *et al*⁴⁰ were the first to apply laser desorption to organic mass spectrometry. In these early experiments, the sample was irradiated directly by the laser beam. Infrared (e.g. Nd/Yag – neodymium/yttrium aluminium garnet), pulsed CO₂ and UV lasers

were generally used. The main problem with this direct ionization mode, known as laser desorption ionization (LDI) or thermal laser desorption, is that the degree of energy transfer is difficult to control and this may lead to thermal degradation of the sample. Another problem with LDI is that not all compounds absorb at the laser wavelength. As a result of these drawbacks LDI is only applicable to a limited number of compounds, usually those with molecular masses below 1000 Da. Time-of-flight,⁴¹ magnetic sector⁴² and quadrupole ion trap⁴³ mass spectrometers have all been coupled to LDI.

In 1988 Tanaka *et al*⁴⁴ and Karas and Hillenkamp⁴⁵ independently, but almost simultaneously made a major breakthrough in the use of LDI. Both groups published mass spectra of monomeric or multimeric organic ions with masses up to 67 kDa. The ionization of such high mass analytes was achieved through embedding the sample into a matrix which absorbed the energy of a pulsed laser beam. This technique became known as matrix-assisted laser desorption/ionization (MALDI) and has led to the application of mass spectrometry to the analysis of proteins and other biomolecules with masses in excess of 300 kDa. The success of this technique is due to the mixing of the sample with a matrix. Tanaka employed metallic nanoparticles (finely divided platinum 10 nm size) suspended in glycerol as a matrix, in contrast, Hillenkamp used a chemical matrix of highly absorbing small organic molecules, particularly nicotinic acid. Of the two methods the later is generally considered to be superior, as a strong signal could be obtained from a few picomoles of sample, whilst with Tanaka's method several nanomoles of sample were required.

1.2.3.1 Mechanisms of MALDI

In MALDI samples are mixed with matrix and irradiated with a laser beam of short pulses ca 10-20 ns duration. The host matrix must absorb energy at the wavelength of the laser

radiation. The most commonly used laser is the nitrogen laser with a wavelength of 337 nm. Absorption of the energy from the laser beam leads to rapid heating and evaporation of the matrix. The resultant gas-phase plume contains analyte molecules, which become ionized via gas-phase proton transfer reactions. The MALDI process is illustrated in Figure 1.11.

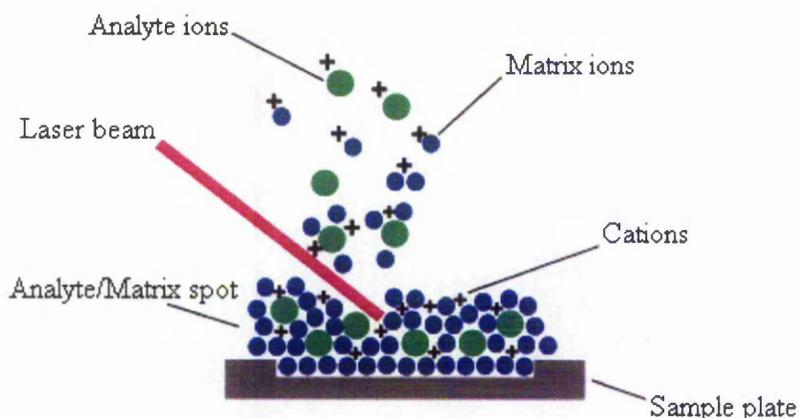


Figure 1.11: A diagram of the MALDI ionization process ⁶

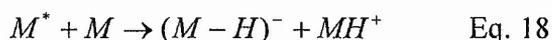
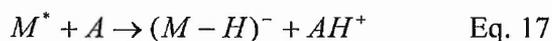
The mechanism of ion formation is not fully understood and several processes have been proposed in the literature. ⁴⁶⁻⁵⁰ There are two classes of ion formation mechanisms, namely primary and secondary. Primary ions are the initial ions formed from neutral molecules in the sample and are often matrix derived. Secondary ionization mechanisms are those which produce ions *via* subsequent reactions, which maybe either matrix-matrix or matrix-analyte reactions.

The most straightforward explanation of primary ion formation in MALDI is multiphoton ionization. Excitation of the matrix occurs *via* the absorption of a photon. The excited state can either absorb a second photon ejecting an electron leading to the formation of a matrix radical cation or return to the ground state to be re-excited.



This method is thought to be unlikely, as the ionization potential required for this process to occur is higher than the energy available from two nitrogen laser photons, since the photon fluxes typically used in MALDI is between 8-12 eV.⁵¹ However, multiple excited matrix molecules may pool their energies to form one matrix radical cation or one highly excited species.^{52,53}

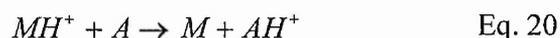
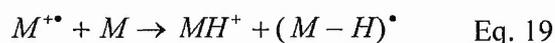
The excited-state proton transfer (ESPT) theory dates back to the origins of MALDI,⁴⁵ the acidity of the matrix is believed to be higher for the excited state as opposed to the ground state. This enables proton transfer reactions to occur between the excited matrix molecules and neutral molecules, forming protonated and deprotonated matrix molecules, Equation 17 and 18. These can react with the analyte molecules yielding positive and negative analyte ions respectively.



The main problem with this theory is that ESPT activity has not been established for common matrices and known ESPT compounds have not performed well as matrices.⁵⁴ Breuker *et al*⁵⁵ suggested the disproportionation pathway where protonated and deprotonated matrix molecules are formed through the excitation of a matrix cluster. Alternative primary mechanisms include thermal ionization^{47, 56-58} and desorption of preformed ions where ions are already present in the solid sample and are released by the laser pulse.⁵⁹⁻⁶¹

Proton transfer reactions are thought to be the most important secondary reactions occurring in the MALDI process, since the majority of analytes are observed as protonated species.⁶²

Radical matrix ions formed during the photoionization may transfer a proton to a neutral matrix molecule (M) which then acts as an intermediate to protonate analyte molecules (A), equation 19 and 20.



Electron transfer reactions may also occur where the charge is transferred from the matrix radical cation to the analyte forming an analyte cation. This type of reaction is less common and can cause analyte neutralization.

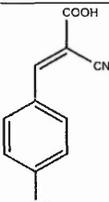
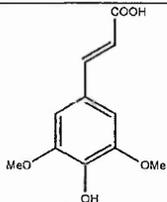
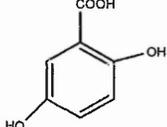
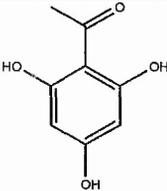
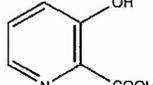
An important feature of MALDI is that molecules are desorbed in the gas phase without undergoing thermal degradation and it is possible to carry out positive and negative ion analyses. MALDI predominantly generates singly charged molecular ions $[M+H]^+$, although, occasionally it is possible to identify doubly and triply charged protonated ions of low abundance. A common feature of MALDI spectra of biological samples is the formation of Na^+ and K^+ adducts which may be preformed in the solid matrix prior to laser ablation or formed by ion-molecule reactions in the gas-phase.

1.2.3.2 MALDI Matrices

The matrix has several important roles in MALDI, as well as absorbing the energy of the laser it also serves to protect the analyte by minimising sample damage caused by direct laser ablation. Upon irradiation, the matrix is vaporized and therefore some of the co-crystallised sample is ejected into the gas phase, allowing the matrix to participate in the ionization mechanisms. The matrix also reduces intermolecular forces minimising aggregation of sample molecules because the matrix is present in vast excess (usually a matrix: analyte ratio of 1000:1 is used). Matrices may be described as either “hot” or “cold” depending upon the amount of energy imparted on to the analyte at the time of ionization. For example CHCA is a “hotter” matrix than DHB as more energy is transferred to the analyte ions inducing more fragmentation. A possible explanation for “hot” and “cold” matrices is the temperature at

which they sublime.⁶³ Common matrices for various types of biological analytes are described in Table 1.1.

Table 1.1: Common MALDI matrices and their biological applications

Matrix	Structure	Application
α -cyano-4-hydroxy-cinnamic acid (CHCA)		Peptide and protein digests (Analytes <10 kDa)
4-hydroxy-3,5-dimethoxycinnamic acid (Sinapinic acid, SA)		Large polypeptides and proteins (Analytes > 10 kDa)
2,5-Dihydroxybenzoic acid (DHB)		Protein digests and proteins, Oligosaccharides
2,4,6-trihydroxyacetophenone (THAP)		Oligonucleotides < 3 kDa
3-hydroxy picolinic acid (3-HPA)		Oligonucleotides > 3 kDa

1.2.4 Atmospheric Pressure MALDI

Matrix-assisted laser desorption/ionization at atmospheric pressure (AP-MALDI) is based on the same principles as MALDI carried out *in vacuo*, in which the sample is mixed with an excess of matrix and irradiated with a pulsed laser beam. The matrix absorbs energy causing rapid heating and the resultant expansion plume carries the analyte into the gas-phase leading to ionization by gas phase proton transfer or cationization. AP-MALDI therefore has many

features in common with vacuum MALDI, including application to a wide range of biomolecules and polymers and use with TOF-MS and ion trapping analysers.

1.2.4.1 AP-MALDI Combined with Time-of-flight Mass Spectrometry

The first report of AP-MALDI combined with mass spectrometry was in 1999, when ionization at atmospheric pressure was combined with an orthogonal acceleration time-of-flight mass spectrometer (TOFMS).⁶⁴ The technique was demonstrated for the investigation of the peptide composition of paralytic poisons, contoxins, from a molluscivorous snail. The same instrumental configuration was later demonstrated for the analysis of a peptide mixture, the tryptic digest of bovine fetuin and the oligosaccharide difrucosyllacto-N-tetraose.⁶⁵ More recently, AP-MALDI incorporating a high repetition rate Nd:YAG laser (1kHz) has been coupled with a quadrupole-time-of-flight mass analyser (QqTOF).⁶⁶ In comparison with low repetition rate nitrogen lasers (< 30 Hz) which are usually used for AP-MALDI, the new configuration showed a considerable increase in sample throughput. This may be useful in proteomics applications that often require high sample throughput. On-line coupling of HPLC with AP-MALDI/TOFMS has been reported for the analysis of polymer and peptide standards.⁶⁷ The HPLC effluent was mixed with a matrix solution and the analyte/matrix mixture droplet formed at the end of the capillary was desorbed and ionized by a high repetition rate Nd:YAG laser beam focused onto the drop. Ions formed were guided *via* electric fields into the inlet of the mass spectrometer. Results for the analysis of angiotensin II obtained with the Nd:YAG laser were compared to those obtained with a 20 Hz nitrogen laser. The signal-to-noise ratio was improved by a factor of two using the higher repetition rate laser. Standing and coworkers⁶⁸ introduced an intermediate pressure MALDI source, operating at pressures near 1 Torr, with orthogonal injection of desorbed ions into a quadrupole TOF-MS. Comparable results to vacuum MALDI were obtained, except that the effect of target position

and laser conditions were less critical and fewer metastable fragments were observed at elevated pressures than for vacuum MALDI.

AP-MALDI has also been combined with atmospheric pressure ion mobility spectrometry (IMS) for the study of peptides, pharmaceuticals and synthetic polymers⁶⁹ and with atmospheric pressure IMS combined with an orthogonal acceleration reflector TOFMS for the analysis of dipeptides and biogenic amine standards.⁷⁰ Solid phase microextraction (SPME) has been coupled with AP-MALDI and the combination interfaced with an ion mobility mass spectrometer operated at ambient pressure and quadrupole/TOF-MS.⁷¹ Analytes were deposited onto the end of a silanized fiber optic which was then either inserted into the IMS source region or aligned with the MS inlet.

1.2.4.2 AP-MALDI Combined with Ion Trap and Fourier Transform Mass Spectrometry

The development of AP-MALDI interfaced with a quadrupole ion trap enables mass spectrometry and multi-stage tandem mass spectrometric analysis to be carried out on AP-MALDI generated ions.⁷²⁻⁷⁴ The MSⁿ capabilities of the AP-MALDI/QIT MS configuration is particularly important in the analysis of biological samples, because it yields sequence and structural information. In early experiments the AP-MALDI ion source was coupled to the ion trap *via* an open interface with reflection geometry, in which the laser ablation and ion extraction occur on the same side of the sample as the incident laser beam. A typical arrangement is shown in Figure 1.12. A flow of nitrogen gas was directed across the target surface to aid transportation of ions into the capillary inlet of the mass analyser. An increase in sensitivity was obtained when the gas stream was employed,^{72, 73} although this was not found to improve sensitivity in one configuration.⁷⁴ However, a moisture-free environment for desorption may be created by introducing dry nitrogen gas. In all configurations, sensitivity was affected by the laser threshold, the position of the target, the voltages applied to the target

and MS inlet and the location and flow rate of the nitrogen gas. A counter current heated gas stream directed towards the target plate was reported to increase sensitivity by two orders of magnitude for tryptic peptides.⁷⁵

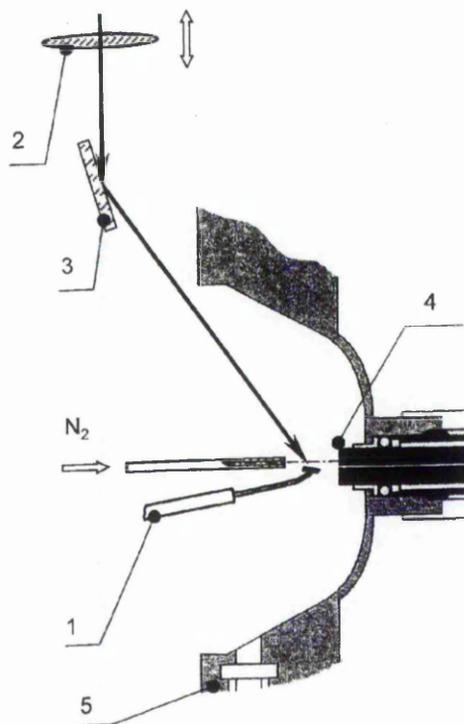


Figure 1.12: Schematic diagram of an AP-MALDI interface: 1, target holder; 2, quartz lens; 3, mirror; 4, stainless steel capillary inlet; 5, inlet flange of the atmospheric pressure interface of an ion trap mass spectrometer⁷²

A glass tube interface which also utilised reflection geometry, Figure 1.13, was reported by Glish *et al.*⁷⁶ The sample was coated on the inside of a glass capillary and desorbed by introducing an optical fibre coupled to an Nd:YAG laser into the other end of the tube. A heating coil was later added to this design to decrease cluster formation and to increase sensitivity.⁷⁷

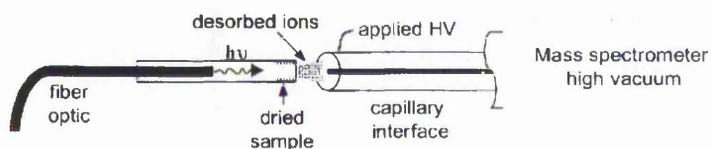


Figure 1.13: Schematic diagram of a glass tube AP-MALDI interface⁷⁶

An on-axis reflection geometry was used in the first commercial AP-MALDI interface, which was introduced in 2001, for the ThermoFinnigan LCQ ion trap mass spectrometer.⁷⁸ A schematic of this interface is shown in Figure 1.14. In this configuration the target plate is in line with the atmospheric pressure inlet of the ion trap and an extended capillary is fitted to aid transportation of ions into the heated capillary inlet.

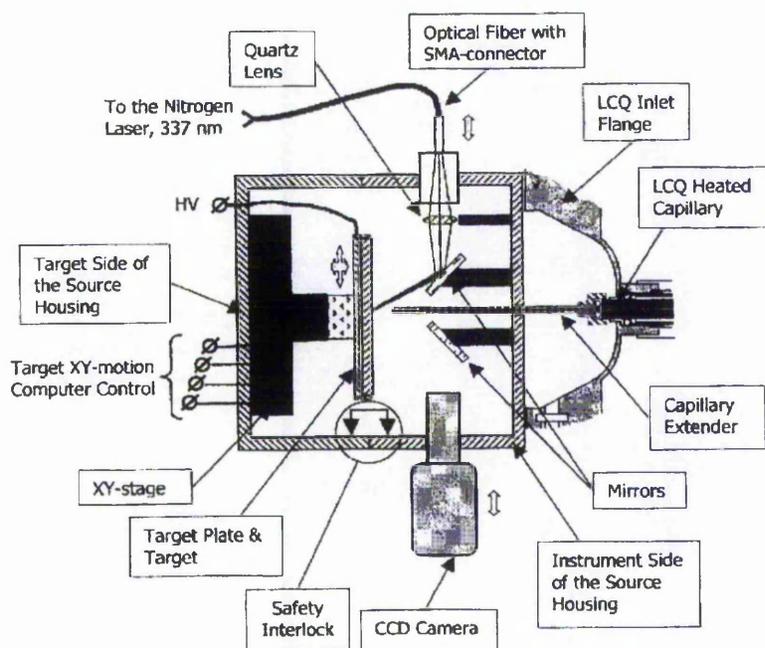


Figure 1.14: Schematic diagram of a commercially available AP-MALDI source⁷⁸

Transmission geometry for AP-MALDI is shown in Figure 1.15, in which the sample is deposited onto a glass slide and the incident laser beam passes through the glass and irradiates the sample from behind.⁷⁹

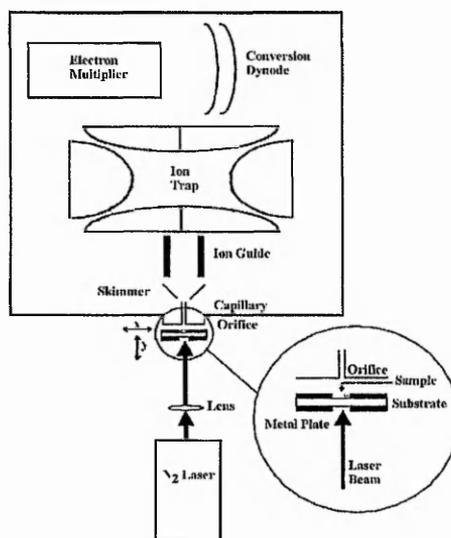


Figure 1.15: Schematic diagram of a transmission geometry AP-MALDI source ⁷⁹

Callahan *et al* ^{79, 80} compared AP-MALDI in transmission and reflection geometry, concluding that the latter results in analyte-matrix clusters, but higher signal intensities. Higher laser intensity was required for transmission geometry (150-190 $\mu\text{J}/\text{pulse}$) than reflection (25 $\mu\text{J}/\text{pulse}$). Transmission geometry AP-MALDI has been used to examine the ionic components of the target plume produced in matrix-assisted pulsed laser evaporation (MAPLE).⁸¹ MAPLE uses a similar process to MALDI for the deposition of polymer thin films. The polymers are embedded into the matrix, irradiated with a pulsed laser beam and carried into the gas-phase by the resulting plume, which occurs when the laser energy is absorbed by the matrix. A substrate is placed in the path of the plume to capture the ejected polymer molecules leading to the formation of polymer thin films.

The transmission efficiency of ions into the vacuum of the mass spectrometer has been improved through the introduction of pulsed dynamic focusing (PDF).⁸² AP-MALDI usually operates with a static electric field applied to the target plate to aid transportation of ions to the capillary inlet, but in PDF a pulsed high voltage extraction field is applied to the target, where ground potential is reached just before ions enter the vacuum. Reducing the applied voltage

prior to ions entering vacuum allows them to follow the gas stream directed into the mass spectrometer inlet, whereas in conventional AP-MALDI the ions follow the electrical field lines often resulting in collision with the inlet. PDF is reported to improve sensitivity by up to an order of magnitude.⁸²

An increase in sequence coverage was reported for a tryptic digest desorbed from porous silicon (DIOS) targets compared with conventional AP-MALDI targets.⁸³ The analytes are retained by the etched silicon wafer and the absorptivity of laser energy leads to ionization. As matrix application is eliminated, analyte/matrix cluster ions are not observed in the spectra, which leads to an improvement in sensitivity. A further advantage of DIOS is the ease of sample preparation. Combining the peptide mass fingerprint data from both AP-MALDI and AP-DIOS led to an increase in sequence coverage with nearly double the number of peptide matches obtained, compared to each technique employed separately, as different peptides were identified by the two techniques. For the lower molecular mass region (280-1000 Da), AP-DIOS produces significantly more peptide matches than AP-MALDI. However, the opposite is true for the higher mass region, 1000-1800 Da. The two techniques have also been used for the identification of amphetamines in forensic samples.⁸⁴

AP-MALDI has been combined with a digital ion trap (DIT), which allowed an extended mass range to be observed.⁸⁵ The advantage of this approach is the capability of performing mass spectrometric analysis of higher mass ions, which fall outside the mass range of conventional ion traps. Tandem mass spectrometry of the triply charged insulin- β -chain ion (m/z 1166) and mass spectrometry of horse heart myoglobin (m/z 16980.8) was reported, demonstrating the application to intact proteins.

A complimentary technique to AP-MALDI exhibiting increased sensitivity for the analysis of small organic analytes has recently been reported.⁸⁶ The technique consists of an atmospheric pressure laser desorption/chemical ionization (AP-LD/CI) source, which employs an IR laser

to generate gas phase neutral molecules. Ionization is then carried out by proton-transfer chemical ionization using a corona discharge. A 150-fold increase in the signal was observed for spirerone and reserpine compared to IR AP-MALDI. AP-LD/CI and IR AP-MALDI, have also been compared for peptide analysis, with a 100 fold increase in signal intensity reported for the former technique.⁸⁷ In addition, this study showed that the incorporation of trifluoroacetic acid into the matrix solution for IR-AP-MALDI leads to a 10 fold increase in the intensity of the protonated molecule.

AP-MALDI has been combined with Fourier transform mass spectrometry (FTMS) for the analysis of complex peptide mixtures,⁸⁸ oligosaccharides⁸⁹ and RNA.⁹⁰ Spectra obtained via AP/MALDI-FTMS are comparable to ESI-FTMS with respect to resolution and accuracy. No metastable decay was observed or significant formation of matrix adducts, which were problematic with vacuum MALDI-FTMS.⁹¹ A standard vacuum MALDI source has been modified to allow transient high pressure during laser ablation and prior to mass analysis.^{92, 93} This work, although not strictly AP-MALDI, combined the high sensitivity of the vacuum source geometry with the much lower fragmentation inherent in the high pressure approach for the analysis for oligosaccharides.

1.2.4.3 Applications of AP-MALDI

AP-MALDI-TOF has been applied to the study of polymers,⁶⁷ native peptides,⁶⁴ tryptic peptides,⁶⁶ peptide mixtures⁹⁴ and oligosaccharides.⁶⁵ IMS analysis has been used for the study of peptides, pharmaceuticals and polymers.⁶⁹⁻⁷¹

AP-MALDI interfaced with ion trap mass spectrometry has been applied to the analysis of intact proteins,⁸⁵ peptide mixtures,⁷² tryptic peptides^{75, 78, 82, 95} and peptide derivatives.⁷³ The identification and characterization of post-translationally modified peptides⁹⁶ and phosphate loss from phosphopeptide ions has been reported.^{97, 98} Complex peptide mixtures have also

been studied using AP-MALDI combined with FTMS.⁸⁸ The use of AP-MALDI /QIT for the determination of acetylation sites in histone proteins, sulfate modifications, phosphorylation and peptide-peptide interactions has been described.⁹⁹ Acetylation modifications were identified through the use of tryptic digestion and tandem mass spectrometry of the resulting peptide ions.

Although the majority of experimentation utilizing AP-MALDI/QIT has focused upon peptides and proteins, the technique has also been applied to oligosaccharides^{74, 99} and synthetic polymers.^{79,100} Wei and Co-workers¹⁰¹ reported improved signal-to-noise ratio and sensitivity for neuropeptides when the matrix/analyte mixture was electrospray deposited onto the target surface compared to other sample deposition techniques. This application method leads to an increase in sample homogeneity and greater reproducibility. Spectra obtained by both AP-MALDI/QIT and vacuum MALDI-TOF mass spectrometry were similar except that vacuum MALDI showed greater sensitivity.

IR lasers incorporated into AP-MALDI sources have been applied to the analysis of oligosaccharides,¹⁰²⁻¹⁰⁴ peptides,^{87, 105} glycolipids¹⁰⁶ and posttranslational modified peptides¹⁰⁷ as discussed in section 1.2.4.5.

1.2.4.4 Comparison of AP-MALDI and Vacuum MALDI

AP-MALDI and vacuum MALDI have many features in common, including application to biomolecules and polymers, minimal sample clean-up, ease of sample preparation, tolerance of interference and salts and potential for multiple analyses from the same sample spot. The matrices employed, ratio of matrix to analyte and laser beam energy densities at the target surface are also similar in many cases.⁶⁵ Keough *et al*⁷³ demonstrated that the AP-MALDI tandem mass spectra obtained for sulfonic acid derivatized tryptic peptides are similar to postsurface decay spectra obtained *via* vacuum MALDI. However, there are some important

differences between the two ionization techniques. One obvious advantage of AP-MALDI, compared to vacuum MALDI, is that vacuum compatible samples are not needed for AP-MALDI and the ion source is easily interchangeable with other atmospheric ionization sources, such as an electrospray and APCI, on the same instrumental platform. AP-MALDI is also compatible with use of liquid matrices, without the difficulties of source contamination experienced with vacuum MALDI.¹⁰⁸ Turney *et al*⁹⁴ investigated AP-MALDI with liquid matrices consisting of a solid MALDI matrix (e.g. α -cyano-4-hydroxycinnamic acid) doped into a liquid medium consisting of a suitable solvent and diethanolamine (DEA). Results showed improved shot-to-shot reproducibility in comparison with solid state samples as well as prolonged signal lifetime. These improvements in reproducibility and lifetime with liquid matrices were attributed to diffusion from the solution rejuvenating ablated spots within the target. Other groups have experimented with liquid matrices for IR-AP-MALDI and these are discussed in section 1.2.4.5.

In comparison with vacuum MALDI, where fragment ions may sometimes be observed,^{109, 110} there is evidence that AP-MALDI is a softer ionization process. This makes the technique a better tool for protein analysis⁶⁵ in cases where facile fragmentation is observed in vacuum MALDI. For example, fragments formed during laser desorption of poly(ethylene glycol) mixtures can be observed in vacuum MALDI spectra.¹⁰⁹ Comparison of vacuum MALDI and AP-MALDI spectra of sulfo- and phosphopeptides showed that AP-MALDI was capable of detecting intact molecular ions of these labile species⁶⁴ whereas only dephosphorylated ions were observed for phosphopeptides using vacuum MALDI operated in the positive ion mode. Gabelica recently reported an extensive investigation of the internal energies of MALDI generated ions under vacuum and atmospheric pressure conditions, using benzylpyridinium cations as probes.¹¹¹ These studies showed that ion effective temperatures found in AP-MALDI were lower than those observed in vacuum MALDI. However, the ion effective

temperature for the benzylpyridinium ions was strongly dependant on the matrix, with the relative hardness of the matrix (CHCA > DHB > SA) correlating with the matrix sublimation temperature.

Mehl and co-workers⁹⁵ compared AP-MALDI/QIT-MS with MALDI-TOF and microcapillary liquid chromatography/tandem mass spectrometry (μ LC/MS/MS) for automated protein identification. Protein sequence coverage was lowest for AP-MALDI although sample throughput was reported to be 10-fold greater than for μ LC/MS/MS. The study concluded that AP-MALDI was a more selective and robust technique than MALDI-TOF MS.

AP-MALDI mass spectra, unlike vacuum MALDI, are usually characterised by the presence of analyte-matrix cluster ions formed by the desorption process. These ions arise as a result of collisional cooling, which stabilises the cluster ions allowing them to be transferred into the mass analyzer. In vacuum MALDI there is no equivalent cooling process, so dissociation of the adducts occurs. Analyte-matrix ions of the type $[X + nM + H]^+$ are observed for peptides ionized by AP-MALDI, where X represents the peptide and M represents the matrix. The occurrence of such clusters leads to a decrease in sensitivity because the available ion current is distributed over several ions. Creaser *et al*¹⁰⁰ compared AP-MALDI-QIT with vacuum MALDI-TOF mass spectrometry for the analysis of synthetic polymers in the presence of alkali metal salts. Clusters of the type $[PEG + matrix^- + 2 metal]^+$ and $[PEG + metal]^+$ were observed in AP-MALDI for all matrix/cation combinations. In contrast, vacuum MALDI resulted in the presence of only alkali metal adducts and no matrix adduct formation was observed. However, more sodium adduct interference was observed for vacuum MALDI indicating that AP-MALDI/QIT maybe more tolerant to the presence of salts. The degree of cluster formation was highly dependent on the matrix and cation used to dope the sample. Complex formation was most effective when cationised with potassium ions and least effective with lithium ions. Esculetin matrix resulted in the lower intensities for the matrix

adduct peaks than α -cyano-4-hydroxycinnamic acid. It was possible to reduce the intensity of matrix/analyte adducts by the use of a higher capillary inlet temperature or by collisionally-induced dissociation (CID) in the mass spectrometer interface.^{72, 100} Comparing AP-MALDI spectra obtained for PEG 1500, from an esculetin matrix doped with sodium chloride with and without application of in-source CID, showed a reduction in the intensity of the matrix adduct peaks by approximately 50 % following application of CID.

1.2.4.5 IR AP-MALDI

AP-MALDI coupled to an ion trap mass spectrometer has been adapted to incorporate an IR laser.¹⁰⁵ In comparison to a nitrogen UV laser (337 nm), the use of an IR laser with a 3 μ m wavelength showed improved performance with a wide variety of matrices, for example glycerol, water and aqueous α -cyano-4-hydroxycinnamic acid (CHCA). UV and IR AP-MALDI both resulted in less fragmentation for peptides than that observed with vacuum MALDI. Mass spectra of peptides in the mass range up to 2000 Da were obtained directly from aqueous solutions using this technique. In contrast to UV AP-MALDI, no matrix-analyte clusters were observed in the IR AP-MALDI spectra. Infrared AP-MALDI/QITMS has also been applied to the analysis of carbohydrates,⁹⁹ post-translationally modified peptides¹⁰⁷ and the fragmentation of cationized phosphotyrosine containing peptides.⁹⁸ Sugar-sugar complexes were studied by IR-AP-MALDI using a glycerol liquid matrix, because the near physiological conditions were suitable for non-covalent complex formation.¹⁰³ The majority of solid MALDI matrices are unsuitable because their acidic nature disrupts these non-covalent complexes. The use of liquid IR-AP-MALDI has also been studied for the analysis of sialylated carbohydrates¹⁰² and the fragmentation of cationized carbohydrates.¹⁰⁴ Frozen samples have been investigated using IR-AP-MALDI for the analysis of peptides, carbohydrates and glycolipids.¹⁰⁶ The use of water as the IR absorbing matrix allows samples to be analysed at

near physiological conditions, but water readily evaporates under IR irradiation and freezing the sample extended the lifetime of the analysis. A Peltier-cooled sample stage was used to generate the frozen samples, which consists of a ceramic thermoelectric (Peltier) cooling plate adhered to a stainless steel target plate, with the heating side of the device attached to an aluminium heat sink. Tandem MS/MS of the oligosaccharide 6'-sialyllactose was carried out demonstrating the capability for obtaining structural information from frozen samples.

1.3 Analysis of Peptides by Mass Spectrometry

Mass spectrometry is a versatile technique for the analysis of proteins and peptides, allowing not only accurate mass measurement but also providing sequencing information through the use of tandem mass spectrometry.¹¹²⁻¹¹⁶ The most common ionization methods used to analyse proteins and peptides are MALDI and ESI. These techniques are characterised by an absence of fragment ions, which enables complex analyte mixtures to be analysed without the need for pre-separation. Sequence elucidation of a peptide from a complex mixture is achieved through the use of tandem mass spectrometry, since a single peptide of a specific m/z may be isolated and fragmented using CID, as described in section 1.1.5.2.

When peptides are subjected to CID at low energy fragmentation commonly occurs at the amide bond along the backbone, generating a ladder sequence, Figure 1.16. The charge may be retained at the N-terminus resulting in the formation of fragment ions designated as a_n , b_n and c_n , or at the C-terminus when x_n , y_n and z_n type ions are observed. This fragmentation nomenclature was first described by Roepstorff and Fohlmann in 1984.¹¹⁷ The amino acid sequence can be determined by calculating the differences in masses of adjacent sequence ions of each series.¹¹⁸

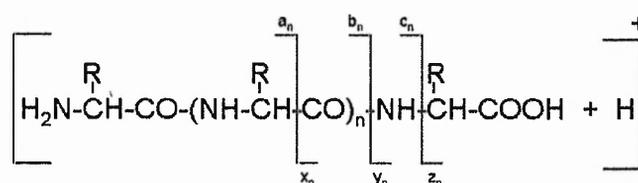


Figure 1.16: Peptide Fragmentation resulting from Low energy CID MS/MS

When the cleavage occurs one carbon away from the amide bond in either direction with the charge retained by the C-terminus an x_n or z_n ion is formed.¹¹⁹ If the charge remains on the N-terminus then a_n or c_n ions are formed. Other product ions may be formed through the loss of water, carbon monoxide, ammonia or amino acid-specific low mass ions. The loss of both the

N and C terminuses results in internal ion formation, Figure 1.17. High energy CID may provide further structural information through the side-chain-specific fragments, Figure 1.18.



Figure 1.17: Internal fragment formation by multiple cleavage of the peptide backbone

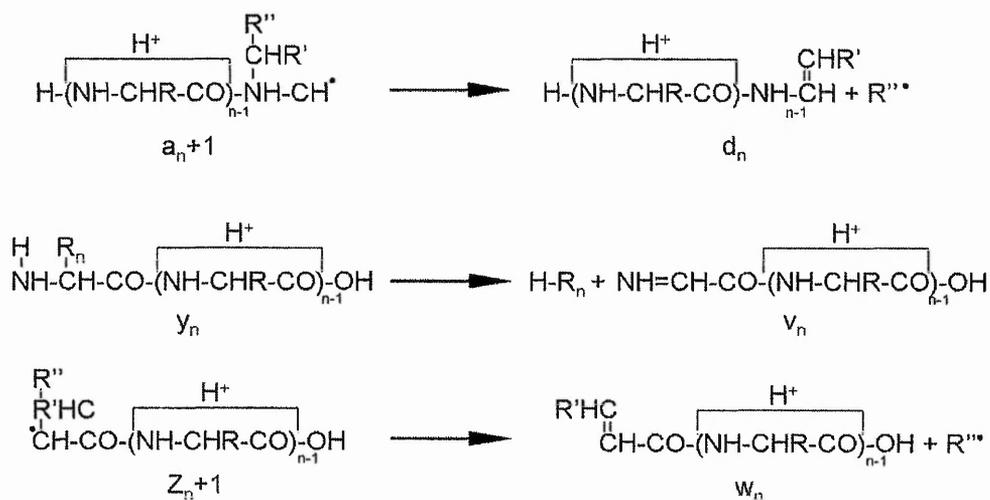


Figure 1.18: Fragmentation ions observed in high energy peptide ion fragmentation

The fragmentation process is influenced by the position of basic amino acids, for example arginine, lysine, proline and histidine, since protonation occurs at more basic sites.

Location of the basic residue towards the C terminal results in the formation of y ions, whereas location on the N terminal side results in predominantly b ions. Small peptides that do not contain basic amino acids tend to show similar abundance of b and y ions in their MS/MS spectra, whereas longer chains tend to favour b ion formation.

The interpretation of fragmentation spectra can be complicated as it is often difficult to establish whether a set of ions corresponds to the b or y series. To overcome this problem, peptide modification may be carried out so that a single ion series is favoured in the fragmentation pathway. For example derivatization with N-succinimidyl-2-(3-pyridyl) acetate encourages formation of b type ions.¹²⁰ Other useful derivitization methods include the

introduction of ^{18}O labelling of the C terminal of tryptic peptides during digestion allowing the two series to be easily distinguished in the MS/MS spectrum, as the y ions are identified by the presence of ^{18}O in their isotopic distribution.¹²¹

Another major problem encountered with peptide sequencing using mass spectrometry is the difficulty in distinguishing between isomeric and isobaric amino acids. For example, leucine (L) and isoleucine (I) both have a monoisotopic mass of 113.08, and lysine (K) and glutamine (Q) are isobaric with a mass of 128.0. High energy CID and chemical derivitization have both been used to overcome these problems. High energy CID may be used to promote d and w fragments allowing identification of L and I.¹²² Acetylation of the amine group of lysine using acetic anhydride differentiates between K and Q because the masses of the K containing fragments are shifted to higher mass by 42 mass units.¹²³

1.4 Proteomics and Strategies for Proteome Analysis

The term proteome refers to the complete profile of proteins expressed in a tissue, cell or biological system at a given time. Proteomics is generally described as the study of the proteome. Other definitions of proteomics include the analysis of proteins and their functions, the study of all the proteins expressed in a cell, large scale analysis of the structure and function of proteins as well as protein-protein interactions and the systematic analysis of the protein expression of healthy and diseased tissues. The main differences between proteomics and protein biochemistry are summarised in Table 1.2. Proteomics is the study of multi-protein systems, the main purpose is the characterisation of the behaviour of the system as opposed to the behaviour of single components.

Table 1.2: Proteomics and Protein Chemistry

Proteomics	Protein Chemistry
Complex mixtures of proteins	Individual proteins
Partial sequence analysis	Full sequence analysis
Emphasis on protein and biomarker identification	Emphasis on structure and function
Systems biology	Structural biology

Proteomics is particularly challenging due to the complexity of the proteome which is a dynamic system that constantly changes. The vast number of proteins involved is problematic as proteins may be present in many forms due to posttranslational modifications. Concentrations ranges over which these proteins are present ($< 1 \text{ pg mL}^{-1}$ to $> \text{mg mL}^{-1}$) also causes difficulties. The range of plasma proteins is shown in Figure 1.19.

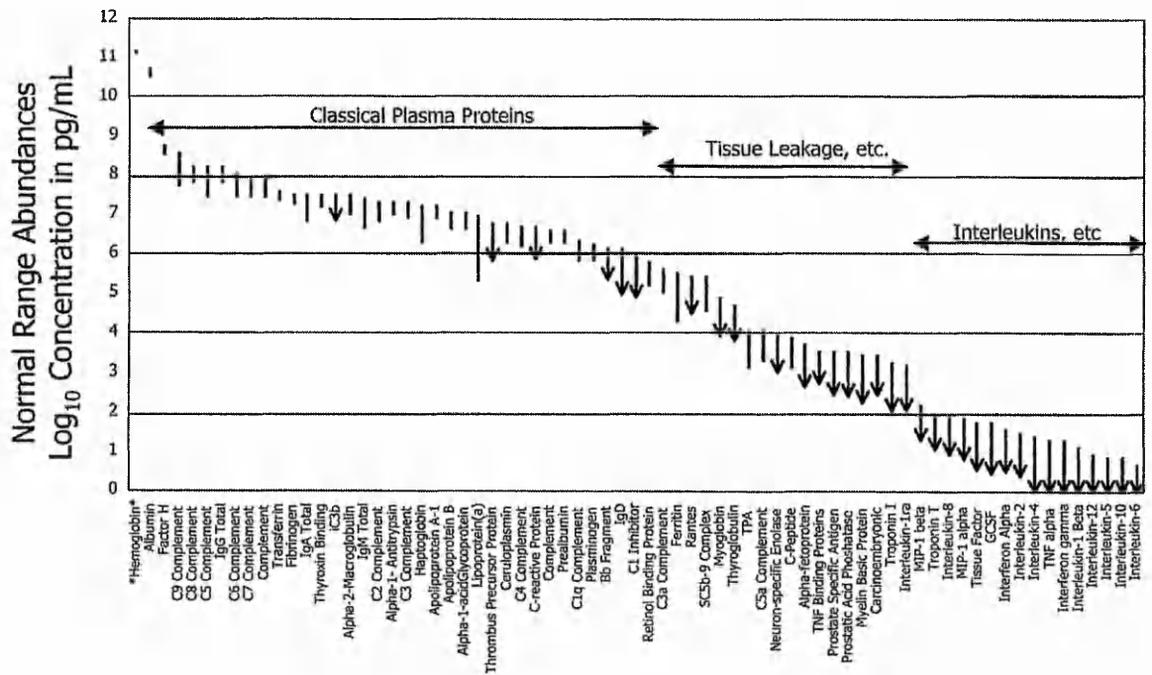


Figure 1.19: Abundances of plasma proteins ¹²⁴

The aims of proteomic research may be classified into four main categories. Protein mining is the identification of as many proteins as possible in a sample. Protein expression profiling is the comparison of specific proteins/peptides in healthy and diseased states. Protein network mapping is the identification of interacting proteins and protein modification mapping is the identification of protein post-translational modifications. Various proteomic tools, including mass spectrometry, database search algorithms and protein separation techniques may be used to achieve these aims and are outlined below.

1.4.1 Proteomic Strategies Using Mass Spectrometry

Mass spectrometry is a particularly attractive approach because of the potential for automation, high sensitivity, low sample volume requirements, and the capability of providing accurate mass measurements and sequencing information.

There are two main approaches to protein identification. The “Top down” approach focuses upon analysis of intact proteins and the fragmentation of these proteins in the gas phase. Accurate protein molecular weight can be determined by direct mass spectrometry analysis. If MS/MS is carried out on the intact protein, the entire sequence is available allowing complete sequencing of the protein and any associated modifications.¹²⁵⁻¹²⁷ This method was initially demonstrated for ESI generated ribonuclease A ions, using a triple quadrupole instrument.¹²⁸ Since then many instruments have been used, including MALDI-TOF which generates the fragment ions via in-source decay (ISD). The problem with the top-down approach to proteomics is that MS/MS data cannot be generated for the large proteins. The alternative approach commonly used for the identification of proteins, the “bottom-up” approach, either involves direct proteolytic digestion of the proteome followed by separation of the fragments or the pre-separation of protein mixtures prior to proteolytic digestion. These strategies are shown schematically in Figure 1.20. In both cases the resulting peptides are analysed by mass spectrometry. The most commonly used strategy is to generate a peptide mass fingerprint (PMF), this involves protein separation by 1D or 2D- sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), removal and subsequent in-gel enzymatic digestion (usually with trypsin) of a selected protein spot. The resulting peptides are eluted from the gel plug and analysed by mass spectrometry. The combination of SDS-PAGE with MALDI-TOF provides a powerful separation technique with high mass accuracy for protein identification *via* a PMF.

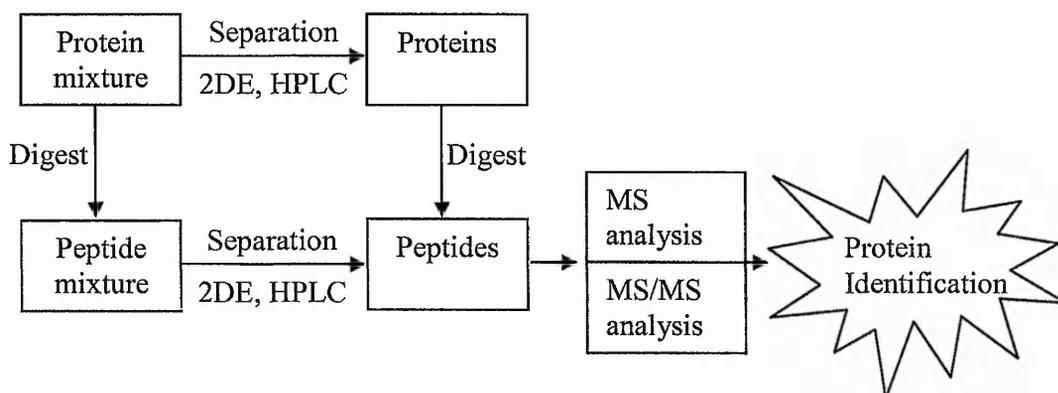


Figure 1.20: Schematic of bottom up proteomics

1.4.2 Protein Separation

Separation of the protein mixtures is necessary to reduce the complexity of the sample prior to analysis. The main protein separation techniques include 1D and 2D- sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and chromatography, for example HPLC.

1D SDS-PAGE separates proteins on the basis of molecular weight, since SDS binds to the protein imparting negative charge to the protein in proportion to molecular weight. Application of a high voltage to the gel causes the SDS-protein complexes to separate. The main problem associated with 1D gels is poor resolution, a band may contain more than one protein. To overcome this problem 2D-SDS-PAGE may be used. In this approach proteins are separated first according to the charge (isoelectric point) and then according to mass of the protein. Visualisation of the protein spots is possible through the application of dyes, such as coomassie brilliant blue and silver staining. The ability to image map and compare proteomes, allowing changes in patterns to be observed, is the main advantage of 2D-SDS-PAGE. The molecular weight of the sample proteins can be estimated using co-migrated standards with known molecular weights. The main problem with this method is limited accuracy, SDS-PAGE only provides an estimate of the molecular mass and not exact mass. Other disadvantages of the technique include low throughput as it is difficult to automate, variations

between gels and relatively long analysis time. Denaturing conditions are used and hence the 3D structure of the protein is lost and therefore protein-protein interactions cannot be studied *via* this technique. Several large hydrophobic proteins are incompatible with the isoelectric focusing (IEF) step, protein precipitation may occur leading to smearing of the proteins in the IEF strip rather than focusing into distinct bands. This smearing effect is also observed in the subsequent SDS-PAGE. Another disadvantage of 2D gels is the limited dynamic range for protein detection, since low abundant proteins may not be detected. Despite the drawbacks of 2D-SDS-PAGE the method is widely used, as complex samples can be visualised and quantitative estimates of proteins and their molecular weights can be made. The separated protein spots may be cut from the gel, enzymatically digested and separated by HPLC, their retention times determined and peptide sequence obtained by mass spectrometry or Edman chemistry.¹²⁹ This is similar to peptide ladder sequencing whereby an enzyme is used to digest the terminal amino acids from peptides. Disadvantages of Edman sequencing include the lengthy sample preparation, the relatively large amounts of protein required and the peptides must be pure for Edman sequencing.

The proteolytic digestion of the proteins is carried out by an endoprotease, such as trypsin or chymotrypsin which cleaves the protein at specific sites forming several smaller peptide chains whose sequence can be determined. Trypsin cleaves the protein backbone at the C-terminal side of arginine (R) and lysine (K), unless a proline (P) residue follows in the C terminal direction. As the protein is cleaved in a predictable manner, the resulting fragments form a characteristic peptide mass fingerprint (PMF) for the protein when analysed by MALDI MS.¹³⁰ If protein separation is carried out prior to MS, identification is achieved by matching the peptide masses obtained from the mass spectrum against calculated masses for predicted peptides based on sequence data derived from protein databases as described in Section 1.4.3.

^{131, 132} PMF has been used to identify many proteins and is usually carried out using MALDI-TOF mass spectrometry. ¹³³⁻¹³⁸ However, the PMF approach may fail to identify a protein, if protein separation is incomplete leading to peptide fragments from two or more proteins in the MALDI-MS spectrum. When the PMF approach fails the tryptic peptides may be subjected to PSD or tandem mass spectrometry (MS/MS),¹²³ this is particularly useful to ascertain the identity and sequence of proteins in a complex mixture. If the proteolytic digestion is carried out without prior protein separation then peptide separation by LC and subsequent MS/MS (LC/MS/MS) is required. This approach, sometimes referred to as “shotgun” proteomics was first reported by Yates.¹³⁹ One benefit of the method is the avoidance of difficult separation techniques, such as 2D gel electrophoresis. However, the main drawback with this method is that as the complexity of the sample increases the sequence coverage obtained usually decreases due to incomplete separation of all the peptides during chromatography and suppression of low abundant peptides leads to a decreased ability to identify peptides for MS/MS analysis. A further limitation of this bottom up approach is the acquisition of unassignable MS/MS spectra, for example the product ion spectra may lack sufficient ions for an identification.¹⁴⁰ However, the identification of the sequence of a single tryptic peptide can be sufficient to unambiguously identify the protein, providing the peptide is unique to that protein.

Both the traditional proteomic method of 2D-SDS-PAGE followed by digestion and MALDI-TOF analysis of the resulting peptides and the shotgun method have disadvantages. As previously mentioned the gels are difficult to reproduce and are biased towards high abundant proteins. LC MS/MS may result in the loss of valuable peptides through co-elution of peptides. Therefore the aim of the work reported in Chapter 4 includes the development of a proteomic method for the analysis of serum samples.

1.4.3 Protein Identification using Database Searching

Search algorithms such as Sequest (ThermoFinnigan, San Jose, CA, USA) ¹⁴¹ and Mascot (MatrixScience, London) are used widely to identify proteins through interpretation of mass spectrometric and tandem mass spectrometric data of digested peptides. PMF database searches compare the experimental masses for each of the peptides derived from proteolytic digestion of proteins with the theoretical masses for the predicted amino acid sequences. This results in a number of possible proteins that the PMF peptides may be derived from. Scoring algorithms are applied to rank these proteins and indicate the most significant match. A number of parameters are usually entered into the search engine interface together with the peptide mass data set, these include the mass accuracy tolerance and possible peptide modifications. The most important consideration is the mass accuracy tolerance, if this is set too high the number of false positives is increased. However, valid matches may be missed if the window is too narrow.

Proteins may also be identified from MS/MS sequence data using the same databases. Experimentally obtained peptide fragments from proteolytic digests are compared with theoretical b_n and y_n fragment masses for peptides derived from proteins in the database. These values have been calculated for each protein following the specific cleavage rules for each protease. Mascot uses a probability-based scoring system, which ranks the data according to the probability of the match being a random occurrence. The higher the Mascot score the lower the probability of the hit being a coincidence. Identification is only possible if data searched is for a known protein recorded in the database. This is the major problem with database search engines.

1.4.4 Biomarker Identification - Comparative Proteomics

Information provided through analysis of the proteome is important in the identification and screening of diagnostic markers, known as biomarkers, *in vivo* and *in vitro* systems. These markers help to enhance our understanding of disease states, identify and trace tumour progression and aid vaccine development. For example, Alzheimer's disease and down syndrome are associated with changes in the expression of voltage dependent anion-selective channel proteins.¹⁴² In the future biomarker identification may result in the development of tailored therapy for individuals.

Protein expression profiling is the detection and measurement of the expression of a set of proteins in a sample and subsequent comparison with other samples. This area of proteomics is particularly important because alterations in the protein expression may be indicative of disease states. One of the main techniques used for protein expression profiling is to carry out 2D-SDS-PAGE separately on two samples, for example a disease state and healthy control, and compare the resulting gel patterns. 2D electrophoresis is particularly suitable for the analysis of cancerous tissue, since protein expression of normal and diseased tissue from the same patient can be compared. Types of cancer investigated by this technique include lung,¹⁴³ thyroid,¹⁴⁴ colon,¹⁴⁵ kidney¹⁴⁶ and bladder.¹⁴⁷

Problems associated with SDS-PAGE gels were discussed earlier, but the main limitation of the method for expression profiling is the reproducibility of the gels, as the principle of the method is the comparison of the variations in intensity and occurrence of the bands within the two gels. Software packages have been developed for this technique that attempt to compensate for the gel-to-gel variations, with some success. Protein expression profiling by 2D gels is also relatively low throughput and time consuming. To help overcome the problems associated with gel-to-gel variations a technique known as differential in-gel electrophoresis (2D-DIGE) was introduced in 1997.¹⁴⁸ The method involves the labelling of complex protein

mixtures with fluorescent dyes prior to 2D electrophoretic separation. The advantage of this technique is that up to three different samples can be mixed together and separated on a single gel. Each sample is labelled with a different dye, for example a control sample may be labelled with Cy 3 dye and a disease state sample with Cy 5 dye. These cyanine dyes have different excitation and emission wavelengths and are matched for charge and molecular weight ensuring that the same protein in each sample will diffuse to the same location in the 2D gel. Sequential excitation of the two dyes allows visualisation of the gel patterns for the two proteomes. Although 2D-DIGE allows the protein expression ratio to be observed from a single gel and the option to use internal standards to improve inter-gel variation, the method is still considered low throughput and requires relatively large amounts of sample. The technique also needs to be combined with MS in order to identify the biomarker proteins.

1.4.4.1 Biomarker Identification by Mass Spectrometry

Wall *et al*¹⁴⁹ compared traditional two-dimensional gel electrophoresis with MALDI-MS for monitoring protein expression in bacterial cell cultures and concluded that the later provided better resolution and faster analysis. Whole cells have also been employed in protein expression studies, offering the advantage of observing chemical composition under more natural environmental conditions.¹⁵⁰⁻¹⁵² MALDI is also capable of proteome profiling from thin tissue samples. For example Jespersen and co-workers¹⁵³ studied neuropeptides from pituitary tissue sections. Caprioli¹⁵⁴ employed direct tissue analysis to obtain protein profiles of mouse colon tumours and compared these with healthy colon tissue in order to discover twelve biomarkers which can be used in diagnostics.¹⁵⁴ More recently, the prognosis of brain tumour patients has been reported using this direct tissue analysis methodology.¹⁵⁵

MALDI, in conjunction with protein chip technology, also referred to as surface enhanced laser desorption/ionization (SELDI) has been utilised for identification of potential cancer

biomarkers from serum proteins. In the SELDI approach, proteins are immobilised on to a chemically modified affinity surface, for example C₁₆, with non-specifically bound impurities washed away.¹⁵⁶ Many types of cancer samples have been analysed using SELDI-TOF MS, including ovarian,^{157, 158} prostate,¹⁵⁹⁻¹⁶² head and neck,¹⁶³ cervix,¹⁶⁴ lung,¹⁶⁵ pancreas¹⁶⁶ and breast cancer.¹⁶⁷ In most cases biomarkers were identified by a simple comparison of the MALDI spectrum generated from each disease state or control.

SELDI-MS has been combined with artificial neural networks (ANNs) to validate biomarker correlation with disease progression, particularly for cancers.¹⁶⁸ For example, Ball *et al*¹⁶⁹ identified mass spectral peaks which are down regulated as astrocytomas (brain tumour) progresses from low to high grade. ANNs are electronic models which are based on the neural structure of the brain. This form of artificial intelligence can therefore model highly complex systems. The basic structure of an ANNs network is shown in Figure 1.21. The input layer consists of a series of nodes which in the case of mass spectral data are intensities at given *m/z* values. The input layer, hidden and output layers are connected *via* weighted links. At the start of the analysis the weight links consist of random numbers. The ANNs models are developed by iteratively changing the weights in response to the predictive error.

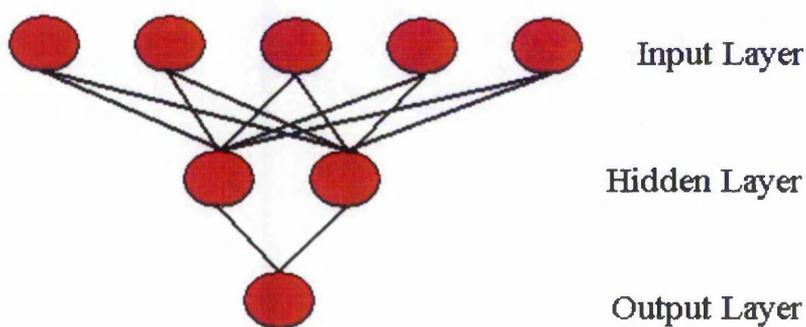


Figure 1.21: A simple neural network diagram.

Predictions are made by using the interconnecting weights between layers to mathematically modify input values, thus learning the relationships between cause and effect. Through parameterisation, trained ANN models (representing biological systems) may be analysed to

determine the importance of biomarkers represented within them and thus, used to identify the optimal subset of predictive biomarkers indicative of a specific class. Other types of bioinformatics have been used to predict clinical outcomes in cancer, for example Petricoin *et al*^{157,170} used a clustering bioinformatics program to obtain a proteomic pattern which discriminated between cancer and non-cancerous samples. Of the 66 non-malignant samples, 63 were correctly recognized as not cancerous and all 50 cancerous samples were correctly identified. This corresponds to a profiling specificity of 95 % and a sensitivity of 100 %.

SELDI has been shown to generate proteomic patterns which can discriminate between controls and disease states. However, the technique highlights variations between states based on differences in MS ion intensities and does not allow identification of specific biomarker proteins. Other criticisms of this technology include the lack of reproducibility. For example, Diamandis¹⁷¹ compared the reports from three prostate cancer studies^{159, 160, 172} and found that even though two of the studies used identical chip surfaces and MS instruments there was little agreement in the peaks selected as diagnostic markers for discriminating between control and prostate cancer serum samples. Diamandis¹⁷¹ also questions the sensitivity of SELDI for dealing with serum biomarkers, since the most well known protein biomarker for prostate cancer, prostate specific antigen (PSA), was not highlighted as being discriminatory in any of the SELDI studies. This suggests that the species observed as potential markers are of much higher concentration than that of traditional cancer biomarkers. Petricoin and Liotta¹⁷³ have suggested the variability between these studies could be due to differences between clinics, such as sample handling procedures. In order to overcome this bias a standard operating procedure is needed which includes sample handling, collection and storage.

Although the variations in populations may be observed using traditional gel based proteomics and SELDI, neither approach allows the identification of the biomarker proteins. Combining 2D gel separation with proteolytic digestion of the selected protein bands and subsequent LC-

MS/MS can lead to successful identification of the targeted proteins. For example, DIGE has been used in combination with MALDI MS for the analysis of protein expression in breast cancer cell lines.¹⁷⁴ Alternatively, LC-MS and isotope tags may be used for quantitative protein expression profiling.¹⁷⁵ The two proteome samples are tagged separately. One sample is labelled with a reagent containing light isotopes (e.g. ^1H , ^{12}C , ^{14}N or ^{16}O) and the other sample is labelled with the same reagent containing heavy isotopes (e.g. ^2H , ^{13}C , ^{15}N or ^{18}O). Labelling may be carried out before, during or following proteolytic digestion with the resultant peptides providing relative quantification by comparison of the ion intensity ratio for the heavy and light isotopically labelled peptides. An example of a multifunctional tagging reagent is an isotope-coded affinity tag (ICAT), which was first reported by Gygi *et al.*¹⁷⁶ There are three constituent groups to the ICAT reagent, Figure 1.22, a thiol reactive iodoacetamide functional group which allows the reagent to bind with free cysteinyl thiols in the proteins. A light or heavy labelled linker which contains either hydrogen or deuterium respectively and biotin, which has high affinity for binding to avidin.

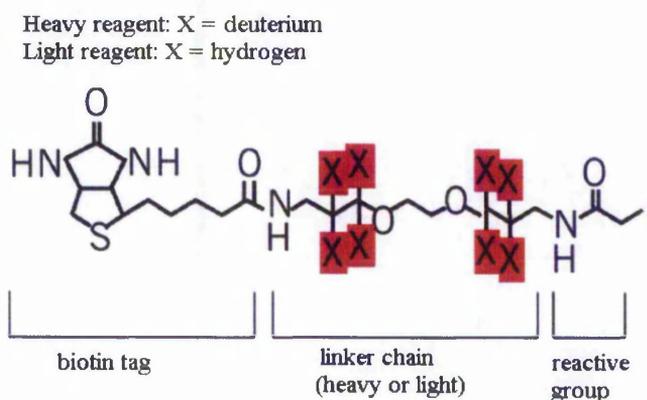


Figure 1.22: Thiol -reactive isotope-coded affinity tag (ICAT) reagents

The two protein mixtures, control and test, are labelled with the ICAT reagents. One is treated with the light ICAT reagent, the other with the heavy ICAT reagent and the two proteomes are then combined and digested. This results in a small number of peptides with tags. The mixture

is passed through an avidin affinity column which binds only the labelled peptides, leading to a less complex sample. The labelled peptides are then eluted and analysed by LC-MS/MS. The two tagged peptides will differ in mass by 8 Da allowing them to be distinguished. The full scan spectrum provides relative quantitative information and the MS/MS data provides peptide sequence information and protein identification. Advantages of the ICAT technique compared with 2D-SDS-PAGE combined with MALDI-TOF for comparative proteomics includes greater confidence in protein assignment because of the MS/MS data. ICAT also offers the ability to detect low abundant proteins which is not possible with 2D gels. A major limitation of the ICAT technique is that not all proteins contain cysteine residues. Only peptides containing cysteines residues are detected and therefore non-cysteine containing peptides will not be retained by the avidin column leading to a loss of valuable information. Other limitations are that the ICAT reagent can interfere with database searching, the reagents are expensive and throughput is lower than strategies that do not involve labelling.

1.5 Research Aims and Outcomes

The objective of the research described in this thesis was the development of novel methods based on AP-MALDI and vacuum MALDI for proteomic analysis and biomarker identification. The thesis is divided into six chapters outlined below:

1. Chapter 2 describes the development and evaluation of a modified AP-MALDI ion source and associated motion control software for peptide analysis by ion trap MS.
2. Chapter 3 describes the novel combination of capillary LC with AP-MALDI-QIT mass spectrometry and its application to the analysis of a protein tryptic digests. This LC/AP-MALDI approach is compared with LC/MALDI-TOF and LC/ESI-qTOF for the identification of tryptic peptides.
3. Chapter 4 describes the development and validation of a novel, bottom-up proteomics methodology for the analysis of low molecular weight serum peptides using tryptic digestion in combination with MALDI-TOF and AP-MALDI/QIT.
4. Chapter 5 describes the application of the methods developed in Chapter 4 to native and tryptic serum peptides from stage IV melanoma patients and control samples, and the use of artificial neural networks (ANNs) and tandem MS to identify biomarkers that distinguish between stage IV melanoma and control serum.
5. Chapter 6 describes conclusions to the thesis and details further work.

1.6 References

- 1) Thompson J.J., Rays of positive Electricity and their Application to Chemical Analysis, Longman Green and Co. Ltd., London (1913)
- 2) Wiley W.C and McLaren J.B., *Rev. Sci. Instrum.*, **26**, 1150-1157 (1955)
- 3) Imrie D.C., Pentney J.M. and Cottrell J.S., *Rapid Commun. Mass Spectrom.*, **9**, 1293-1296 (1995)
- 4) Moniatte M., vanderGoot F.G., Buckley J.T., Pattus F., van Dorsselaer A., *FEBS Lett.*, **384**, 269-272 (1996)
- 5) Onnerfjord P., Nilsson J., Wallman L., Laurell T., Marko-Varga G., *Anal. Chem.*, **70**, 4755-4760 (1998)
- 6) University of Bristol [online] available at: www.chm.bis.ac.uk [accessed Jan 2006]
- 7) Lennon J.J., Walsh K.A., *Prot. Sci.* **6**, 2446-2453, (1997)
- 8) Kaufmann R., Spengler B., Lutzenkirchen F., *Rapid Commun. Mass Spectrom.*, **8**, 781-755, (1994)
- 9) Vestal M.L., Juhasz P and Martin S.A., *Rapid Commun. Mass Spectrom.*, **9**, 1044 (1995)
- 10) Koefeler H.C., Gross M.L., *J. Am. Soc Mass Spectrom.*, **16** , 406-408 (2005)
- 11) Michalet S., Favreau P., Stoecklin R., *Clinical Chem. Lab. Med.*, **41**, 1589-1598 (2003)
- 12) Xu, Chong-Feng; Lu, Yun; Ma, Jinghong; Mohammadi, Moosa; Neubert, Thomas A. *Mol. Cel. Proteomics*, **4**, 809-818 (2005)
- 13) Schevchenko A., Loboda A., Shevchenko A.E., Ens W., Standing K.G., *Anal. Chem.*, **72**, 2132-2141, (2000)
- 14) Loboda A.V., Krutchinsky A.N., Bromirski M., Ens W., Standing K.G., *Rapid Commun. Mass Spectrom.*, **14**, 1047-1057, (2000)

- 15) Martin R et al *Proceedings of the 53rd ASMS Conference on Mass Spectrometry and Allied Topics*, San Antonio, Texas,(2005)
- 16) Bilsborough S., Loftus N., Taniguchi J., Miseki K., *Spectra Analyse.*, **239**, 35-38 (2004)
- 17) Wolfgang P., Steinwedel, H, *Z. Naturforsch.*, **8a**, 448-450 (1953)
- 18) Cristoni S., Bernardi R., *Mass Spectrom. Rev.*, **22**, 369-406 (2003)
- 19) Badman E.R., Cooks G.R., *J Mass Spectrom.*, **35**, 659-671 (2000)
- 20) University of Adelaide, Australia [online] available at:
<http://www.chemistry.adelaide.edu.au/external/soc-rel/content/quadrupo.htm>
[accessed Jan 2006]
- 21) Mathieu E., *J. Math. Pures. Appl.* **13**, 137 (1868)
- 22) March R.E., Hughes R.J., *Quadrupole Storage mass spectrometry.*, Wiley-Interscience Publication, New York, USA (1989)
- 23) March RE., Todd JFJ., *Practical aspects of ion trap mass spectrometry*, Vol 3, Taylor & Francis Ltd CRC Press, London (1995)
- 24) Jonscher K.R., Yates J.R., *Anal. Biochem.*, **244**, 1-15, (1997)
- 25) March R.E., *J. Mass Spectrom.*, **32**, 351-369 (1997)
- 26) March, R.E., Hughes, R.J., (1989) in *Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications* (Winefordneer, J.D and Kolthoff, I.M., Eds.), Vol. 102, Wiley, New York
- 27) Stafford jnr G.C., Kelley P.E., Syka J.E.P., Reynolds W.E., Todd J.F., *Int. J. Mass Spectrom. Ion Processes*, **60**, 85-98 (1984)
- 28) Stafford G.C., Taylor D.M., Bradshaw S.C. and Syka J.E.P., *Proceedings of the 35th ASMS Conference on Mass spectrometry and Allied Topics*, Denver CO (1987)

- 29) Louris J.N., Cooks R.G, Syka J.E.P., Kelley P.E., Stafford G.C., Todd J.F.J., *Anal. Chem.*, **59**, 1677-1685 (1987)
- 30) Barber M., Bordoli R.S., Sedgwick R.D., Tyler A.N., *J. Chem. Soc. Chem Commun.*, **7**, 325-327 (1981)
- 31) Dole M, Hines R.L., Mack R.C., Mobley R.C., Ferguson L.D., Alice M.B., *J. Chem. Phys.*, **49**, 2240-2249 (1968)
- 32) Fenn J.B., Mann M., Meng C.K., Wong S.K., Whitehouse C., *Science*, **246**, 64-71 (1989)
- 33) Hardouin J., Lange C.M., *Curr. Org. Chem.*, **9**, 317-324 (2005)
- 34) Yan, Z., Caldwell, G.W., *Recent Research Developments in Biophysics & Biochemistry*, **3**, 117-139 (2003)
- 35) Meng F., Forbes A.J., Miller L.M., Kelleher N.L., *Mass Spec. Rev.*, **24**, 126-134, (2005)
- 36) Bonner P.L.R., Lill J.R., Hill S.C., Creaser C.S., Rees R.C., *J. Immun. Meth*, **262**, 5-19 (2002)
- 37) Weiskopf A., Vouros P., Harvey D.J., *Anal. Chem.*, **70**, 4441-4447 (1998)
- 38) Schweda E.K.H., Li J.J., Moxon E.R., Richards J.C., *Carb. Res.*, **337**, 409-420 (2002)
- 39) Kebarle P., Peschke M., *Analytic Chimica Acta*, **406**, 11-35 (2000)
- 40) Posthumos M.A., Kistemaker P.G., Meucelaar H.L.C., Ten Voer de Brauw M.C., *Anal. Chem.*, **50**, 985-991 (1978)
- 41) Olthoff J.K., Lys I., Demirev P., Cotter R.J., *Anal. Instrum.*, **16**, 93-115 (1987)
- 42) Wu Z.C. and Fenselau C., *Org. Mass Spectrom.*, **28**, 1034-1038 (1993)
- 43) Creaser C.S and O'Neill K.E., *Anal. Proc.* **30**, 246-248 (1993)

- 44) Tanaka K., Waaki H., Ido H, Akita S and Yoshida T., *Rapid Commun. Mass Spectrom.*, **2**, 151-153 (1988)
- 45) Karas M and Hillenkamp F., *Anal. Chem.*, **60**, 2299-2301 (1988)
- 46) Gluckmann M. and Karas M., *J. Mass Spectrom.*, **345**, 467-477 (1999)
- 47) Driesewerd K., Schurenberg M., Karas M., Hillenkamp F., *Int. J. Mass Spectrom. Ion Proc.*, **141**, 127-148 (1995)
- 48) Chan T., Colburn A.W., Derrick P.J., Gardiner D.J., Bowden M., *Org. Mass Spectrom.*, **27**, 188-194 (1992)
- 49) Wang B., Driesewerd K., Bahr U., Karas M., Hillenkamp F., *J. Am. Soc. Mass Spectrom.* **4**, 393-397 (1993)
- 50) Vertes A., Irinyi G., Gijbels R., *Anal. Chem.*, **65**, 2389-2393 (1993)
- 51) Karabach V., Knochenmuss R., *Rapid Commun. Mass Spectrom.*, **12**, 968-974 (1998)
- 52) Johnson R.E., *Int. J. Mass Spectrom. Ion Proc.* **139**, 25-38 (1994)
- 53) Ehring H and Sundqvist B.U.R., *J. Mass Spectrom.*, **30**, 1303-1310 (1995)
- 54) Ehring H., Karas M., Hillenkamp, F., *Org. Mass spectrum.*, **27**, 427-480 (1992)
- 55) Breuker K., Knochenmuss R., and Zenobi R., *Int. J. Mass Spectrom.*, **184**, 25-38 (1999)
- 56) Fujii T., *Eur. J. Mass Spectrom.*, **2**, 91-114 (1996)
- 57) Mowry C.D., Johnston M.V., *J. Phys. Chem.*, **98**, 1904-1909 (1994)
- 58) Cotter R.J., *Anal. Chim. Acta.*, **195**, 45-49 (1987)
- 59) Liao P.C and Allison J., *J. Mass Spectrom.*, **30**, 408-423 (1995)
- 60) Zhu L., Parr G.R., Fitzgerald M.C., Nelson C.M., Smith L.M., *J. Am. Chem. Soc.*, **117**, 6048-6056 (1995)

- 61) Lehmann E., Knochenmuss R., Zenobi R., *Rapid Commun. Mass Spectrom.*, **11**, 1483-1492 (1997)
- 62) Knochenmuss R., Zenobi R., *Chem. Rev.* **103**, 441-452, (2003)
- 63) Karas M., Bahr U., Strupat K., Hillenkamp F., Tsaibopoulos A., Prarnanik B.N., *Anal. Chem.*, **67**, 675-679 (1995)
- 64) Wolfender, J.L., Chu F.X., Ball H., Wolfender F., Fainzilber M., Baldwin, M.A., Burlingame A.L., *J. Mass Spectrom.*, **34**, 447-454 (1999)
- 65) Laiko V.V., Baldwin M.A., Burlingame A.L., *Anal. Chem.*, **72**, 652-657 (2000)
- 66) McLean J.A., Russell, W.K., Russell D.H., *Anal. Chem.*, **75**, 648-654 (2003)
- 67) Daniel J.M., Ehala S., Friess S.D., Zenobi R., *Analyst*, **129**, 574-578 (2004)
- 68) Krutchinsky A.N., Loboda A.V., Spicer V.L., Dworschak R., Ens W., Standing K.G., *Rapid Commun. Mass Spectrom.*, **12**, 508-518 (1998)
- 69) Bramwell C.J., Creaser C.S., Reynolds J.C., Dennis R., *Int. J. Ion Mob. Spectrom.*, **5**, 87-90 (2002)
- 70) Steiner W.E., Clowers B.H., English W.A., Hill H.H., Jr. *Rapid Commun. Mass Spectrom.*, **18**, 882-888 (2004)
- 71) Tong H., Sze N., Thomson B., Nacson S., Pawliszyn J., *Analyst*, **127**, 1207-1210 (2002)
- 72) Laiko V.V., Moyer S.C., Cotter R.J., *Anal. Chem.*, **72**, 5239-5243 (2000)
- 73) Keough T., Lacey M.P., Strife, R.J., *Rapid Commun. Mass Spectrom.*, **15**, 2227-2239 (2001)
- 74) Creaser C.S., Reynolds J.C., Harvey D., *Rapid Commun. Mass Spectrom.*, **16**, 176-184 (2002)
- 75) Miller C.A., Donghui Y., Perkins P.D., *Rapid Commun. Mass Spectrom.*, **17**, 860-868 (2003)

- 76) Danell R.M., Glish G.L., An Atmospheric Pressure MALDI Probe for use with ESI Source Interfaces, In: *Proceedings of the 48th Annual ASMS Conference on Mass Spectrometry and Allied Topics*, Long Beach, CA , **2000**.
- 77) Danell RM. Glish, GL., Heating to Maximize AP-MALDI Performance: Evidence for Desolvation, In: *Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics*, Chicago, IL, **2001**.
- 78) Doroshenko V.M., Laiko V.V., Taranenko N.I., Berkout V.D., Lee H.S., *Int. J. Mass Spectrom.*, **221**, 39-58 (2002)
- 79) Galicia M.C., Vertes A., Callahan J.H., *Anal. Chem.*, **74**, 1891-1895 (2002)
- 80) Callahan J.H., Galicia,M.C., Vertes A., Atmospheric Pressure Matrix-Assisted Laser Desorption Ionization (AP-MALDI) With an Ion Trap Mass Spectrometer, In: *Proceedings of the 48th ASMS conference on mass spectrometry and allied topics*, ASMS, Long Beach, CA, **2000**.
- 81) Callahan J.H., Galicia M.C., Vertes A.J., *Appl. Surface Sci.*, **130**, 197-198 (2002)
- 82) Tan P.V., Laiko V.V., Doroshenko V.M., *Anal. Chem.*, **76**, 2462-1742 (2004)
- 83) Laiko V.V., Taranenko N.I., Berkout V.D., Musselmann B.D., Doroshenko V.M., *Rapid Commun. Mass Spectrom.*, **16**, 1737-1742 (2002)
- 84) Pihlainen K., Grigoras K., Franssila S., Sami R, Kotiaho T., Kostianen R., *J. Mass Spectrom.*, **40**, 539-545 (2005)
- 85) Ding L., Sudakov M., Brancia F.L., Giles R., Kumashiro S.J., *Mass Spectrom.*, **39**, 471-484 (2004)
- 86) Coon J.J., McHale K.J., Harrison W.W., *Rapid Commun. Mass Spectrom.*, **16**, 681-685 (2002)
- 87) Coon J.J., Harrison W.W., *Anal. Chem.*, **74**, 5600-5605 (2002)

- 88) Kellersberger K.A., Tan P.V., Laiko V.V., Doroshenko V.M., Fabris D., *Anal. Chem.*, **76**, 3930-3934 (2004)
- 89) Zhang J., LaMotte L., Dodds E.D., Lebrilla C.B., *Anal. Chem.* **77**, 4429-4438 (2005)
- 90) Kellersberg K.A., Yu E.T., Merenbloom S.I., Fabris D., *J. Am. Soc. Mass Spectrom.*, **16**, 199-207 (2005)
- 91) Buchanan M.V., Hettich R.L., *Anal. Chem.*, **65**, 245A-259A (1993)
- 92) O'Connor P.B., Mirgorodskaya E., Costello, C.E., *J. Am. Soc. Mass Spectrom.* **13**, 402-407 (2002)
- 93) O'Connor P.B., Costello C.E., *Rapid Commun. Mass Spectrom.*, **15**, 1862-1868 (2001)
- 94) Turney K., Harrison W.W., *Rapid Commun. Mass Spectrom.*, **18**, 629-639 (2004)
- 95) Mehl J.T., Cummings J.J., Rohde E., Yates N., *Rapid Commun. Mass Spectrom.*, **7**, 1600-1610 (2003)
- 96) Marzilli L.M., Moyer S.C., Cotter R.J., Analysis of Posttranslational Modifications by Atmospheric Pressure MALDI/Ion Trap Mass Spectrometry, In: *Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics*, Chicago, IL, May **2001**.
- 97) Moyer S., Cotter R., Woods A., *J. Am. Soc. Mass Spectrom.*, **13**, 274-283 (2002)
- 98) Moyer S.C., Von Seggern C.E., Cotter R.J., *J. Am. Soc. Mass Spectrom.*, **14**, 581-592 (2003)
- 99) Moyer S.C., Marzilli L.A., Woods A.S., Laiko V.V., Doroshenko V.M., Cotter R.J., *Int. J. Mass Spectrom.*, **226**, 133-150 (2003)
- 100) Creaser C.S., Reynolds J.C., Hotelling A.J., Nichols W.F., Owens K.G., *Eur. J. Mass Spectrom.*, **9**, 33-44 (2003)

- 101) Wei H., Nolkranz K., Powell D.H., Woods J.H., Ko M-C., Kennedy R.T., *Rapid Commun. Mass Spectrom.*, **18**, 1193-1200 (2004)
- 102) Von Seggern C.E., Moyer S.C., Cotter R.J., *Anal. Chem.*, **75**, 3212-3218 (2003)
- 103) Von Seggern C.E., Cotter R.J., *J. Am. Soc. Mass Spectrom.*, **14**, 1158-1165 (2003)
- 104) Von Seggern C.E., Zarek P.E., Cotter R.J., *Anal. Chem.*, **75**, 6523-6530 (2003)
- 105) Laiko V.V., Taranenko N., Berkout V., Yakshin M., Prasas C., Lee H., Doroshenko V.M., *J. Am. Soc. Mass Spectrom.*, **13**, 354-361 (2002)
- 106) Von Seggern C.E., Gardner B.D., Cotter R.J., *Anal. Chem.*, **76**, 5887-5895
- 107) Moyer S.C., Woods A.S., Cotter R.J., Fragmentation of Phosphotyrosine Containing Peptides by Atmospheric Pressure MALDI Ion Trap Mass Spectrometry, In: *Proceedings from 2nd Mass Spectrometry Conference Applied to Biological Warfare Agents*, Warwick, UK, **2002**.
- 108) Dale M.J., Knochenmuss R., Zenobi R., *Anal. Chem.*, **68**, 3321-3329 (1996)
- 109) Doroshenko V.M., Cotter R.J., *J. Mass Spectrom.*, **31**, 602-615 (1997)
- 110) Doroshenko V.M., Cornish T.J., Cotter R.J., *Rapid Commun. Mass Spectrom.*, **6**, 753-757 (1992)
- 111) Gabelica V., Schulz E., Karas M., *J. Mass Spectrom.*, **39**, 579-593 (2004)
- 112) Cotter R.J., *Anal. Chem.*, **64**, 1027A (1992)
- 113) Roepstorff P., *Trends Anal. Chem.*, **12**, 413-421 (1993)
- 114) Eckart K., *Mass Spectrom. Rev.* **13**, 23-55 (1994)
- 115) Feistner G.J., Faull K.F., Barofsky, D.F., Roepstorff P., *J. Mass spectrom.*, **30**, 519-530 (1995)
- 116) Burlingame A.L., Boyd R.K., Gaskell S.J., *Anal. Chem.*, **68**, 599R-651R (1996)
- 117) Roepstorff P., Fohlmann J., *Biomed Mass Spectrom.*, **11**, 601-604 (1984)

- 118) Eng J.K., McCormack A.L., Yates J.R. III, *J. Am. Soc. Mass Spectrom.*, **5**, 976-989 (1994)
- 119) Hunt D.F., Henderson R.A., Shabanowitz J., Sakaguchi K., Michel H., Sevilir N., Cox A.L., Appell e., Engelhard V.H., *Science*, **255**, 1261-1263 (1992)
- 120) Cardenas M.S., van der Heeft E., de Jong A.P.J.M., *Rapid Commun. Mass Spectrom.*, **11**, 1271-1278 (1997)
- 121) Shevchenko A., Chernushevich I., Ens W., Standing K.G., Thomson B., Wilm M., Mann M., *Rapid Commun. Mass Spectrom.*, **11**, 1015-1024 (1997)
- 122) Vath J.E and Biemann, K., *Int J. Mass spectrum. Ion Processes*, **100**, 287-299 (1990)
- 123) Hunt D.F., Yates, J.R., Shabanowitz J., Winston, S., Hauer C.H., *Proc. Natl. Acad. Sci. USA*, **83**, 6233-6237 (1986)
- 124) Anderson NL., Anderson NG., *Mol. Cell. Proteomics*, **1**, 845-867 (2002)
- 125) Meng, F., Cargile, B.J., Miller L.M., Forbes A.J., Johnson J.R., Kelleher N.L., *Nature Biotechnol.*, **19**, 952-957 (2001)
- 126) Reid G.E., Stephenson, J.L., McLuckey S.A., *Anal. Chem.*, **74**, 577-583 (2002)
- 127) Kelleher N.L., Zubarev R.A., Bush, K., Furie B.C., McLafferty F.W., Walsh C.T., *Anal. Chem.*, **71**, 4250-4253 (1999)
- 128) Loo J.A., Edmonds C.G., Smith R.D., *Science*, **248**, 210-204 (1990)
- 129) Chait B.T., Wang R., Beavis R.C., Kent S.B., *Science*, **263**, 89-92 (1993)
- 130) Pappin D.J.C., Bleasby A.J., *Current Biol.*, **3**, 327-332, (1993)
- 131) James P., Quadroni M., Carafoli E., Gonnet G., *Biochem. Biophys. Res. Commun.*, **195**, 58-64, (1993)
- 132) Mann M., Højrup P., and Roepstorgg P., *Biol. Mass Spectrom.*, **22**, 338-345 (1993)

- 133) Zhang Z-T., Geng Y-P., Zhou L., Lai B-C., Si L-S., Wang Y-L., *World J. Gastroenterol.*, **11**, 4679-4684 (2005)
- 134) Yoon S-W., Kim T-Y., Sung M-H., Kim C.J., Poo H., *Proteomics*, **5**, 1987-1995 (2005)
- 135) Tien C-C., Chen J-B., Wang C-C., Lee W-C., *Enzyme Microb. Technol.*, **33** (4), 488-491 (2003)
- 136) Flensburg J., Belew M., *J. Chromatography A.*, **1009**, 111-117, (2003)
- 137) Eklund P., Andersson H.O., Kamali-Moghaddam M., Sundstrom L., Flensburg J., *J. Chromatography A.*, **1009**, 179-188, (2003)
- 138) Ishino M., Aoto H., Sasaki H., *Tumour Res.*, **37**, 23-31 (2002)
- 139) Link A.J., Eng J., Schieltz D.M., Cormack E., Mize G.J., Morris D., Garvick B.M., Yates JR III., *Nat. Biotechnol.*, **17**, 676-682 (1999)
- 140) Simpson R.J., Connelly L.M., Eddes J.S., Pereira J.J., Moritz R.L., Reid G.E., *Electrophoresis*, **21**, 1707-1711 (2000)
- 141) Qin J, Herring C.J. and Zhang X., *Rapid Commun. Mass Spectrom.*, **12**, 209-216 (1998)
- 142) Yoo B.C., Fountoulakis M., Cairns N. and Lubec G., *Electrophoresis*, **22**, 172-179 (2001)
- 143) Okuzawa K., Franzen B., Lindholm J., Linder S., Hirano T., Bergman T., Ebihara Y., Kato H., Auer G., *Electrophoresis.*, **15**, 382 (1994)
- 144) Lin J.D., Huang C.C., Weng H.F., Chen S.C., Jeng L.B., *J. Chromatogr. B. Biomed. Appl.*, **667**, 153-160 (1995)
- 145) Ward LD., Hong J., Whitehead RH., Simpson RJ., *Electrophoresis.*, **11**, 883-891 (1990)

- 146) Sarto C., Marocchi A., Sanchez JC., Giannone D., Frutiger S., Golaz O., Wilkins M.R., Doro G., Cappellano F., Hughes G., Hochstrasser DF., Mocarelli P., *Electrophoresis*, **18**, 599-604 (1997)
- 147) Celis J.E., Ostergaard M., Basse B., Celis A., Lauridsen J., Ratz G.P., Anderson L., Hein B., Wolf H., Orntoft T.F., Rasmussen H.H., *Cancer Res.*, **56**, 4782-4790 (1996)
- 148) Unlu M., Morgan M.E., Minden JS., *Electrophoresis* **18**, 2071-2077 (1997)
- 149) Wall D.B., Lubman D.M., and Flynn S.J., *Anal. Chem.*, **71**, 3894-3900 (1999)
- 150) Easterling M.L., Colangelo C.M., Scott R.A. and Amster I., *Anal. Chem.*, **70**, 2704-2709 (1998)
- 151) Arnold R.J., Karty J.A., Ellington A.D. and Reilly J.P., *Anal. Chem.*, **71**, 1990-1996 (1999)
- 152) Demirev P.A., Ho Y.P., Ryzhov V. and Fenselau C., *Anal. Chem.*, **71**, 2732-2738 (1999)
- 153) Jeperen S., Chaurand P., Van Srien F.J.C., Spengler B and Van der Greef J., *Anal. Chem.*, **71**, 660-666 (1999)
- 154) Reyzer M.L. and Caprioli R.M., *Mod. Drug Disc.*, 45-48 (2002)
- 155) Schwartz S.A., Weil R.J., Thompson R.C., Shyr Y., Moore J.H., Toms S.A., Johnson M.D., Caprioli R.M., *Cancer Res.* **65**, 7664-7681, (2005)
- 156) Merchant M., Weinberger S.R., *Electrophoresis*, **21**, 1164-1174 (2000)
- 157) Petricoin III E.F., Ardekani, A.M., Hitt B.A., Levine P.J., Fusarco V.A., Steinberg S.M., Mills G.B., Simone C., Fishman D.A., Kohn E.C., Liotta L.A., *Lancet* **359**, 572-577 (2002)
- 158) Rai A.J., Zhang, Z., Rosenweig, J., Shih Ie M., Pham T., Fung E.T., Sokoll L.J., Chan D.W., *Arch. Pathol. Lab. Med.* **126**, 1518-1526 (2002)

- 159) Adam B.L., Qu Y., Davis J.W., Ward M.D., Clements M.A., Cazares L.H., Semmes O.J., Schellhammer, P.F., Yasui Y., Feng, Z., Wright Jr G.L., *Cancer Res.* **62**, 3609-3614 (2002)
- 160) Petricoin III E.F., Ornstein D.K., Pawletez C.P., Ardekani A., Hackett P.S., Hitt B.A., Velasco A., Trucco C., Wiegand L., Wood K., Simone C.B., Levine P.J., Linehan W.M., Emmert-Buck M.R., Steinberg S.M., Kohn E.C., Liotta L.A., *J. Natl. Cancer Inst.* **94**, 1576-1578 (2002)
- 161) Banez L.L., Prasanna P., Sun L., Ali A., Zou Z., Adam B.L., McLeod D.G., Moul J.W., Srivastava S., *J. Urol.* **170**, 442-446 (2003)
- 162) Li J., White N., Zhang Z., Rosenzweig J., Mangold L.A., Partin A.W., Chan D.W., *J. Urol.* **171**, 1782-1787 (2004)
- 163) Wadsworth J.T., Somers K.D., Cazares L.H., Malik G., Adam B.L., Stack Jr B.C., Wright Jr G.L., Semmes O.J., *Clin. Cancer Res.*, **10**, 1625-1632 (2004)
- 164) Wong Y.F., Cheung T.H., Lo K.W., Wang, V.W., Chan C.S., Ng T.B., Chung T.K., Mok S.C., *Cancer Lett.* **211**, 227-234 (2004)
- 165) Xiao X., Liu D., Tang Y., Guo F., Xia L., Liu J., He D., *Dis. Markers* **19**, 33-39 (2003)
- 166) Koopmann J., Zhang Z., White N., Rosenzweig J., Fedarko N., Jagannath S., Canto M.I., Yeo C.J., Chan D.W., Goggins M., *Clin. Cancer Res.*, **10**, 860-868 (2004)
- 167) Lancashire L.J., Mian S., Ellis I.O., Rees R.C., Ball G., *Proteomics* **2**, 15-29 (2005)
- 168) Mian S., Ball G., Hombuckle J., Holding F., Carmichael J., Ellis I.m Ali S., Li G., McCardle S., Creaser C., Rees RC., *Proteomics* **3**, 1725-1737 (2003)
- 169) Ball G., Mian S., Holding F., Alliborne RO., Lowe J., Ali S., Li G., McCardle S., Ellis IO., Creaser C., Rees RC., *Bioinformatics* **18**, 395-404 (2002)

- 170) Petricoin EF., Zoon KC., Kohn EC., Barrett JC., Liotta LA., *Nat. Rev Drug Discov.* **9**, 683-695 (2002)
- 171) Diamandis EP., *Clin. Chem.*, **49**, 1272-1843 (2003)
- 172) Qu Y., Adam, B-L., Yasui, Y., Ward, M.D., Cazares L.H., Schellhammer P.F., Feng Z., Semmes O.J., Wright G/L., *Clin. Chem.*, **48**, 1835 (2002)
- 173) Petricoin III E., Liotta LA., *Clin. Chem.*, **49**, 1276-1278 (2003)
- 174) Gharbi S., Gaffney P., Yang A., Zvelebil M.J., Cramer R., Waterfield MD., Timms J.F., *Mol. Cell Proteomics.*, **1**, 91-98 (2002)
- 175) Brancia F.L., *Cur. Anal. Chem.*, **2**, 1-7, (2006)
- 176) Gygi S.P., Rist B., Gerber S.A., Turecek F., Gelb M.H., Aebersold R., *Nature Biotechnol.*, **17**, 994-999 (1999)

Chapter 2

Development and Modification of A Commercial Ionization Source for AP-MALDI

2.1 Introduction

This chapter describes the development and modification of an atmospheric pressure ionization source for matrix-assisted laser desorption ionization (AP-MALDI). The AP-MALDI studies reported in this thesis were carried out using a quadrupole ion trap mass spectrometer (ThermoFinnigan LCQ classic) fitted with an early version of a commercial AP-MALDI ionization source (Mass Technologies, Burtonsville, MD, USA). The commercial AP-MALDI ion source was supplied by the manufacturer without a laser and control software. Therefore initial stages of the work were concerned with the implementation of a versatile AP-MALDI ion source based on an existing in-house nitrogen laser.

Previous studies have reported the use of an in-house AP-MALDI source consisting of a stainless steel target interfaced with a quadrupole ion trap.¹ These studies demonstrated the capability of AP-MALDI-QIT combined with a high powered nitrogen laser ($\lambda = 337$ nm, 1.4 mJ per 600 ps pulse) for the analysis of polymers² and oligosaccharides.³ However, the main limitation with this previous in-house AP-MALDI source was that only one sample could be applied to the target plate at a time. In comparison the commercially available source used in this thesis has a 96 well target plate allowing rapid analysis of a large number of samples.

The work carried out in this thesis involved the modification of a commercial AP-MALDI source to allow operation with a high powered nitrogen laser ($\lambda = 337$ nm, 1.4 mJ per 600 ps pulse) (Laser Photonics, Orlando, FL) and independent PC control of the sample stage motion. The development of the hardware and the software program are discussed below.

Evaluation of the AP-MALDI ion source was carried out on a test set of compounds, including the peptides gp 70 and neurotensin. The performance of the source was also examined using an ubiquitin digest and optimisation of sample preparation methods were carried out.

2.2 Hardware Configuration

A PC-based motion control system has four main components namely; a controller, an amplifier, motors and a feedback system, as shown in Figure 2.1. The controller board inserted into the control PC used in this work was a PCI-7342 (National Instruments, UK).

The controller generates control signals and receives feedback from the motors, such as activation of limits, encoders and home reference switches. This information is passed through an external motor drive, in this case an MC-4SA (National Aperture, New Hampshire, USA), which provides the current and voltage required to power the motors.

The motors are located inside the AP-MALDI source (Figure 2.1). The motors move the stage which holds the sample target plate to the required location. This is usually so that the focal point of the laser is directed at the sample to be analysed, shown in Figure 2.2. High voltage (HV) is applied to the target plate (typically 2.-2.7 kV) to assist the transport of the produced ions towards the inlet of the mass spectrometer. Sample deposited onto the target plate is irradiated with UV light from the nitrogen laser which is coupled to the source *via* a fibre optic. The laser beam is focused by the quartz lens and directed onto the sample by the mirror. A CCD camera coupled to a monitor allows the target plate motion to be monitored and the sample ablation process to be observed.

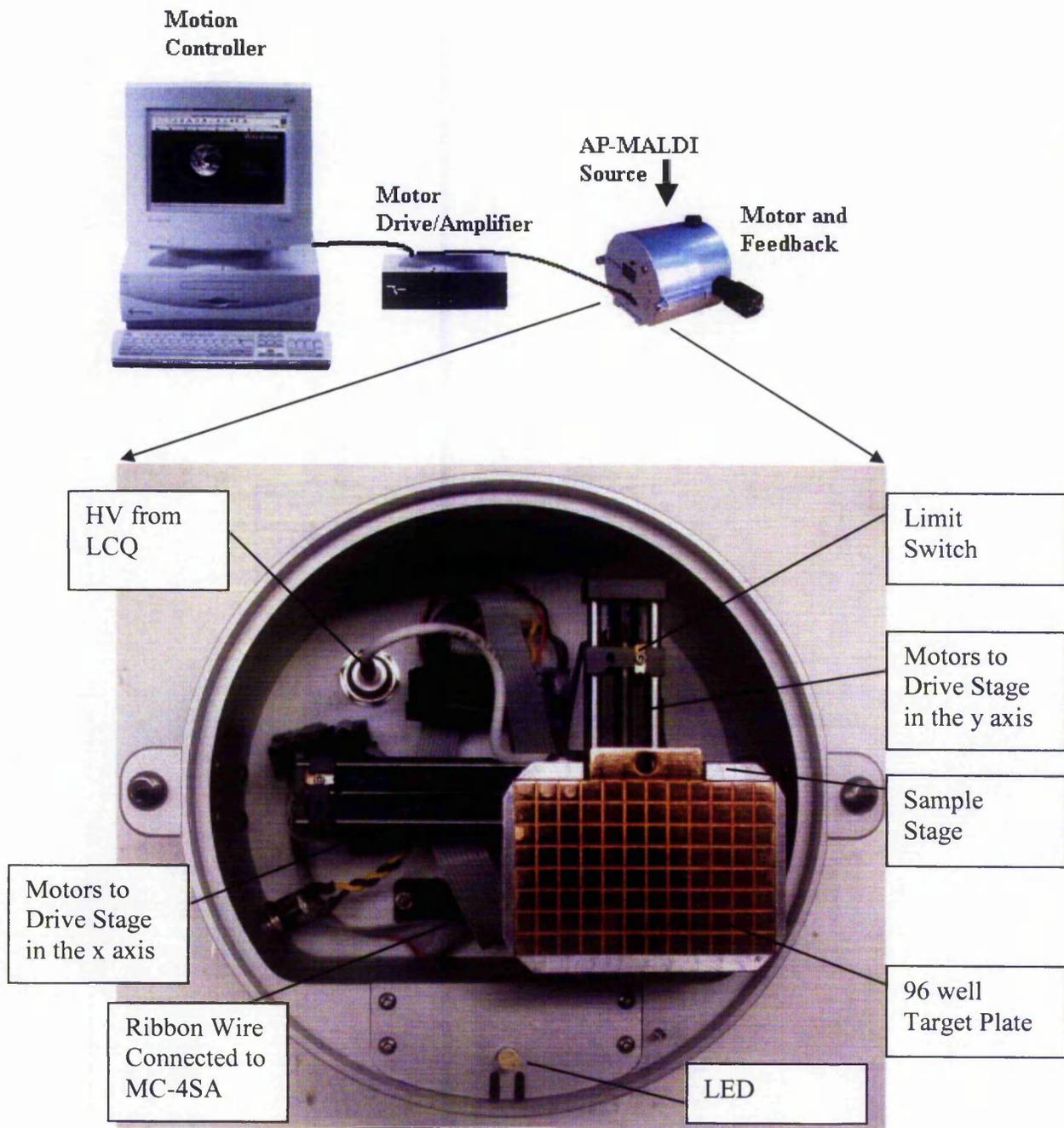


Figure 2.1 – AP-MALDI motion control set-up

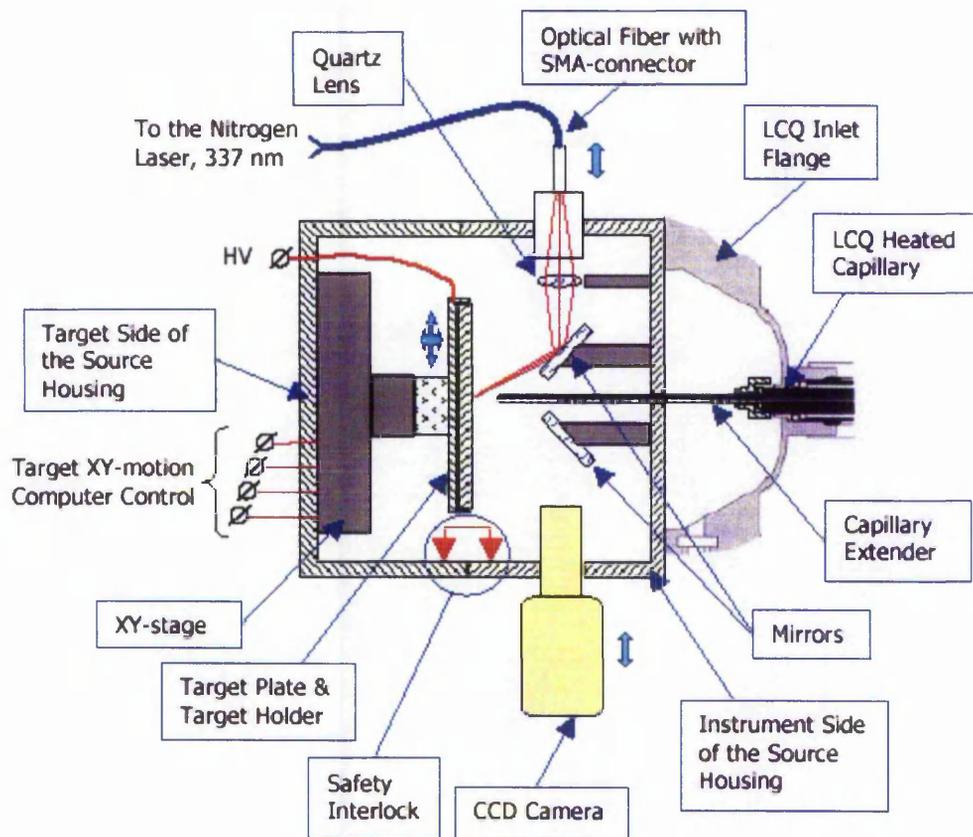


Figure 2.2: Schematic of AP-MALDI source.

2.3 LabVIEW

The control software was developed using LabVIEW 6.1, which is a graphical programming language known as G programming. This means that icons are utilised instead of text to create the program and the flow of the data determines the execution. LabVIEW programs are known as virtual instruments (VIs) since their appearance and operation are like physical instruments. There are three main components to a VI, firstly the front panel, which is the user interface and contains all the interactive inputs and displays. The block diagram contains the graphical representation of the programming code. Every control or indicator on the front panel has a corresponding terminal on the block diagram. Each terminal is connected via wires to produce a complete wiring diagram for the application. The final component to a VI is the icon and connector pane this identifies the VI.

2.4 Software Development

2.4.1 AP-MALDI Plate Coordinates

A virtual AP-MALDI target plate was created on the front panel through a series of Booleans, shown in Figure 2.3a. These are logic functions where the only values allowed are true or false. These two states are often referred to as on or off respectively. Each Boolean on the front panel (labelled 1-96) represented a sample position on the target plate, Figure 2.3b.

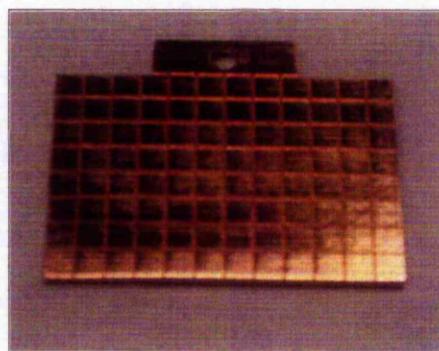
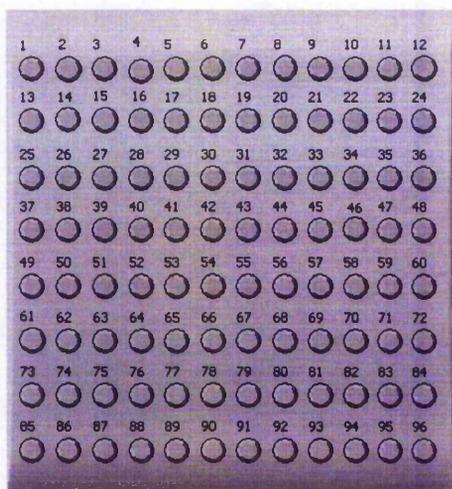


Figure 2.3 a) Virtual target plate

b) AP-MALDI target plate

The boolean true or false value was converted into a digital output (i.e. 1 or 0) through the use of a converter (Boolean to 0, 1). Each boolean was assigned x, y values corresponding with positions on the MALDI target plate. These were entered as constants on the wiring diagram, Figure 2.4. The inputs corresponding to the x coordinates were wired into 8 compound arithmetics (summing devices) with 12 booleans assigned to each block. This function performs arithmetic on one or more numeric, array, cluster, or boolean inputs, in this case the mode of operation was addition. The inputs were split into 8 blocks for ease of wiring. The outputs for the eight summing devices were then fed into another compound arithmetic. For example when the control on the front panel corresponding to position $x = 2$ and $y = 1$ was selected the input for the x compound arithmetic was 2. As no other target

positions were selected the other inputs to the compound arithmetic were all 0, hence the output from the summing device was 2. The output from the first compound arithmetic was entered into the second device. Again, as no other target positions were selected the output was 2. These devices were needed so that each boolean utilised the same output, in this case a digital display on the front panel which showed the coordinates. The y inputs were wired in exactly the same manner. A time function was included in order to increase the motor position by the required amount and the total value entered on a digital display on the user interface.

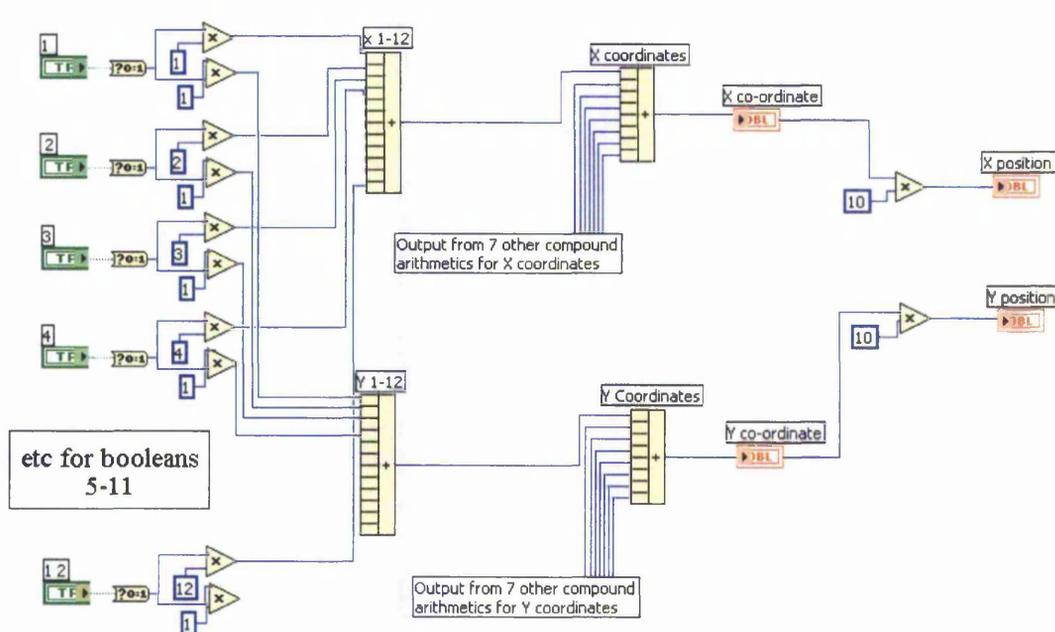


Figure 2.4: Example section of code for controlling the AP-MALDI target position

2.4.2 Development of Basic Motion Control

A two-axis simple motion control program was developed to control the motors.

The graphical programming code is illustrated in Figure 2.5 and the user interface shown in Figure 2.6. Target positions (x and y) were entered in the front panel and the stage moved to the corresponding positions. Combing the simple motion program with the plate control program, allowed the target position to be entered *via* the boolean controls as described in Section 2.4.1. The target position was fed from the plate coordinates to the multi-start command which authorised and controlled the movement of the motors in the

stage, this resulted in the movement of the AP-MALDI stage to the corresponding sampling site. This was usually the focal point of the laser, which was located close to the end of the ion trap heated capillary extension, so that desorbed ions could be extracted into the mass spectrometer, Figure 2.2.

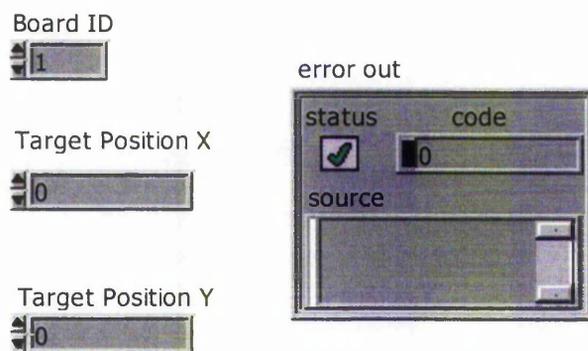


Figure 2.5: User Interface for simple motion program

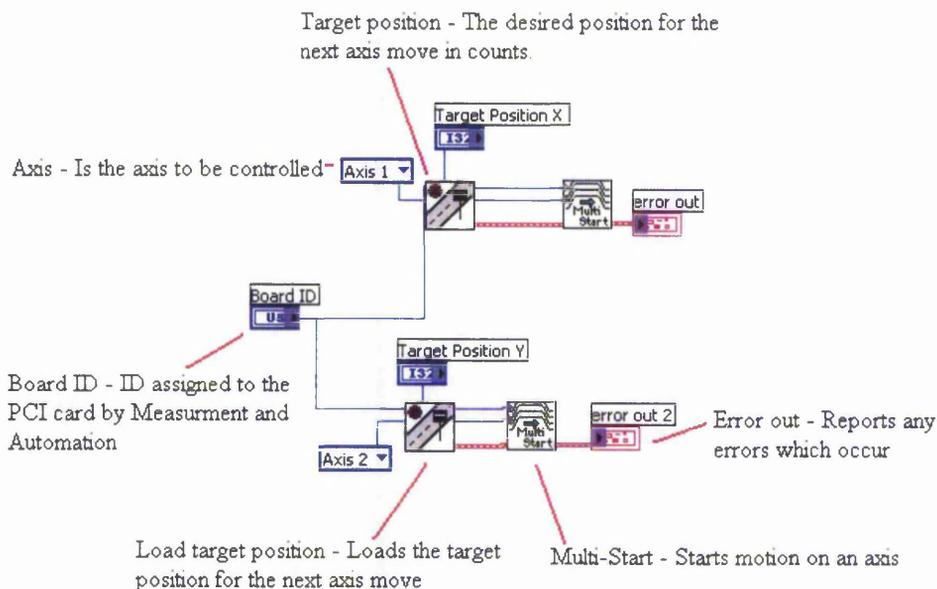


Figure 2.6: Program code for simple motion program

2.4.3 Motion Control Development

A more sophisticated motion control element was developed to control the movement of the AP-MALDI stage. This included the incorporation of velocity controls to adjust the speed at which the stage moved. The motion control software was developed by utilising part of the two-axis contouring spiral VI example supplied by national instruments as part of the LabVIEW package. The graphical programming code was adapted so that the complete motion module, Figure 2.7, included a configure vector space VI, which was used to group axes into a vector space. In this particular instance, the vector space defined an x, y (2D) coordinate space. The physical axes were associated with the logical x, y, and axes to control motion in the vector space. The default values for x and y axes were 1 and 2 respectively. A load velocity VI was included to specify the maximum trajectory velocity for the vector space. The upper range limit allowed for this application was 10000 counts/s due to the physical limitations of the encoder inputs and outputs. The sign of the number of counts loaded specified the direction of travel. A load acceleration/ deceleration VI was included in order to control the maximum rate of acceleration and deceleration for the individual axis and the vector space. Acceleration is defined as how quickly the axis comes up to speed and is limited in order to prevent stress on the motor. The typical operating value was 200 motor revolutions/s/s (RPS/s). The x, y target positions entered from the plate control program were entered into the load vector space position VI, which took the values and loaded them for the next vector space move. The multi-start VI simply started motion on a single axis, a single vector space, multiple axes, or multiple vector spaces using the most recently entered values of velocity, acceleration and target positions. The positions of the axes were constantly monitored by the read vector space position VI. This returned the instantaneous positions of the axes in the vector space and, since servo motors were utilised returned the feedback in counts. The other monitoring function built into the program was the check move complete status VI, which ensured that the move had been finalised and reported whether or not the specified move or moves were complete.

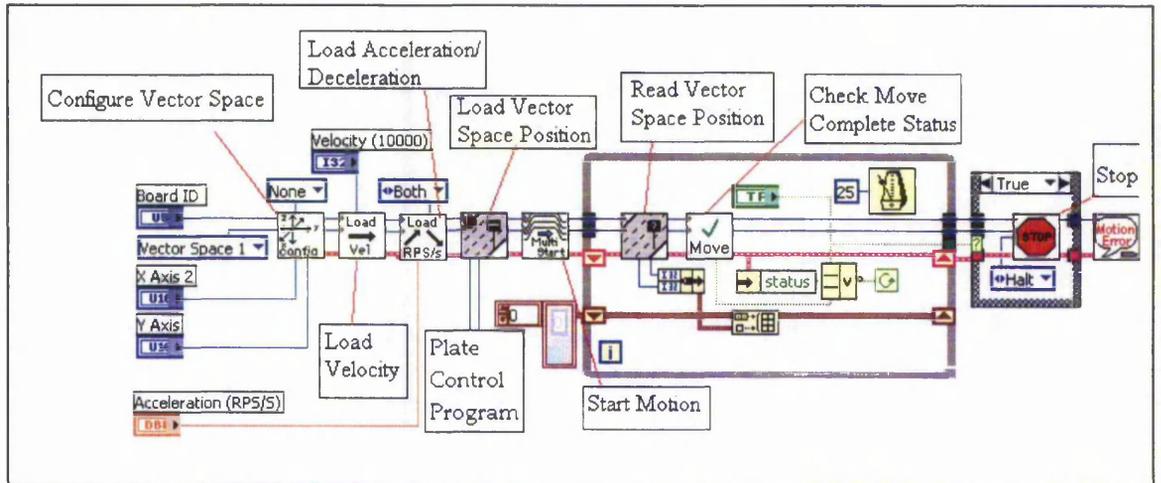


Figure 2.7 – Code for motion control element

The front panel was updated to allow user-friendly control of the AP-MALDI stage *via* an LED display to indicate the target position, an over-ride stop facility, velocity and acceleration controls. The programming code incorporating the new developments is shown in Figure 2.8. A manual function was also added, where the target position can be entered and the mode of operation can be selected, for example, the relative or absolute coordinate in respect of the target position. The ability to move quickly between adjacent samples was included through the shift function, which moves the stage from one sample to the next. A fine tune function was also incorporated into the program, as the ability to control the position of the AP-MALDI target and sample spot precisely in relation to the fibre optic was essential in order to obtain maximum signal and to be able to move to sweet spots within the sample. The fine tune control allowed the sample position to be adjusted in both x and y directions.

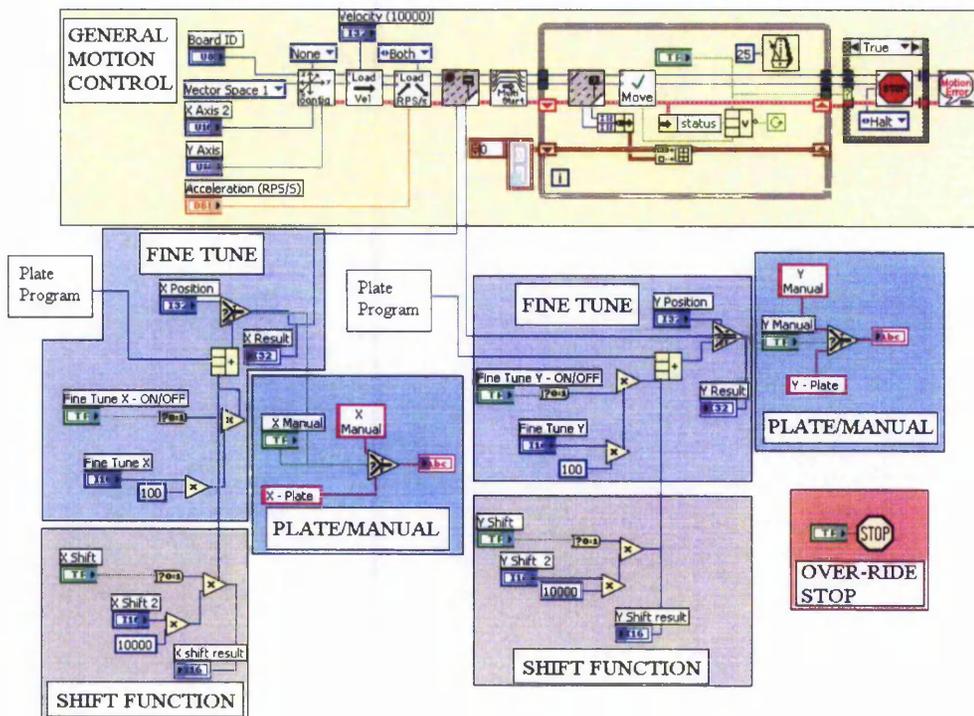


Figure 2.8: AP-MALDI wiring diagram showing shift functions and fine tune additions

Figures 2.9 and 2.10 show further modifications which were made to allow more fluent movement within a sample spot by using slide controls on the front panel for the fine tune operation. Displays were created to enable the current target position to be shown, x positions were assigned 1-12 and y positions as A-H. The LED plate display was adapted in order to indicate samples completed. For example Figure 2.9 shows that the first four samples on the top row have been analysed. The position corresponding to C1 on the AP-MALDI target plate is currently aligned with the fibre optic allowing the sample to be desorbed from the target plate and therefore analysed by the mass spectrometer.

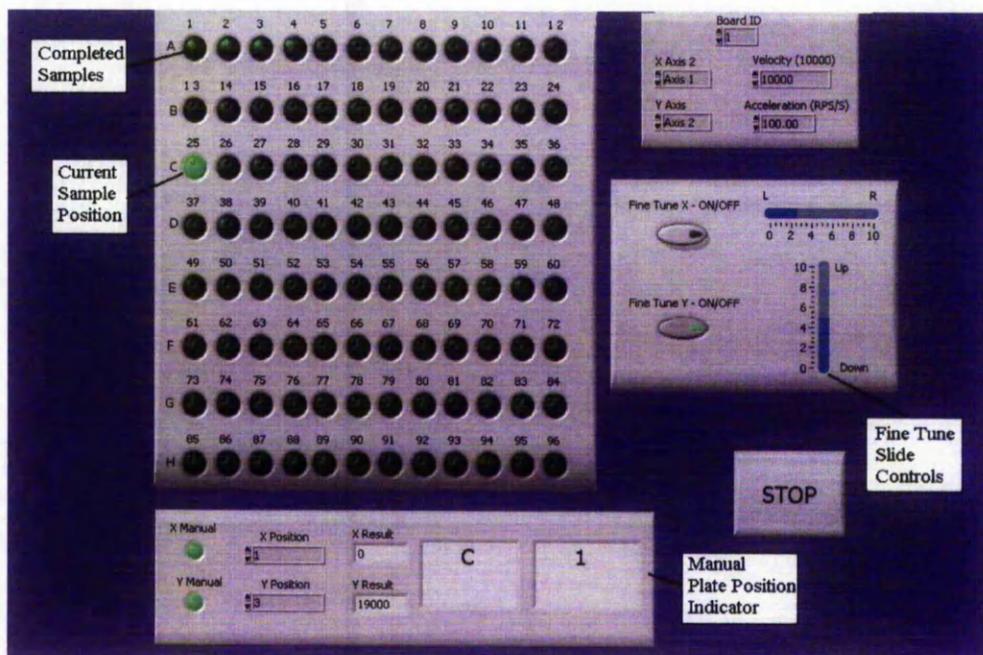


Figure 2.9: Front panel of the AP-MALDI motion control software used for subsequent experiments

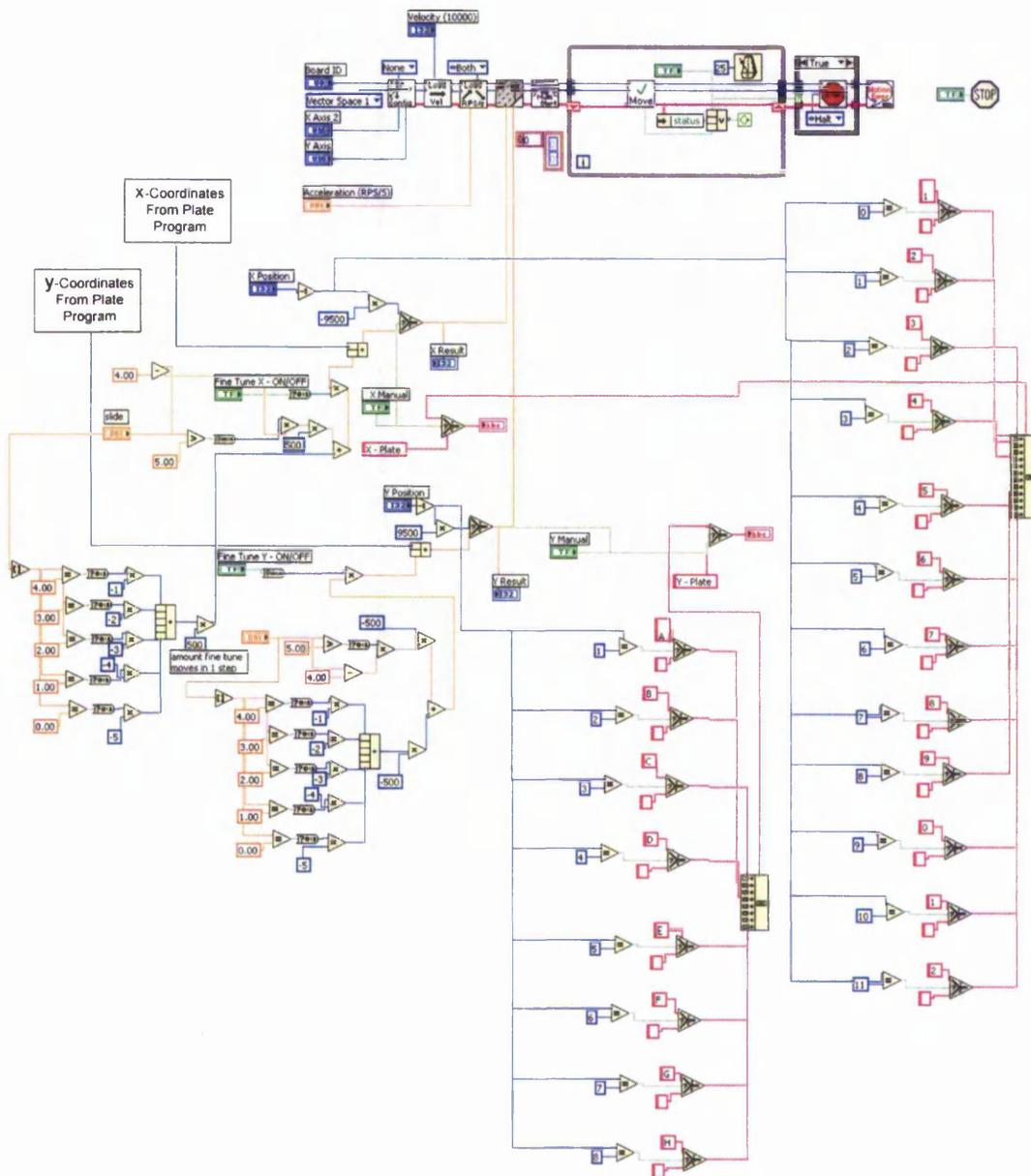


Figure 2.10: Programming code for the AP-MALDI motion control software used for subsequent experiments

2.4.4 Joystick Control

The addition of a joystick (MC-4B-JOYSTICK, National Apertures) provides an alternative method of controlling the x, y stage. The advantage of the joystick control is that it allows maximum operator control flexibility. The drawbacks to this method are that no automation is possible, and as there are no set coordinates for the stage to move to, the operator needs to be aware of the position of the stage in relation to the laser so that the correct sample is analysed.

2.5 Evaluation of AP-MALDI Ion Source and Control Software

2.5.1 Materials

α -Cyano-4-hydroxycinnamic acid (CHCA) matrix was purchased from Sigma-Aldrich Chemical Co. Ltd. (Poole, Dorset, UK) and was used without further purification. Sequencing grade modified trypsin (product number V5111) was purchased from Promega Ltd (Southampton, UK). All solvents were obtained from Fisher (Loughborough, Leicestershire, UK).

2.5.2 AP-MALDI-QIT

All experiments were carried out using the modified AP-MALDI source interfaced with a Finnigan LCQ quadrupole ion trap mass spectrometer (ThermoFinnigan, San Jose, CA, USA). The AP-MALDI target was held at + 2.7 kV using the LCQ high voltage power supply. The ion trap automatic gains control (AGC) was deactivated, heated capillary inlet temperature set to 350 °C and the gate time set to 500 ms. The nitrogen laser (Laser Photonics, Orlando, FL) with performance characteristics $\lambda = 337$ nm, 1.4 mJ per 600 ps pulse width, was operated continuously at 5 Hz and laser power was regulated by adjusting the alignment of the fibre and laser until the threshold power was attained.

2.5.3 Effects of Sample Preparation Technique and Application to Target

α -cyano-4-hydroxycinnamic acid (CHCA) matrix solutions were prepared as 10 mg mL⁻¹ solutions in both 80 % acetonitrile/water + 0.1 % trifluoroacetic acid and 50 % acetonitrile/water + 0.1 % trifluoroacetic acid. The synthetic peptide gp70 (SPSYVYHQF) was dissolved in water to concentrations of 5 pmol/ μ L and 2 pmol/ μ L. The sample was applied to the target plate using the following methods, (i) thin layer method, matrix (1 μ L) was applied to the target and allowed to air dry, followed by the addition of sample (1 μ L) which was also air dried. (ii) Thin layer method where the layers were reversed so that sample (1 μ L) was dispensed first, air dried and matrix (1 μ L) applied and dried. (iii) Dried-

droplet method, sample (1 μ L) and matrix (1 μ L) were mixed on the target plate and allowed to air dry. (iv) Sandwich method, matrix (0.5 μ L) dispensed onto the target and air dried, followed by the subsequent layers of sample (1 μ L) and matrix (0.5 μ L). The sample application methods were repeated using the different matrix solutions.

2.5.4 Analysis of gp70 and neurotensin

The matrix, α -cyano-4-hydroxycinnamic acid (CHCA), obtained from Sigma-Aldrich, was used without further purification and prepared as a 10 mg mL⁻¹ solution in 50 % acetonitrile/water + 0.1 % trifluoroacetic acid (TFA). Peptide samples, gp70 (SPSYVYHQF) and neurotensin (PGlu-LYENKPRRPYIL) were dissolved in d.d H₂O to a concentration of 20 pmol μ L⁻¹.

Samples were applied to the target plate using the thin-layer method,⁴ matrix solution (1 μ L) was spotted onto the plate and allowed to air dry followed by the addition of peptide solution (1 μ L) which was also air dried. The target plate was then inserted into the AP-MALDI source and analysed by mass spectrometry and tandem mass spectrometry (MSⁿ).

2.5.5 Proteolytic Digestion of Ubiquitin

Lyophilized trypsin gold was reconstituted in 50 mM acetic acid to 1 μ g μ L⁻¹ and stored at -80° C until required. Ubiquitin (1500 pmol), ammonium bicarbonate (50 μ L of 100 mM), trypsin (2 μ L) and distilled water (23 μ L) were placed into an eppendorf and allowed to digest overnight at 37 °C. The reaction was quenched by the addition of trifluoroacetic acid to give a final concentration of 0.1 %. The resultant peptides were analysed by AP-MALDI/QIT mass spectrometry using the optimised method for sample preparation and application to target.

2.6 Results and Discussion

The modified AP-MALDI source was evaluated using peptide standards and a ubiquitin tryptic digest. The set of results were obtained using the motion control program developed in Section 2.4.3. Using the CCD camera and a TFT screen, it was possible to assess the homogeneity of each sample spot and look for sweet spots (areas where consistently high signal is obtained) which appear as slightly darker patches within the sample. Since the developed software contained the fine tune function it was possible to move fluently and precisely within the sample to target the sweet spots generating high quality spectra.

2.6.1 Sample Preparation Techniques

The effect of the concentration of the acetonitrile on the AP-MALDI/QIT spectra was investigated for 50 % and 80 % acetonitrile. It was observed that matrix prepared with 50 % acetonitrile solution is more effective than the 80 % for all preparation techniques, containing a higher proportion of sweet spots and more homogeneous crystallisation, which can be attributed to the slower evaporation of the solvent and hence slower crystal formation. Significant improvement in signal was observed using the lower acetonitrile concentration. Figure 2.11 shows the mass spectra for 2 pmol gp70 from a 50 % acetonitrile matrix solution) and from an 80 % acetonitrile solution both applied to the target using the sandwich method. The average ion intensity of the $[M+H]^+$ ion obtained for 2 pmol gp70 using 50 % acetonitrile is 1.1×10^6 in comparison with 3.2×10^5 for 80 % acetonitrile.

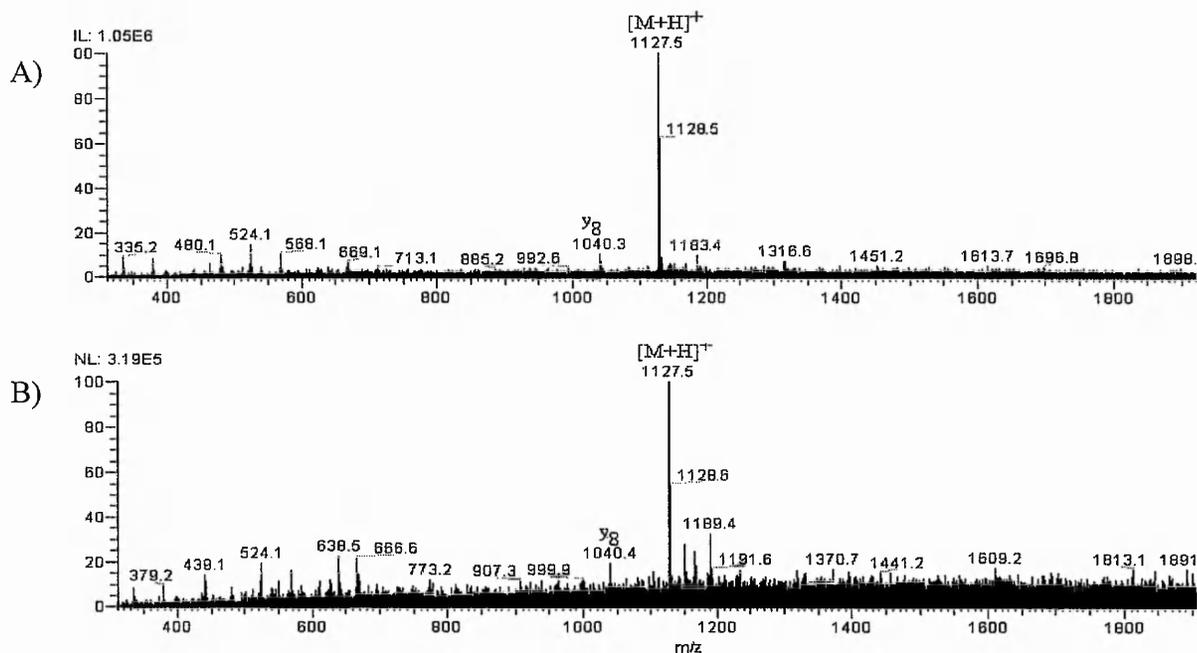


Figure 2.11: Mass spectra of 2 pmol gp70 A) from 50 % acetonitrile CHCA matrix solution B) from 80 % acetonitrile CHCA matrix solution

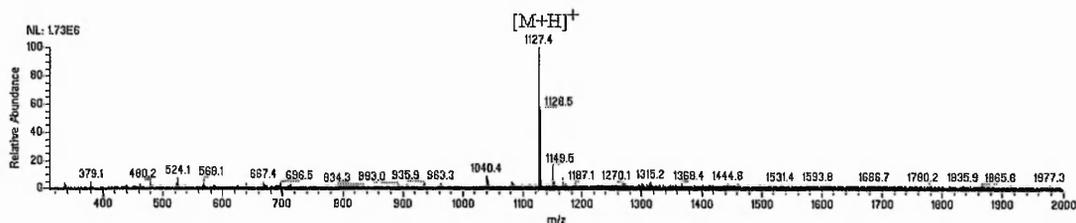
The AP-MALDI/QIT spectra of gp70 (2 pmol) were obtained using the following four different techniques to apply the sample to the target plate:

- 1) Thin layer method, matrix applied to the target and allowed to dry, followed by the addition of sample.
- 2) Thin layer method where the layers were reversed, so that sample was dispensed first and air dried then a matrix layer added on top.
- 3) Dried-droplet, sample and matrix mixed together on the target plate and allowed to air dry
- 4) Sandwich method, the matrix was deposited onto the target plate and allowed to dry followed by the addition of a sample layer and another matrix layer.

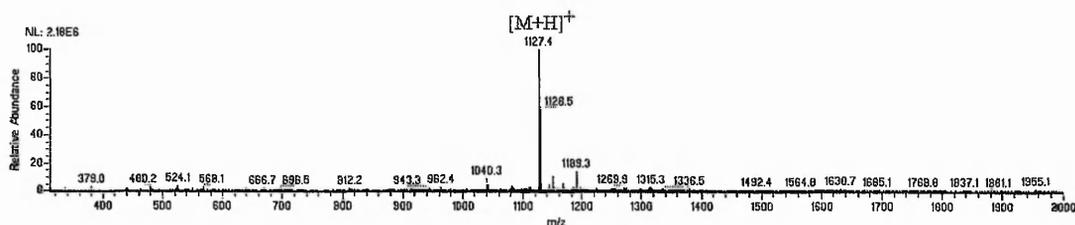
Figure 2.12 shows that all the application methods produced good quality spectra; based on multiple analysis. However, consistently high signal was obtained using the thin layer method, where sample was deposited onto the target plate and allowed to dry before the

subsequent addition of a matrix layer. Therefore this preparation method was used for all further investigations.

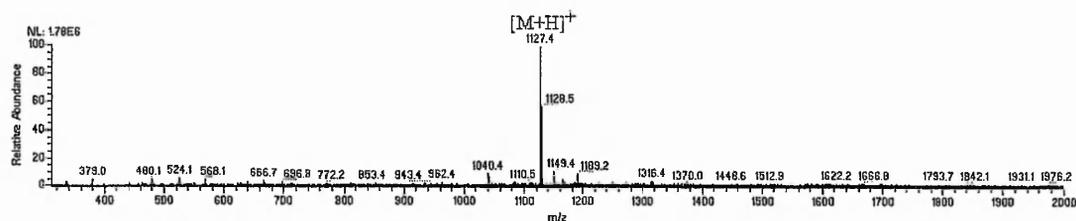
A)



B)



C)



D)

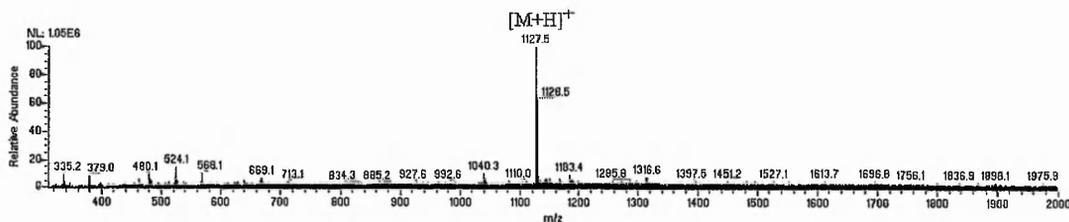


Figure 2.12: AP-MALDI/QIT mass spectra of 2 pmol gp70 applied to the target by A) Thin layer method with CHCA applied to the target followed by sample. B) Thin layer method with sample applied to the target plate followed by CHCA C) Dried-droplet and D) Sandwich method

2.6.2 AP-MALDI-QIT Analysis of gp70 and neurotensin

The resulting mass spectrum for the AP-MALDI-QIT analysis of gp70 (SPSYVYHQF) is shown in Figure 2.13. The peak at m/z 1127.9 is the $[M+H]^+$ ion, whilst that at m/z 1150.9 corresponds to $[M+Na]^+$. The peak at m/z 1040.8 relates to the y_8 fragment ion. The ion observed at m/z 1189.7 is a minor constituent of the mixture, tandem mass spectrometry (MS/MS) identified this as a peptide containing tyrosine, lysine and phenylalanine.

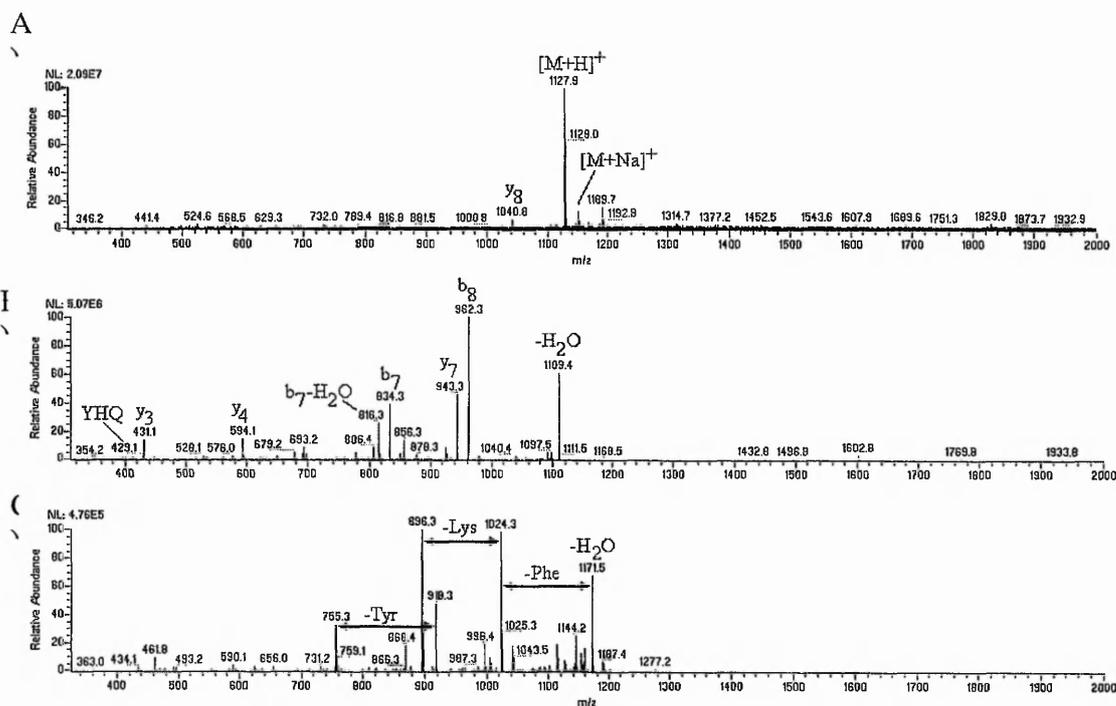


Figure 2.13: AP-MALDI-QIT mass spectrometry of gp70 A) mass spectrum of 20 pmol gp70 B) MS/MS spectrum of the $[M+H]^+$ ion at m/z 1127.9 C) MS/MS of m/z 1189.7

Tandem mass spectrometry was performed on the $[M+H]^+$ ion (m/z 1127.9) producing useful sequence information as shown in Figure 2.13b. The dominant product ion present in the MS/MS spectrum was the b_8 ion at m/z 962.3, with the b_7 and b_7-H_2O ions observed at m/z 834.3 and 816.3 respectively. The loss of water from the $[M+H]^+$ ion was observed at m/z 1109.4. A number of y -type fragment ions were observed in the spectrum with the y_3 , y_4 , y_6 and y_7 ions at m/z 431.1, 594.1, 856.3 and 943.2 respectively giving further

sequence information. An internal fragment corresponding to YHQ was present at m/z 429.1.

The peptide neurotensin (P_{Glu}-LYENKPRRPYIL), was then analysed and the AP-MALDI MS spectrum obtained from a CHCA matrix is shown in Figure 2.14. The dominant peak in the mass spectrum is the protonated molecular ion at m/z 1672.9. Tandem mass spectrometry of this ion produced a product ion spectrum shown in Figure 2.14b, which consists mainly of y ions (y_5 , y_6 , y_7 , y_8 , y_9 and y_9 -NH₃) along with two b ions (b_7 and b_{11}). The most prominent peak in this spectrum results from the loss of water from the precursor ion.

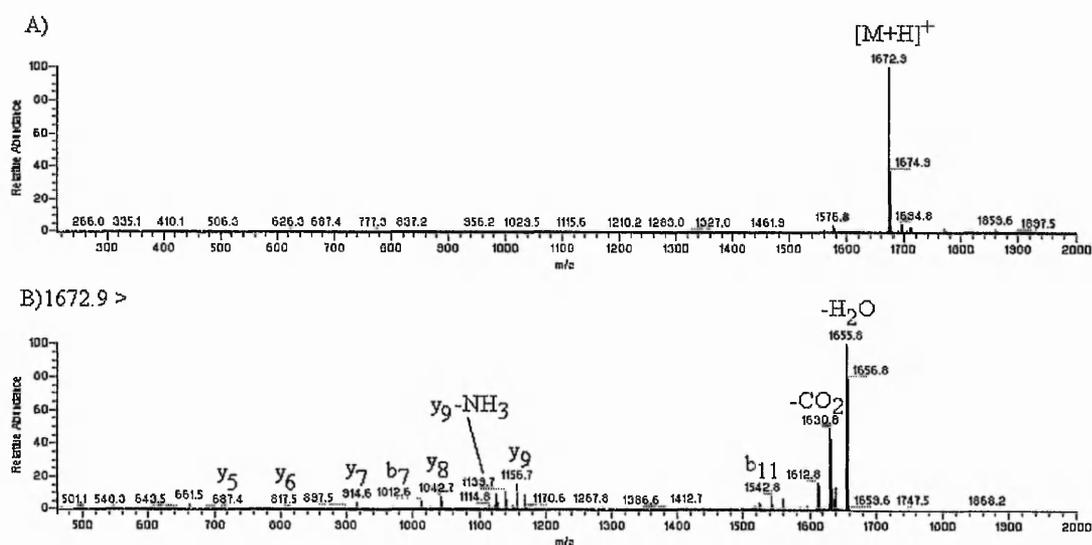


Figure 2.14: AP-MALDI-QIT mass spectrometry of neurotensin A) Mass spectrum obtained for the analysis of 20 pmol neurotensin B) MS/MS spectrum of the [M+H]⁺ ion at m/z 1672.9

2.6.3 AP-MALDI-QIT Analysis of Tryptic Peptides

The peptide mass fingerprint of ubiquitin (1.5 nmol) is shown in Figure 2.15, several peptides associated with the protein can be observed, the sequences of which are shown in Table 1.

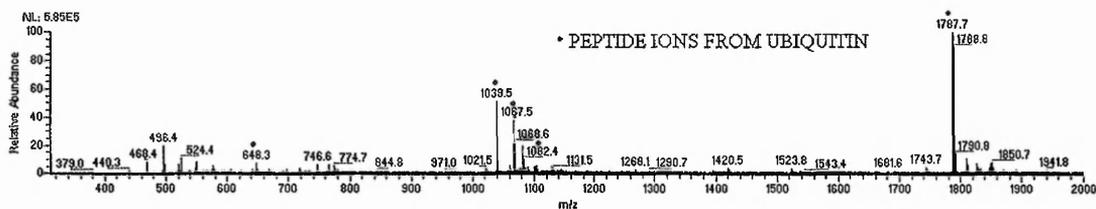


Figure 2.15: Peptide mass fingerprint of ubiquitin digest

Table 1: Tryptic peptides derived from ubiquitin observed in the AP-MALDI-QIT mass spectrum

$[M+H]^+$	Sequence of Fragments
648.3	(R)LIFAGK(Q)
1039.5	(K)EGIPPDQQR(L)
1067.5	(K)ESTLHLVLR(L)
1081.4	(R)TLSDYNIQK(E)
1523.8	(K)IQDKEGIPPDQQR(L)
1787.7	(K)TITLEVEPSDTIENVK(A)

Tandem mass spectrometry was performed on the $[M+H]^+$ ion of the digest fragment TITLEVEPSDTIENVK (m/z 1787.7) resulting in useful sequence information as shown in Figure 2.16. Several y and b ions were observed as well as internal fragments.

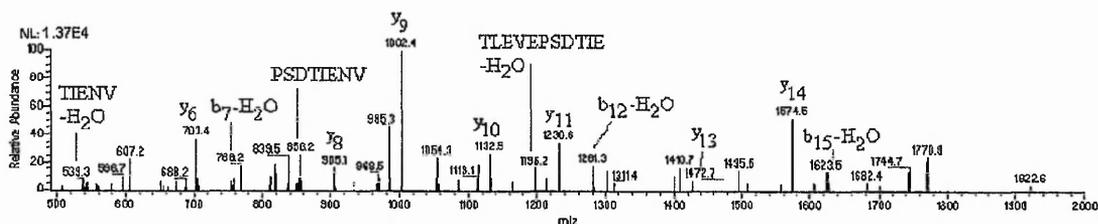


Figure 2.16: AP-MALDI-QIT MS/MS spectrum of tryptic peptide m/z 1787.7 derived from ubiquitin

2.7 Conclusion

A software interface has been developed to control the sample stage contained within the modified AP-MALDI source in relation to the fibre optic. An evaluation of AP-MALDI-QIT for peptide identification and sequencing has been performed using the peptides gp70 and neurotensin, as well as the tryptic digest of ubiquitin. Sample preparation and application of the sample to the target plate have been optimised.

The experimentation has demonstrated the AP-MALDI-QIT is capable of generating high quality sequence information on peptides. This shows that the modified AP-MALDI source has potential as a rapid method for analysing proteomic samples.

2.8 References

- 1) Creaser C.S., Reynolds J.C., Harvey D.J., *Rapid Commun. Mass Spectrom.*, **16**, 176-184 (2002)
- 2) Creaser C.S., Reynolds J.C., Hoteling A.J., Nichols W.F., Owens K.G., *Eur. J. Mass Spectrom.*, **9**, 33-44 (2003)
- 3) Creaser C.S., Reynolds J.C., Harvey D.J., *Rapid Commun. Mass Spectrom.*, **16**, 176-184 (2002)
- 4) Vorm O., Roepstorff P., Mann M., *Anal. Chem.*, **66**, 3281-3287 (1994)

Chapter 3

Capillary LC Interfaced with AP-MALDI/QIT for the Identification of Tryptic Peptides

3.1 Introduction

The direct analysis of proteolytic digests by MALDI mass spectrometry is a well-established method for peptide mapping of proteins separated by 2D SDS-PAGE due to its simplicity and wide application. The main disadvantage of this approach is that proteins may not be identified because of ionization effects that result in weak ion responses for some peptides in the presence of other peptides¹ and the presence of two, or more proteins in a single extracted spot.² Low efficiencies for in-gel digestion and peptide extraction, particularly for membrane proteome analysis, are also important factors affecting gel-based techniques. The combination of capillary liquid chromatography (LC) with mass spectrometry has reduced the latter problem by separating the tryptic peptides prior to mass spectrometric analysis.³ The coupling of ESI on-line to LC⁴ is straightforward, but susceptible to ion suppression in the presence of salts and other contaminants. MALDI is less susceptible to the effects of contaminants and benefits from the ease of spectral interpretation; for this reason it has also been coupled to LC separation.⁵ The application of LC/MALDI for quantitative proteome analysis is also an important area where LC/MALDI can offer some advantages over LC/ESI MS. Off-line methods for interfacing MALDI with LC separations involve collecting fractions or depositing the eluate directly onto the MALDI target plate and the combination has been demonstrated for the assignment of disulfide linkages in polypeptides⁶ and for protein identification,⁷⁻¹⁰ and corticosteroids.¹¹ As previously discussed in Chapter 1 database searching of the peptide mass fingerprint (PMF) data obtained by vacuum MALDI-TOF mass spectrometry can lead to rapid protein identification. However, identification of proteins may fail if too few peptide ions are observed in the PMF¹² and post source decay (PSD) sequence information may be limited to a small number of the more abundant peptides. The application of tandem mass spectrometry (MSⁿ) to peptide derived ions using ion trap or hybrid q-TOF instruments provides both sequence and structural information, which increases the confidence in

protein identification, as a single peptide sequence may be sufficient to confirm the identify of a protein unambiguously.¹³

Atmospheric pressure-MALDI (AP-MALDI) has been interfaced with oa-TOF and ion trap mass spectrometry^{14, 15} and demonstrated to produce high quality mass spectra from low femtomole amounts of analytes.¹⁶ The successful combination of AP-MALDI with a quadrupole ion trap enables multi-stage tandem mass spectrometric analysis to be carried out yielding sequence data that is not readily available using a TOF mass spectrometer.^{14, 17, 18} AP-MALDI-QIT MS has been compared with MALDI-TOF, microcapillary liquid chromatography/ tandem mass spectrometry (μ LC/MS/MS)¹⁹ and nanoLC/MS/MS²⁰ for protein identification. Sequence coverage was lowest for AP-MALDI, although sample throughput was reported as 10-fold greater than for μ LC/MS/MS. The study concluded that AP-MALDI was a more selective and robust technique than MALDI-TOFMS. A recent comparison of AP-MALDI with vacuum MALDI has been carried out using a hybrid quadrupole linear trap (QqLIT) for the analysis of tryptic digests.²¹ Vacuum MALDI showed improved MS/MS performance with improved sequence coverage. However, there has been no systematic investigation of direct MALDI analysis and LC/MALDI for peptide identification using AP-MALDI-QIT and MALDI-TOF.

AP-MALDI has the advantage that the ion source is rapidly interchanged with other atmospheric ionization techniques, for example ESI, and offers a more cost effective solution to protein analysis than other MS/MS methods such as qTOF and TOF-TOF. ESI-qTOF is a well established technique, capable of extremely high sensitivity making it a widely used instrumental configuration in the proteomics field.

This chapter describes the development and evaluation of off-line capillary LC/AP-MALDI-QIT for protein identification using a bovine serum albumin (BSA) proteolytic digest. The technique is compared with LC/MALDI-TOF, LC/ESI-qTOF, and with the direct analysis of digests without prior LC fractionation.

3.2 Experimental

Bovine serum albumin (BSA) was digested as described in section 3.2.2 and the resulting tryptic peptides were spotted onto the vacuum MALDI and AP MALDI target plates with, and without, LC separation, and analysed using TOF and ion trap mass analysers respectively. Samples were also analysed directly by LC/ESI/MS after digestion. The experimental procedure is summarised in Figure 3.1 and detailed below.

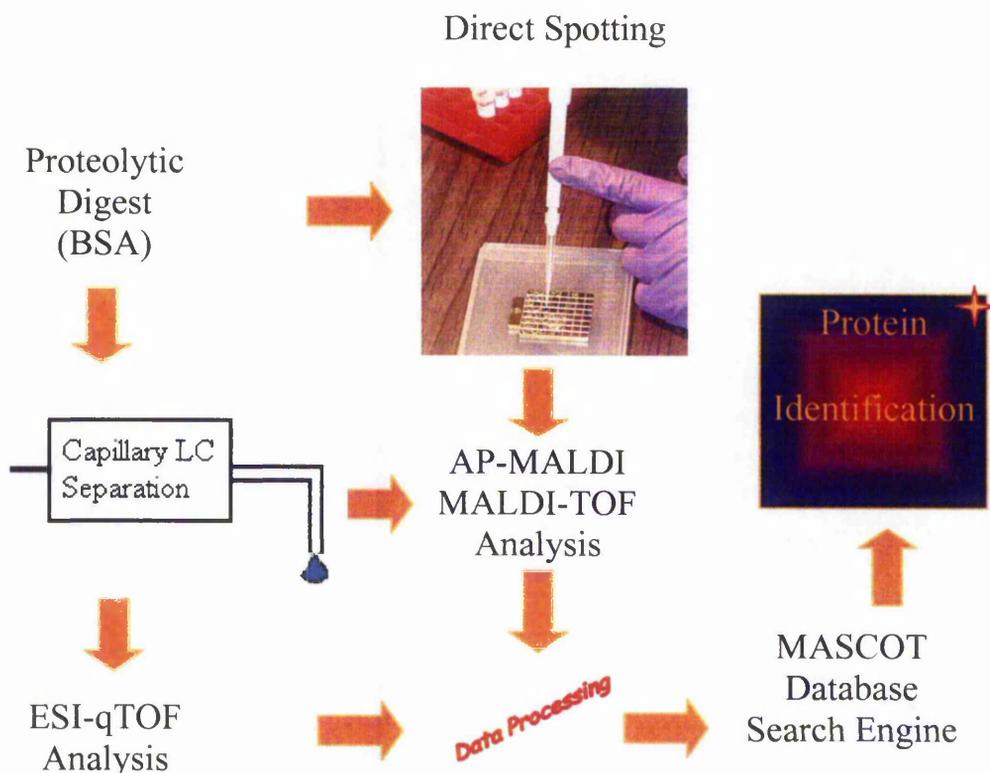


Figure 3.1: Summary of the methodology used for the analysis of tryptic digests with, and without capillary LC separation.

3.2.1 Materials

α -Cyano-4-hydroxycinnamic acid (CHCA) matrix was purchased from Sigma-Aldrich Chemical Co. Ltd. (Poole, Dorset, UK) and was prepared as a saturated solution in acetone for vacuum MALDI and as a 10 mg ml^{-1} solution in 50 % acetonitrile/water + 0.1 % TFA for AP-MALDI experiments. HPLC grade acetonitrile and acetone were obtained from Rathburn (Walkerburn, UK). Ultrapure water ($18.2 \text{ M}\Omega$) obtained from a water purification

system (ELGA, High Wycombe, UK) was used for the preparation of matrix and sample solutions. Sequencing grade modified trypsin (product number V5111) was purchased from Promega Ltd (Southampton, UK)

3.2.2 Sample Preparation and Capillary LC separation

A solution phase tryptic digest of BSA (Sigma, Poole, Dorset, UK) was prepared by reduction and alkylation with dithiothreitol (DTT) and iodoacetamide, followed by overnight digestion with trypsin. LC separations were carried out using a Waters capLC™ capillary LC instrument (Waters, Milford, MA, USA) fitted with an analytical capillary column (Zorbax XDB C18, 7.5 cm x 75 µm i.d., packed in a New Objectives picofrit capillary). Samples (10 µL containing 100 fmol-10 pmol of digest) were loaded directly onto the analytical column in 0.1 % formic acid at 1.0 µL min⁻¹ using a syringe pump. After loading, the column was connected to the Cap LC and peptides were eluted, over a period of 20 minutes at a flow rate of 200-340 nL min⁻¹, using a gradient of 5-35 % B (acetonitrile/water 95:5 v/v containing 0.1% v/v formic acid), solvent A (water/acetonitrile 95:5 v/v containing 0.1% v/v formic acid). The LC column eluate was deposited manually onto adjacent spots on the vacuum MALDI or AP-MALDI target plates in seventeen one-minute wide fractions. The procedure was repeated using protein concentrations in the range 100 fmol to 1 pmol. Matrix was added using the thin layer method. In the case of vacuum MALDI-TOF, the target plates were pre-spotted with matrix (0.3 µL) and LC fractions deposited on top of the matrix. For AP-MALDI the LC eluate was deposited directly onto the plate and a matrix layer applied on top of the sample. For the direct analysis of the BSA digest the samples (1 µL) were deposited onto the target plates without prior LC separation and matrix (1 µL) applied as for LC eluates. The LC separation experiments were carried out in collaboration with P. Green of Syngenta, Jealott's Hill International Research Centre, Bracknell, UK.

3.2.3 AP-MALDI-QIT

AP-MALDI experiments were performed using a Mass Technologies (Burtonsville, MD, USA) AP-MALDI ion source equipped with a 10 Hz, 300 μ J, 337 nm N₂ laser interfaced with a LCQ ion trap mass spectrometer (ThermoFinnigan, San Jose, CA, USA). For all experiments the ion trap automatic gain control was deactivated, the gate time set to 500 ms and the heated capillary inlet temperature was set to 350 °C. The AP-MALDI target plate was held at 2.7 kV and no auxiliary or sheath gas were used.

Data-dependent acquisition was configured so that each full scan (scan range m/z 700-2000, 50 shots accumulated /spectrum, automated acquisition) was followed by MS/MS scans on the three most intense precursor ions. A precursor ion was reselected for MS/MS up to three times within a one minute window, after this period the ion was placed onto a dynamic exclusion list for the remainder of the seven minute run. The collision energy for MS/MS was set to 35 % and the scan range was limited to 500-2000 Da in order to exclude matrix associated ions.

3.2.4 MALDI-TOF

MALDI-TOF experiments were performed using a Bruker Ultraflex mass spectrometer (Bruker Daltonics, Coventry UK) equipped with a 20 Hz, 337 nm N₂ laser. Spectra were acquired manually over the range m/z 600-3500. For each sample 200 shots were acquired in 20 shot intervals. All spectra were acquired using FlexControl (version 2.0) and processed using FlexAnalysis (version 2.0). Resulting data was not smoothed and monoisotopic peaks were selected. For all samples the instrument was operated in positive ion reflectron mode using a delayed extraction of 100 ns.

3.2.5 ESI-qTOF

Mass spectrometry was performed on a quadrupole time-of-flight mass spectrometer (Q-TOF2, Micromass, Manchester, UK), fitted with an embedded PC acquisition system

(EPCAS) upgrade and connected to a Waters CapLCTM, autosampler and ternary pump system (Waters, Milford, MA, USA). One solvent channel was used for loading samples onto the trapping column at 30 $\mu\text{L min}^{-1}$ and the other two channels were used to deliver a solvent gradient directly to the capillary column, by-passing the autosampler. The electrical connection for electrospray was made in a low dead volume union placed at the inlet of the capillary column. Sub-microlitre flow rates of ca. 300 nL min^{-1} through the capillary were delivered by flow splitting with an input flow of 3 $\mu\text{L min}^{-1}$, using a T piece situated between the pumps and the switching valve. The instrument was calibrated using the product ions from the fragmentation of doubly charged glu-fibrinopeptide B (500 $\text{fmol}/\mu\text{L}$, in acetonitrile: water 1:1 v/v containing 0.1% v/v formic acid) directly infused through the capillary column using a syringe pump.

Peptide digest samples (typically 10 μL) were loaded onto the trapping column (C18 Pepmap, 1 mm x 0.3 mm, LC Packings, Dionex) in 0.1 % formic acid at 30 $\mu\text{L min}^{-1}$. The peptides were back-flushed onto the analytical capillary column (Zorbax XDB C18, 7.5 cm x 75 μm i.d., packed in a New Objective picofrit) at flow rate of 200-340 nL min^{-1} and eluted using the same 20 minute gradient used for eluting peptides onto the AP and vacuum MALDI sample plates. Solvents were: A (water/acetonitrile 95:5 v/v containing 0.1 % v/v formic acid), B (acetonitrile/water 95:5 v/v containing 0.1 % by v/v formic acid) and C (0.1 % v/v formic acid in water). The mass spectrometer was configured to detect peptide ions in survey mode. Automatic charge state recognition was used to select up to three ions of charge states 2, 3 or 4 for subsequent fragmentation by MS/MS. Only the three most intense ions of intensity > 10 counts/s were selected. Ions were not selected again for periods of 30 s during the 35 minute gradient. The energy applied to fragment each ion was taken from preset tables and depended on m/z and charge state. MS/MS fragmentation data were collected from each of the three ions for 3 s before the instrument

acquired another survey scan to select further ions for fragmentation. The ESI-qTOF experiments were carried out in collaboration with P. Green, at Syngenta, UK.

3.2.6 Database Searching

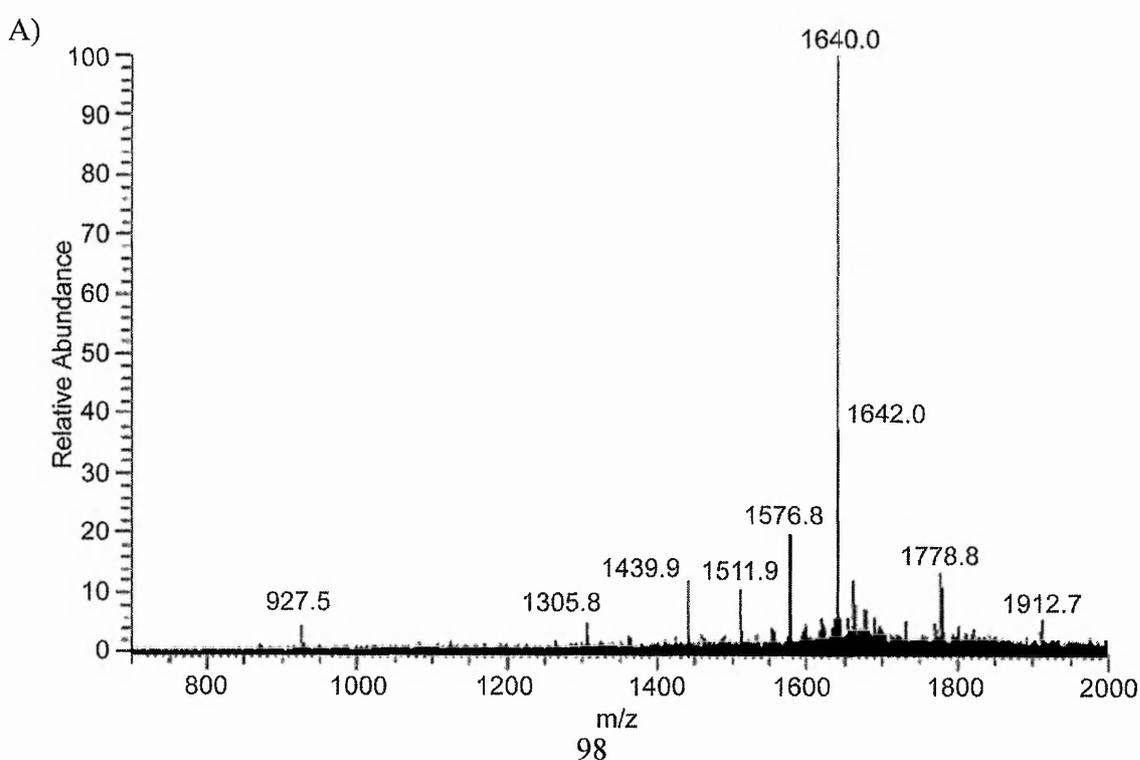
All spectra were processed by MASCOT (Version 2.0, MatrixScience, London) database searching. For ESI-qTOF data, peak lists were created using the PeptideAuto function in ProteinLynx v4.0 (Micromass). AP-MALDI-QIT data was extracted from spectra resulting from each of the LC fractions and manually converted into a combined peak list. An intensity threshold was set in order to minimise interference from noise when the data was submitted for PMF and MS/MS ion searches. Typical search parameters were as follows: consideration of up to 1 incomplete cleavage site per peptide, fixed modification carboxyamidomethylcysteine (CAM), variable modification oxidation of methionine. Data was submitted to MASCOT with the following peptide mass tolerances, for AP-MALDI-QIT: PMF search 0.3 Da and for MS/MS ion search peptide tolerance \pm 500 ppm and MS/MS tolerance \pm 0.6 Da. Peptide tolerances employed for TOF and qTOF data were \pm 75 ppm and \pm 50 ppm respectively with an MS/MS tolerance of 0.1 Da for the QTOF MS/MS ion search.

3.3 Results and Discussion

The objective of this work was to evaluate the potential of capillary liquid chromatography combined with atmospheric pressure MALDI/ion trap mass spectrometry (LC/AP-MALDI-QIT) for protein identification, using a scan routine that generated a peptide mass fingerprint and data-dependent tandem mass spectrometry from each single sample spot. These data are compared with the techniques of capillary LC/vacuum MALDI-TOF and LC/ESI-quadrupole TOF, using well-established instrumental platforms and protocols for the analysis of a BSA tryptic digest. Tryptic digest samples with and without LC separations were spotted onto the MALDI target plates and analysed, whilst samples were directly introduced into the LC/ESI after digestion. The data from all three configurations were searched against the same MSDB database, using MASCOT

3.3.1 LC/MALDI Peptide Mass Fingerprinting

Typical PMF mass spectra for the BSA digest (1 pmol), obtained using AP-MALDI-QIT and vacuum MALDI-TOF for the LC fraction corresponding to an elution window of 22-23 min are shown in Figure 3.2.



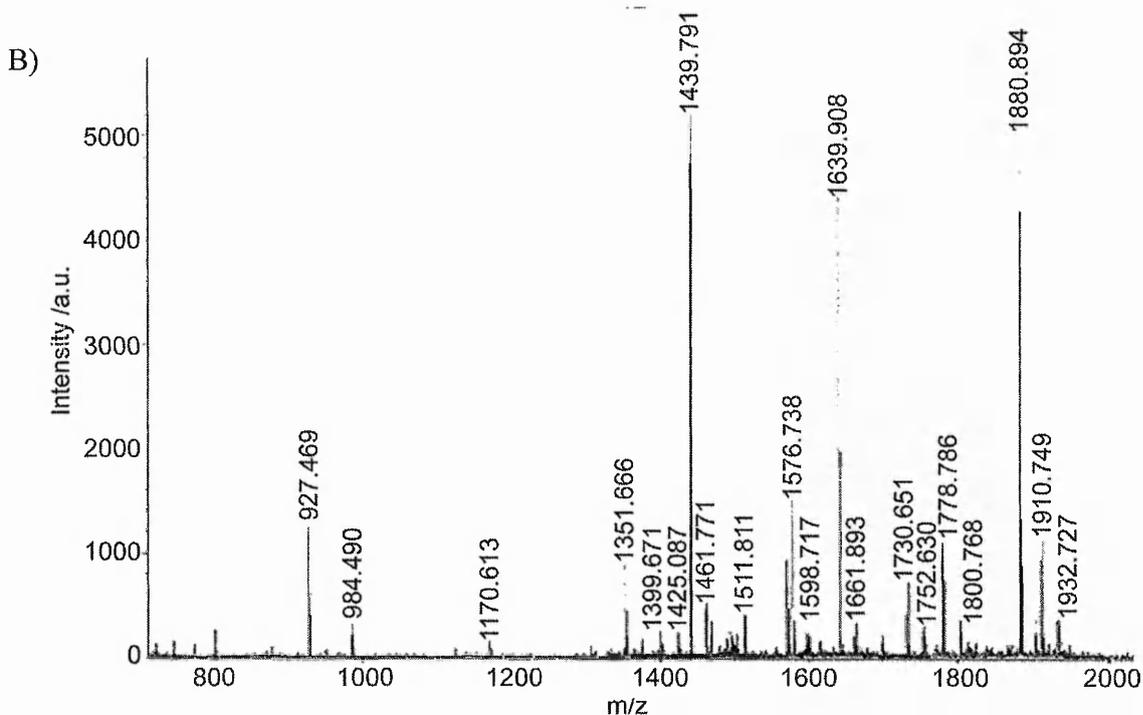


Figure 3.2: PMF mass spectra for a BSA tryptic digest (1 pmol) corresponding to the LC fraction eluting at 22-23 minutes analysed by (a) LC/AP-MALDI-QIT and (b) LC vacuum-MALDI-TOF.

The spectra were acquired on different instrumental platforms and are therefore influenced by parameters such as sample preparation, mass resolution and accuracy, mass range and data file processing. For example, the sample/matrix deposition procedures reflect the optimised protocols for the two techniques, which for AP-MALDI required deposition of the eluate onto the target plate prior to matrix addition, whilst the reverse procedure gave the best results for vacuum MALDI. In both cases ions matching predicted BSA tryptic peptides are present in the full scan spectra although the AP-MALDI spectrum has a lower signal to noise (S/N) ratio. For each of the techniques, peak lists were generated from the seventeen LC fractions, combined and submitted to MASCOT for PMF searching. For the LC/AP-MALDI-QIT data, a signal intensity threshold was employed in order to minimise the effects of the higher background noise observed in the mass spectrum. This was manually selected for each data set to be above inherent noise peaks in the spectra.

LC/AP-MALDI-QIT and LC/MALDI-TOF PMF search results for the same BSA digest are compared in Table 3.1. The identification of peptides derived from 1 pmol of BSA by

AP-MALDI, following LC separation and PMF analysis, was confirmed by a MASCOT score of 72 and sequence coverage of 37 % for the 21 peptides matched. It should be noted that some tryptic peptides observed in the spectra, for example m/z 927 in Figure 3.2, were not identified by MASCOT because mass accuracies lay outside the tolerance used for the search. Lower sequence coverage, numbers of matched peptides and MASCOT scores were achieved at 500 fmol and 100 fmol. Significant values (>66) are 95% certain to be peptides derived from the protein and this scoring system allows varying peptides to be weighted depending on their capacity to be identified.

Table 3.1: AP-MALDI-QIT and MALDI-TOF PMF data for the analysis of a BSA tryptic digest with LC pre-separation

	LC/AP-MALDI-QIT				LC/MALDI-TOF			
	Number of peptides matched	Sequence Coverage /%	MASCOT Score	% Submitted/ Matched	Number of peptides matched	Sequence Coverage /%	MASCOT Score	% Submitted/ Matched
1 pmol	21	37	72	16	31	50	125	16
500 fmol	10	18	42	20	29	44	174	22
100 fmol	6	13	40	24	18	32	121	22

As expected, both LC/AP-MALDI-QIT and LC/MALDI-TOF analyses result in greater number of peptides matched with confidence at the 1 pmol level than at lower concentrations. However, LC/MALDI-TOF gave the highest ranking matches and greater sequence coverage over all concentrations tested. This is probably due to the higher mass range of the TOF instrument compared to the QIT, which allowed higher mass peptides ($> m/z$ 2000) to be detected in the PMF, and to the lower S/N ratio observed for the latter technique. The data from each technique were therefore analysed by taking into account the actual number of ions submitted to the database. This allows the percentage of peptides submitted/matched to be calculated, giving a better indication of the relative performance of the two mass spectrometric methods (Table 3.1). The percentage of submitted/matched peptides for AP-MALDI rose from 16 % to 24 % as the concentration was reduced from 1

pmol to 100 fmol, compared with a 16-22 % increase for vacuum MALDI-TOF, showing that AP-MALDI is equivalent to MALDI-TOF on the basis of peptides submitted. Although MALDI-TOF outperformed AP-MALDI-QIT overall there are, nevertheless, a number of advantages to using the AP-MALDI-QIT technique, such as the ability to rapidly interchange ionization sources, the MSⁿ capabilities of the QIT and some evidence that AP-MALDI is a softer ionization technique than vacuum MALDI resulting in less fragmentation.^{22, 23}

3.3.2 LC/AP-MALDI and LC/ESI Tandem Mass Spectrometry

The use of MS/MS for the analysis of BSA tryptic peptides separated by capillary-LC, was examined using AP-MALDI-QIT and ESI-qTOF. The AP-MALDI tandem mass spectrometric data were recorded during the same acquisition as the PMF data reported above, using a data dependant scan routine. Precursor ions were selected automatically for both ESI-qTOF and AP-MALDI-QIT analysis and the resulting fragmentation spectra were combined for each technique and submitted separately to MASCOT. Data acquired from the automatic selection of precursor ions not associated with BSA was minimised by restricting MS/MS scans to a maximum of three per full scan. Results are summarized in Table 3.2, which compares MS/MS data obtained from LC/ESI-qTOF and LC/AP-MALDI-QIT. A typical LC/AP-MALDI MS/MS spectrum is shown in Figure 3.3 for the precursor peptide ion at *m/z* 1888.7 identified by MASCOT as the BSA derived tryptic peptide HPYFYAPELLYYANK.

Table 3.2: LC/AP-MALDI-QIT and LC/ESI-qTOF tandem mass spectrometric analysis of BSA tryptic digest

	LC/AP-MALDI-QIT			LC/ESI-qTOF		
	Number of peptides matched	Sequence Coverage %	MASCOT Score	Number of peptides matched	Sequence Coverage %	MASCOT Score
10 pmol	26	35	525	-	-	
1 pmol	14	24	229	28	40	1133
500 fmol	5	11	59	22	29	998
100 fmol	0	0	-	13	18	356

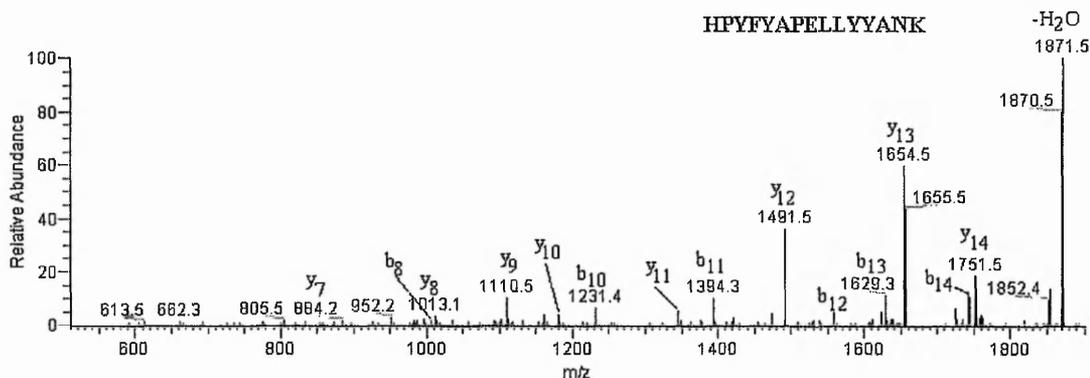


Figure 3.3: LC/AP-MALDI MS/MS product ion spectrum of HPFYAPELLYYANK, (m/z 1888.7) derived from a BSA tryptic digest.

BSA was the highest-ranking match for all concentrations above 500 fmol for both techniques and tandem mass spectrometry greatly enhanced the confidence level for the matched ions compared to the PMF data (Table 3.1). AP-MALDI-QIT MS/MS spectra generated from a single eluate fraction (corresponding to 22-23 min of the LC run) confirmed the assignment of four co-eluting BSA peptides (Table 3.3), which were observed in the PMF, contributing to 6 % of the total protein sequence. Combining LC/AP-MALDI MS/MS data from all LC fractions resulted in 24 % sequence coverage for 1 pmol BSA and 11 % coverage at the 500 fmol level. Sequence coverage and MASCOT scores for the corresponding LC/ESI-q-TOF analysis were higher at both the 1 pmol and 500 fmol concentrations, principally as a consequence of a better S/N ratio. This increased sensitivity allowed sequence coverage to be extended to the 100 fmol level in the ESI experiment. No peptides were matched at the 100 fmol level using AP-MALDI.

Table 3.3: Co-eluting tryptic peptides of BSA (corresponding to 22-23 min of the LC run) observed in the LC/AP-MALDI-QIT PMF of BSA and confirmed by tandem mass spectrometry

Observed m/z	Fragment ion sequence
1640.0	KVPQVSTPTLVEVSR
1576.8	LKPDPNTLCDEFK
1511.9	VPQVSTPTLVEVSR
1439.9	RHPEYAVSVLLR

3.3.3 Direct MALDI Peptide Mass Fingerprinting

In order to establish whether LC separation prior to mass spectrometric analysis improved the number of peptides identified, the sequence coverage and the confidence of correct identification, the AP-MALDI-QIT and MALDI-TOF analysis was performed after direct spotting of the proteolytic digest onto the target plates without pre-separation. The resulting spectra were submitted to MASCOT for PMF and MS/MS ion searching after a manual threshold for ion submission was set to reduce interference from background noise. AP-MALDI and MALDI-TOF PMF analysis of the BSA digest without pre-separation resulted in similar sequence coverage, which was greater than 30 % for all concentrations (Table 3.4).

Table 3.4: AP-MALDI-QIT and MALDI-TOF PMF data for the direct analysis of a BSA digest without LC pre-separation

	AP-MALDI-QIT				MALDI-TOF			
	Number of peptides matched	Sequence Coverage /%	MASCOT Score	% Submitted/ Matched	Number of peptides matched	Sequence Coverage /%	MASCOT Score	% Submitted/ Matched
1 pmol	23	41	78	18	18	34	179	50
500 fmol	18	33	90	27	20	37	204	53
100 fmol	17	32	93	28	25	46	302	71

Surprisingly, the confidence level for the correct identification of BSA derived peptides (MASCOT score) was higher for both techniques when peptides had not undergone prior LC separation (Tables 3.1 and 3.4). Furthermore, AP-MALDI analysis without LC separation showed a greater number of peptides matched and improved sequence coverage compared to LC/AP-MALDI PMF data at all concentrations. In contrast, the number of peptides matched and sequence coverage was only increased for MALDI-TOF analysis at the 100 fmol level. The reasons for these reproducible observations for AP-MALDI-QIT experiments is not clear, but may be due to the acetonitrile gradient of the LC system resulting in differences between the spotting (and hence crystal formation) conditions for

the LC and static samples. However, no systematic bias in peptide identification was observed on the basis of hydrophobicity (based on grand average of hydrophobicity index, termed GRAVY, from ExPASy).²⁴ A negative hydrophobicity value indicates a hydrophilic peptide, whilst a positive value corresponds to a hydrophobic peptide. As expected a greater number of hydrophilic peptides are observed for both AP-MALDI and vacuum MALDI. Hydrophobic peptides are difficult to analyse because they have limited solubility in the aqueous solvents used in MALDI.²⁵ Following LC separation hydrophilic peptides are not seen to the same extent in AP-MALDI in comparison with vacuum MALDI, Figure 3.4. One reason for this discrepancy may be due to incomplete elution of the peptides from the column combined with the low sensitivity of AP-MALDI. Peptide identification was not biased in terms of sequence length, Figure 3.5, or number of positively and negatively charged residues.

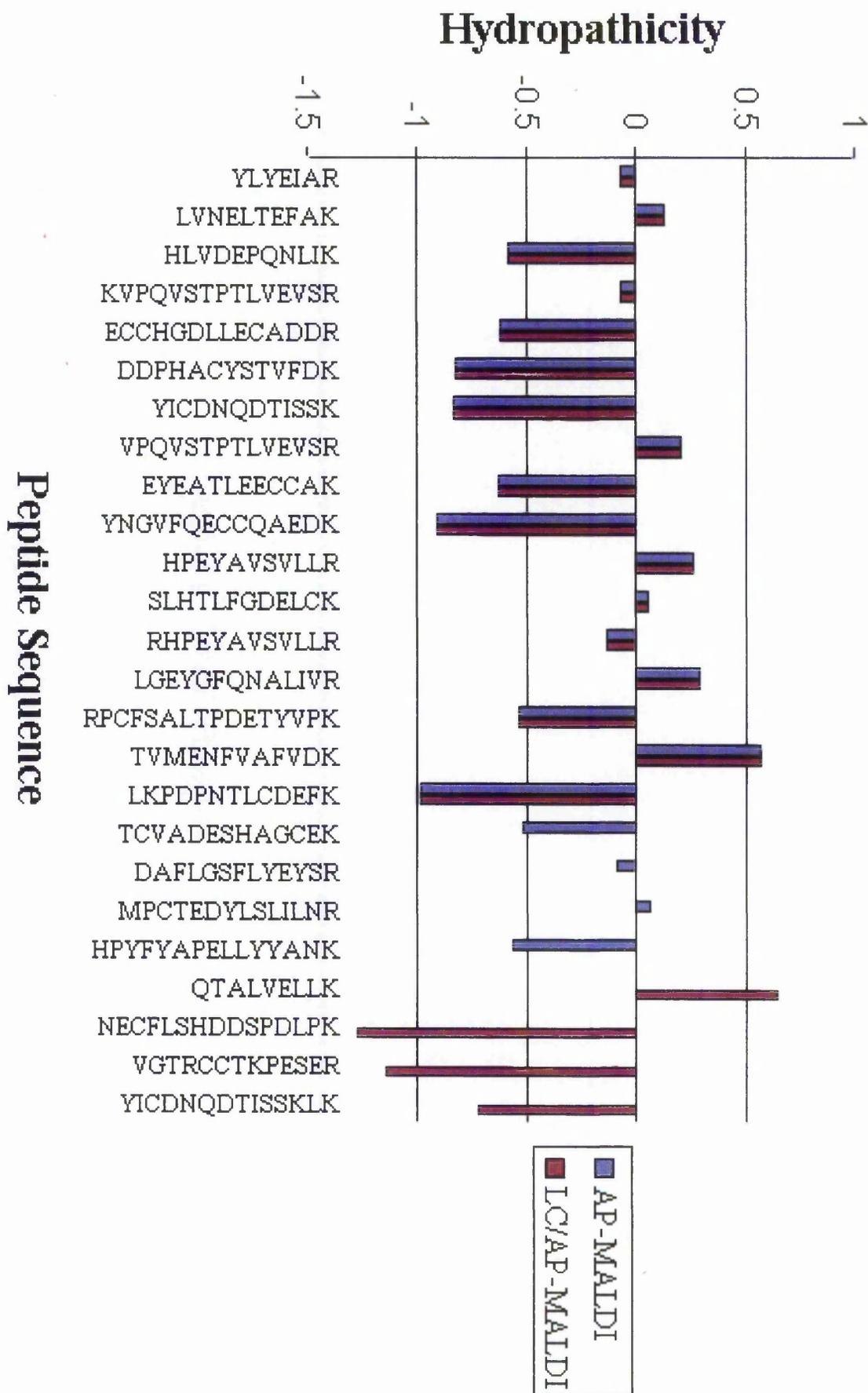


Figure 3.4a: Hydrophobicity values of BSA tryptic peptides (1 pmol) identified by AP/MALDI-QIT, with and without LC separation

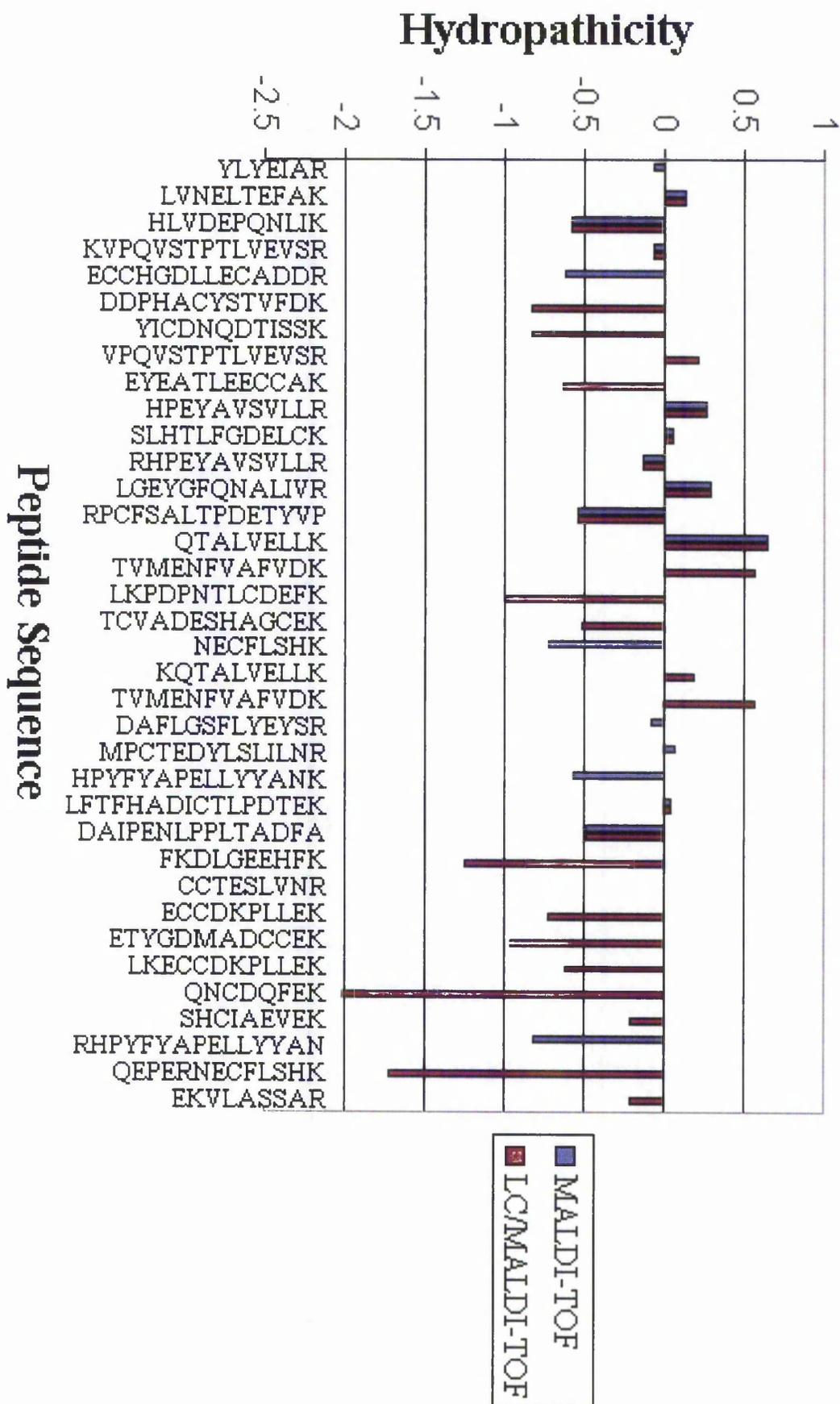


Figure 3.4b: Hydropathicity values of BSA tryptic peptides (1 pmol) identified by vacuum-MALDI-TOF, with and without LC separation.

Number of Amino Acids in Peptide Sequence

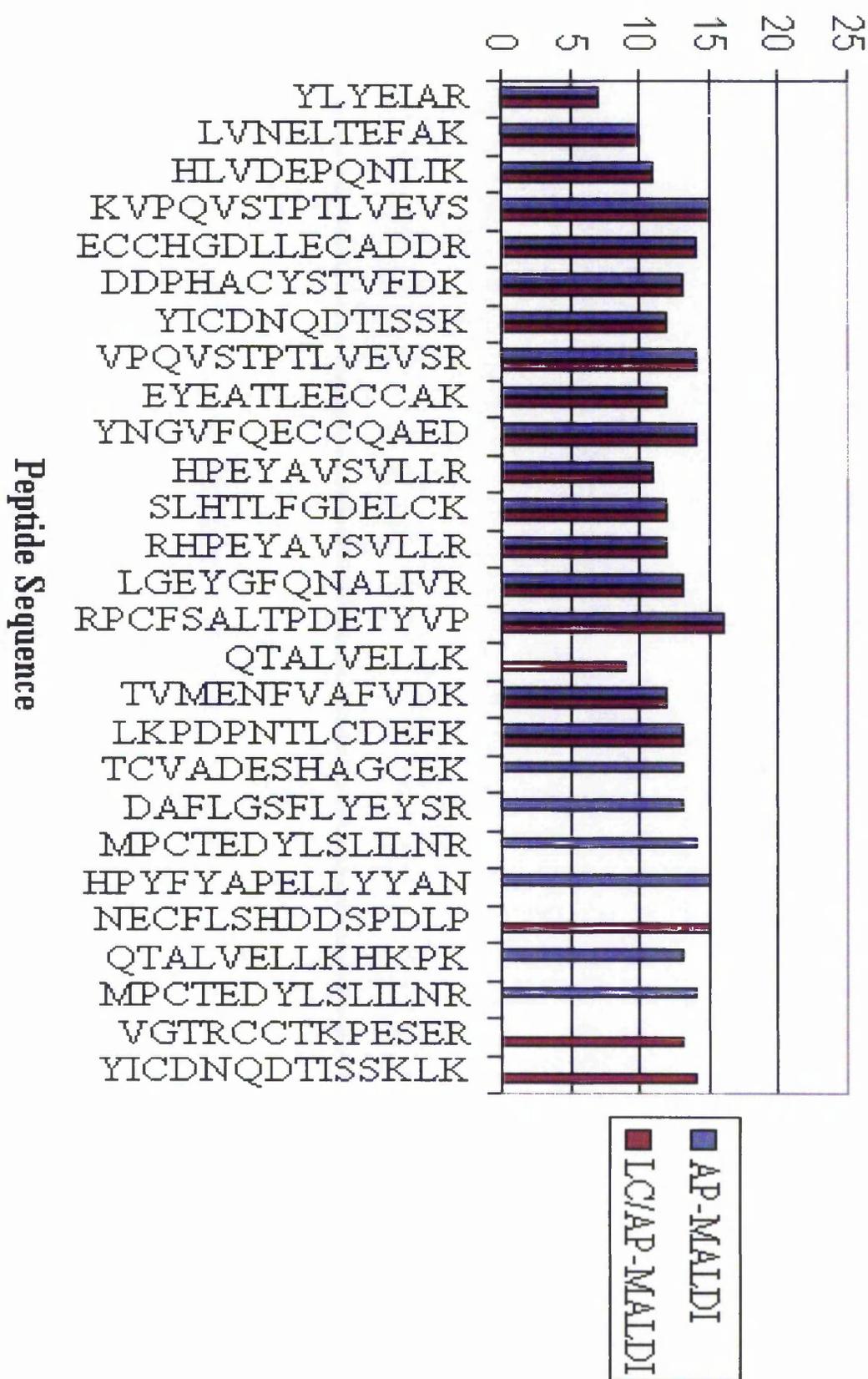


Figure 3.5a: Length of peptides identified at 1 pmol by AP/MALDI-QIT, with and without LC separation

Number of Amino Acids in Peptide Sequence

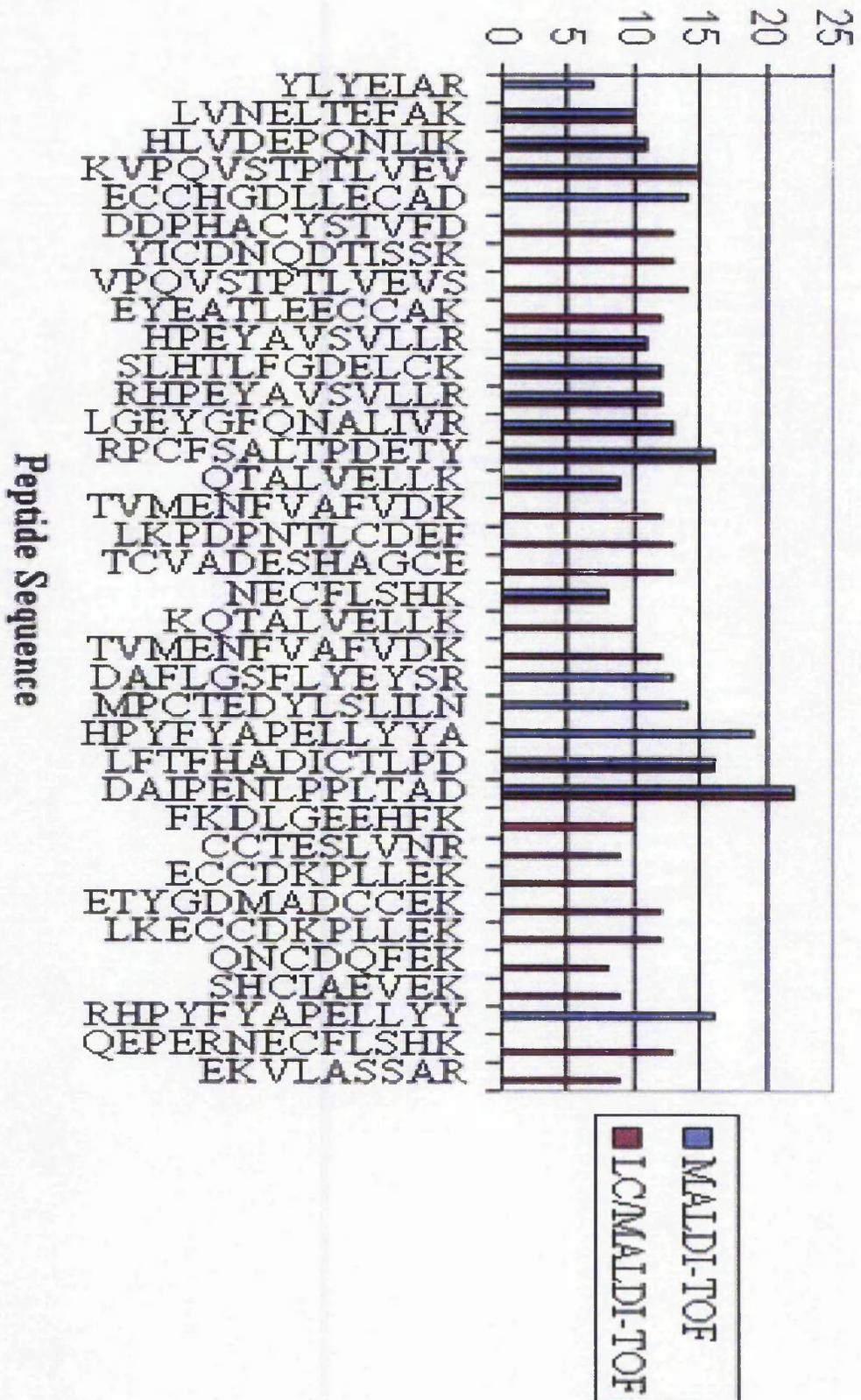


Figure 3.5b: Length of peptides identified at 1 pmol by MALDI-TOF, with and without LC separation.

3.3.4 AP-MALDI-QIT Tandem Mass Spectrometry without LC Pre-separation

AP-MALDI tandem mass spectrometry was carried out on the BSA tryptic peptides without LC separation using the same data dependent acquisition protocol described for LC separated peptides. BSA peptides were identified at all concentrations with sequence coverage in the range 11 – 16 % (Table 3.5).

Table 3.5: AP-MALDI-QIT tandem mass spectrometric analysis of BSA tryptic digest without LC pre-separation

	Number of peptides matched	Sequence Coverage /%	MASCOT Score
1 pmol	8	16	137
500 fmol	7	13	130
100 fmol	8	11	134

As with the LC separated peptide fractions, the MS/MS analysis resulted in lower sequence coverage than direct PMF analysis, but increased confidence in assignment compared to PMF analysis alone. If confidence in protein assignment based on peptides observed in the PMF, rather than sequence coverage, is the prime objective of the analysis, then MS and MS/MS data acquisition combined with direct spotting of the digest provides a rapid approach to protein identification.

3.3.5 Sequence Coverage

A comparison of sequence coverage obtained by the AP-MALDI-QIT, vacuum MALDI-TOF and ESI-qTOF techniques following LC separation is shown in Figure 3.6. The peptides detected for 1 pmol using the various instrumental configurations are summarised in Table 3.6.

MKWVTFISLL LIFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
 FSQYLQOCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPLK PDPNTLCDEF
KADEKKFWGK LYEIIARRHP YFYAPELLEY ANKYNGVFQE COQAEDKGAC
LLPKIETMRE KVLTSSARQR LRCASIQKFG ERAKAWSSVA RLSQKFPKAE
FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCCK NYQEAKDAFL
GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAADDPH ACYSTVFDKL
KHLVDEPQNL IKONCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKOT
ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
STQTALA

Figure 3.6: BSA sequence showing peptides identified at 1 pmol by AP-MALDI-QIT (—), Vacuum MALDI-TOF (—) and ESI-qTOF (—) techniques following LC separation.

Table 3.6: Peptides detected at 1 pmol by AP-MALDI, MALDI-TOF and LC/ESI-qTOF. Values shown are signal-to-noise ratio, peptides shown in bold were identified by Mascot. Peptides identified by AP-MALDI MS/MS are underlined.

Peptide	Hydropathicity	AP-MALDI	LC/AP-MALDI	MALDI-TOF	LC/MALDI-TOF	LC/ESI-qTOF
LYEIIAR	-0.071	14.3	16.7	179.2	102.3	11.3
LVNELTEFAK	0.13	6.7	9.0	29.8	99.2	6.8
HLVDEPQNLIK	-0.582	12.5	25.0	56.2	422.2	21.3
KVPQVSTPTLVEVSR	-0.067	25.0	33.3	58.0	212.7	2.5
ECCHGDLLECADDR	-0.621	16.7	3.0	33.0	23.8	12.4
DDPHACYSTVFDK	-0.823	12.5	25.0	-	109.9	6.4
YICDNQDTISSK	-0.833	4.9	2.5	-	132.6	27.2
VPQVSTPTLVEVSR	0.207	10.6	6.4	-	21.1	9.5
EYEATLEECCA	-0.625	5.6	3.6	-	51.7	39.7
YNGVFOECCQAEDK	-0.914	5.0	3.8	-	27.0	12.9
HPEYAVSVLLR	0.264	20.0	6.7	143.2	293.5	-
SLHTLFGDELCK	0.058	9.4	12.5	29.2	357.0	-
RHPEYAVSVLLR	-0.133	8.3	6.8	280.3	251.1	-
LGEYGFQNALIVR	0.292	10.1	6.7	155.2	903.5	-
RPCFSALTPDETYVPK	-0.537	20.0	20.0	240.2	205.4	-
QTALVELLK	0.644	-	2.3	32.4	352.6	3.4
TVMENFVAFVDK	0.567	9.7	5.0	-	84.7	-
LKPDPNTLCDEFK	-0.985	7.4	14.3	-	71.7	-
TCVADESHAGCEK	-0.515	3.9	-	-	188.2	3.9
NECFLSHK	-0.725	-	-	24.8	317.8	6.4
KQTALVELLK	0.19	-	4.3	-	37.4	6.0
EACFAVEGPK	0.02	-	3.8	-	34.2	7.0
TVMENFVAFVDK	0.567	-	-	-	42.7	-
DAFLGSFLYEYSR	-0.085	5.5	-	278.9	-	-
MPCTEDYLSLILNR	0.064	4.0	-	52.0	-	-
HPYFYAPELLEYANK	-0.567	4.0	-	18.9	-	-

Peptide	Hydropathicity	AP-MALDI	LC/AP-MALDI	MALDI-TOF	LC/MALDI-TOF	LC/ESI-qTOF
NECFLSHDDSPDLPK	-1.275	-	3.4	-	24.8	-
LFTFHADICTLPDTEK	0.044	3.0	-	17.9	85.7	-
DAIPENLPLLTADFAEDKDVCK	-0.491	-	-	23.3	20.3	-
FKDLGEEHFK	-1.25	-	4.0	-	78.0	-
GACLLPKIETMREK	-0.264	-	4.3	-	15.8	-
CCTESLVNR	0.00	-	-	-	101.1	2.6
ECCDKP LLEK	-0.730	-	-	-	95.5	6.1
ETYGDMADCCEK	-0.967	-	-	-	42.4	13.0
LKECCDKP LLEK	-0.617	-	-	-	103.8	2.5
AEFVEVTK	0.175	-	-	-	79.3	13.0
DLGEEHFK	-1.425	-	-	-	271.6	6.7
QNCDQFEK	-2.013	-	-	-	84.1	3.3
SHCIAEVEK	-0.211	-	-	-	353.5	5.9
DAIPENLPLLTADFAEDK	-0.561	18.2	-	-	-	-
QTALVELLKHKPK	-0.523	3.12	-	-	-	-
MPCTEDYLSLILNR	0.064	4.83	-	-	-	-
LSQKFPKAEFVEVTK	-0.38	13.0	-	-	-	-
CCAADDKEACFAVEGPK	-0.124	-	2.63	-	-	-
VGTRCCTKPESER	-1.146	-	3.71	-	-	-
YICDNQDTISSK LK	-0.721	-	3.3	-	-	-
RHPYFYAP ELLYYANK	-0.812	-	-	113.8	-	-
QEPERNECFLSHK	-1.723	-	-	-	41.0	-
EKVLASSAR	-0.211	-	-	-	27.2	-
IETMR	-0.460	-	-	-	-	3.7
KFWGK	-1.260	-	-	-	-	4.2
GACLLPK	0.857	-	-	-	-	13.8
DDPHACYSTVFDK	-0.823	-	-	-	-	13.7
LVTDLTK	0.429	-	-	-	-	15.2
ECCHGDLLECADDR	-0.621	-	-	-	-	12.9
LCVLHEK	0.529	-	-	-	-	6.1

* Highest S/N ratios are shown for LC separated peptides; M denotes oxidation of methionine

A number of peptides were observed that were unique to one of the three techniques. For example, the tryptic peptide VGTRCCTKPESER was one of three peptides identified only by AP-MALDI at the 1 pmol level, whilst four unique peptides were identified by MALDI-TOF and eight were identified by ESI-qTOF. These unique peptides provide 7 %, 7 % and 9 % coverage of the BSA sequence for AP-MALDI-QIT, vacuum MALDI-TOF and ESI-qTOF techniques respectively following LC separation. Expressing the number of unique peptides identified as a percentage of sequence coverage for MALDI-TOF and AP-MALDI-QIT without LC pre-separation, shows that sequence coverage fell from 7 % to 3 % for MALDI-TOF, the corresponding coverage for AP-MALDI-QIT associated with unique peptides decreased from 7 % to 4 % compared to PMF data from LC separated peptides. Sequence coverage can therefore be enhanced significantly by combining results from more than one technique. For example, the greatest sequence coverage obtained for

1 pmol of BSA digest using a single technique was 50 % using LC/MALDI-TOF analysis, which was increased to 70 % by combining all three techniques.

3.4 Conclusions

The combination of capillary-LC with AP-MALDI-QIT mass spectrometry has been demonstrated for protein identification by tryptic peptide analysis. A data-dependant scan routine was used to generate MS and MS/MS spectra from the same sample spot in a single acquisition. A comparison of AP-MALDI-QIT with data obtained by vacuum MALDI-TOF and ESI-qTOF has been carried out, focusing on the performance of these different instrumental platforms using a single protein digest. The results showed higher numbers of matched peptides and sequence coverage for the latter techniques, but a similar ratio of peptides submitted to matched peptides for the database search using PMF data. For direct analysis of the digest without prior LC separation, the sequence coverage obtained for AP-MALDI-QIT MS was comparable with MALDI-TOF. Tandem mass spectrometry of AP-MALDI generated ions is shown to improve the confidence of assignment for peptide identification. The AP-MALDI-QIT, vacuum MALDI-TOF and ESI-qTOF mass spectrometric methods are complimentary with unique peptides observed by each technique. Sequence coverage is significantly improved through combining techniques. PMF data may be straightforward for the identification of a single protein digest, but complications may arise when dealing with mixtures resulting from the digestion of several proteins. AP-MALDI-QIT, with and without prior LC separation, shows potential for protein identification through the analysis of tryptic peptides by combined peptide mass fingerprinting and tandem mass spectrometry from a single MALDI experiment. For more complex samples, the benefits of prior LC separation are likely to become increasingly apparent, as demonstrated by many publications using vacuum MALDI ²⁶.

3.5 References

- 1) Beavis R.C., Chait B.T., *Proc. Nat. Acad. Sci, USA*, **87**, 6873-6877 (1990)
- 2) Zaluzec E.J., Gage D.A., Watson J.T., *Protein Express. Purifi.*, **6**, 109-123 (1995)
- 3) Caprioli R.M., Whaley B., Mock K.K., Cottrell J.S., *Techniques in Protein Chemistry II*, Villafrance, Ed. (Academic Press, New York), **48**, 497 (1991)
- 4) Gelpi E., *J. Chromatogr. A.*, **703**, 59-80 (1995)
- 5) Zhang B., McDonald C., Liang L., *Anal. Chem.*, **76**, 992-1001 (1992)
- 6) Gehrig P.M., Biemann K., *Peptide Res.*, **9**, 308-314 (1996)
- 7) Palm L., Burka L.T., Højrup P., Stevens R.D., Tomer K.B., *Int. J. BioChem. Cell Biol.*, **28**, 1319-1326 (1996)
- 8) Wei-Ling Tan S., Wong S-M., Manjunatha Kini. R., *J. Virol. Methods.*, **85**, 93-98 (2000)
- 9) Lochnit G., Geyer R., *Biomed. Chromatogr.*, **18**, 841-848 (2004)
- 10) Fung K., Askovic S., Basile F., Duncan M.W., *Proteomics*, 3121-3127 (2004)
- 11) Brombacher S., Owen S.J., *Anal. Bioanal. Chem.* **376**, 773-776 (2003)
- 12) Qin J., Fenyo D., Zhao Y., Hall W.W., Chao D.M., Wilson C.J., Young R.A., Chait B.T., *Anal. Chem.*, **69**, 3995-4001 (1997)
- 13) McCormack A.L, Schieltz D.M, Goode B., Yang S., Barnes G., Drubin D., Yates JR III., *Anal. Chem.*, **69**, 767-776 (1997)
- 14) Laiko V.V., Moyer S.C., Cotter R.J., *Anal. Chem.*, **72**, 5239-5243 (2000)
- 15) Laiko V.V., Baldwin M.A., Burlingame A.L., *Anal. Chem.*, **72**, 652-657 (2000)
- 16) Creaser C.S., Reynolds J.C., Hoteling A.J., Nichols W.F., Owens K.G., *Eur. J. Mass Spectrom.*, **9**, 33-44 (2003)
- 17) Keough T., Lacey M., Strife R., *Rapid Commun. Mass Spectrom.*, **15**, 2227-2239 (2001)
- 18) Creaser C.S, Reynolds J.C, Harvey D., *Rapid Commun. Mass Spectrom.*, **16**, 176-184 (2002)

- 19) Mehl J.T., Cummings J.J., Rohde E., Yates N., *Rapid Commun. Mass Spectrom.*, **17**, 1600-1610 (2003)
- 20) Tang N., Miller C.A., Meza J., *Proceedings of the Association of Biomolecular Resource Facilities Annual Meeting*, Savannah, GA, February 5-8, 2005
- 21) Schneider B.B, Lock C., Covey T.R., *J. Am. Soc. Mass Spectrom.*, **1**, 176-182 (2005)
- 22) Doroshenko V.M., Cotter R.J., *J. Mass Spectrom.*, **31**, 602-615 (1997)
- 23) Doroshenko V.M., Cornish T.J., Cotter R.J., *Rapid Commun. Mass Spectrom.*, **6**, 753-757 (1992)
- 24) ExPASy (ProtParam) available at <http://us.expasy.org/tools/protparam.html> [accessed on 25/04/05]
- 25) Green-Church K.B., Limbach P.A., *Anal. Chem.*, **70**, 5322-5325 (1998)
- 26) Lochnit G., Geyer R., *Biomed. Chromatogr.*, **18**, 841-848 (2004)

Chapter 4

Development and Evaluation of Methodology for the Analysis of Serum Peptides Using MALDI-TOF and AP-MALDI-QIT Mass Spectrometry

4.1 Introduction

Biomarkers, which have been described by Adkins as “proteins that undergo a change in concentration or state in association with a biological process or disease,” are extremely important factors for early diagnosis and monitoring progression of various diseases.¹ Mass spectrometry has been shown to be a powerful technique for biomarker identification and MALDI and SELDI mass spectrometry have both been applied to the detection of biomarker ions of intact proteins in sera and plasma.²⁻⁷

Although the majority of these top down approaches have been successful in generating proteomic profiles which distinguish between disease states, it is not possible to determine with confidence the identification of proteins directly from the masses of the biomarker ions. 2D PAGE has been used widely for the separation and comparison of complex protein mixtures, including serum and plasma.⁸⁻¹⁰ However, as previously discussed in Chapter 1, Section 1.4.2, the limitations of gel based methods include poor reproducibility, the requirement for large sample quantities, a limited dynamic range for proteins and the need to extract and digest the separated bands or spots for protein identification by mass spectrometry. The overall objective of the research described in this chapter was to overcome these limitations by the development of a bottom-up approach to identify peptide fragments as biomarkers and derive sequence information allowing the associated proteins to be identified.

Serum is a particularly useful medium for possible biomarker identification because it is easily obtained, has a high protein concentration (60-80 mg mL⁻¹) and many of the protein and peptide constituents are shed from cells and tissues.¹¹ For these reasons specific serum markers are used routinely to identify disease states, for example up-regulation of prostate specific antigen is used as a marker for prostate cancer.¹² Biomarkers are often low molecular weight (LMW) proteins secreted into the blood stream as a result of the disease

process.¹³ Some LMW serum proteins and peptides have been associated with cancer,¹⁴ diabetes¹⁵ and cardiovascular conditions.¹⁶

Large proteins such as albumin, transferrin, immunoglobulins, α -macroglobulin, haptoglobin and antitrypsin are the principal components of serum amounting to approximately 97 % of the total serum proteins, Figure 1.19.¹⁷ The dynamic range of serum proteins is in excess of 10 orders of magnitude posing a severe challenge to traditional proteomic methods for the detection of the low abundant analytes, such as possible biomarkers. The presence of highly abundant proteins makes detection of the smaller, less abundant proteins difficult.^{1, 18} Removal of the high abundant proteins is usually the first step in serum proteome analysis to reduce the complexity of the sample. A number of approaches have been reported for protein removal or depletion including affinity separation,^{1, 19, 20} protein precipitation,²¹⁻²² ultrafiltration,^{20, 23, 24} and chromatographic separation.²⁵

Affinity columns such as anti-human serum albumin antibody columns, which remove the most abundant proteins, are often used prior to mass spectrometric analysis.^{1, 19} Plasma has been separated using other chromatography methods such as ion exchange and reverse-phase liquid chromatography prior to MS analysis.²⁵ However, the major drawback of these techniques is the concomitant removal of potentially important LMW proteins and peptides from the sample as a result of association with other larger species.^{20, 26} Albumin, for instance, has been shown to act as a carrier and transport protein which binds many important species such as hormones, cytokines and lipoproteins.²⁷ The ideal depletion method to allow the analysis of LMW proteins and peptides would therefore remove the high molecular weight proteins of high abundance, whilst releasing the LMW species.

Protein precipitation with acetonitrile has been used successfully in a procedure for the determination of cefuroxime, a cephalosporin (a class of antibiotic), in human serum by LC-MS/MS. ²¹ Chertov *et al* ²² investigated precipitation of proteins and peptides from serum using organic solvents as a sample preparation technique prior to biomarker identification by mass spectrometry. The study compared the effect of different organic solvents including acetonitrile and methanol on serum protein precipitation. Addition of acetonitrile to serum led to the formation of a small volume of precipitate which was easily separated by centrifugation. However, when methanol was used as the precipitating agent a larger, less dense precipitate formed which was harder to remove from the supernatant. SDS-PAGE was used to examine the effect of acetonitrile precipitation on human serum, showing that low molecular weight proteins were retained in the supernatant after precipitation of the high molecular weight species. SELDI MS analysis showed differences between untreated and precipitated serum from the same mouse. Visual spectral differences were observed and ions were detected in the precipitated sample which could not be distinguished in the untreated sample. The study concluded that this was due to the depletion of the high molecular weight proteins which mask low abundant species. However, when mass spectra of untreated sera from tumour bearing and healthy mice were compared with protein precipitated samples from the same mice, some ions present in untreated sera were observed at lower intensities in the protein precipitated samples. This shows the need for caution when preparing samples using protein precipitation prior to MS analysis.

Centrifugal ultrafiltration methods which separate the LMW from high molecular weight proteins/peptides have also been reported in the literature. ^{20, 23} Ultrafiltration devices contain a hydrophilic membrane which is characterised by a nominal molecular weight cut-off (typically 3, 10, 30, 50 and 100 kDa). Molecules above this cut-off should be retained by the membrane, whilst molecules of lower mass are filtered through. These devices have

been used successfully to prepare plasma and sera samples for MALDI-TOF analysis for melanoma biomarker identification.²⁸ Samples were diluted in saline solution (NaCl) and filtered through the ultrafiltration membrane. Salts were removed *via* standard C₁₈ ZipTip clean-up prior to MS analysis. However, the association of LMW proteins with high molecular weight proteins can be a problem. Siegel *et al*²⁴ have used ultrafiltration to screen LMW compounds that are non-covalently bound to retained proteins. A recent investigation showed that protein precipitation is superior to ultrafiltration for the analysis of LMW serum proteins/peptides.²⁹ Although similar LC-MS spectra were obtained, the acetonitrile precipitated sample contained many more LMW species than the ultrafiltrate. Protein-protein interactions are disrupted by the addition of acetonitrile, causing LMW proteins to dissociate from carrier molecules allowing many more LMW species to be detected. Tirumalai *et al*³⁰ also demonstrated the use of acetonitrile protein precipitation to allow the detection of LMW proteins/peptides. This study examined the combination of both protein precipitation and ultrafiltration methods to remove high molecular weight proteins from serum, prior to SELDI-TOF MS analysis. Acetonitrile precipitation and subsequent ultracentrifugation was found to give the highest number of LMW species. SDS-PAGE and subsequent tryptic digest of the fractions was used to evaluate the success of high molecular weight protein removal. Although no peptides from albumin were identified, heavy chain immunoglobulin G (molecular weight 55 kDa) was still present after ultrafiltration. Other reports also show that ultrafiltration is ineffective in removing large species from plasma.³¹ Georgiou *et al* examined plasma samples prepared by ultrafiltration and analysed by 2D gel electrophoresis. Comparison of the filtrate and retentate showed little variation, with high and low molecular weight proteins/peptides present in both fractions, possibly due to the pore size distribution of the cut-off membranes used. On the whole samples analysed without ultrafiltration were more informative than those following ultrafiltration.

The specific aim of the research reported in this chapter was to develop a high-throughput strategy for the detection and identification of LMW serum peptide biomarker ions for melanoma (Chapter 5 Section 5.1) using MALDI-TOF and AP-MALDI/QIT MS. The method development included an investigation of sample clean-up by acetonitrile precipitation, ultrafiltration, ZipTip clean up and simple dilution for native peptide fragments and peptides derived by tryptic digestion of serum proteins. The reproducibility of the methods developed for both native and tryptic serum peptides have been evaluated. A serum aging study was conducted to assess the effects of sample handling conditions on the human serum proteome.

4.2 Experimental

4.2.1 Materials

Recrystallized matrix, α -cyano-4-hydroxycinnamic acid (CHCA), obtained from LaserBio Labs (Sophia-Antipolis Cedex, France) was used without further purification and prepared as a 10 mg mL^{-1} solution in 50 % acetonitrile/water + 0.1 % trifluoroacetic acid (TFA). Peptide Calibration Mix 4 (Proteomix) consisting of bradykinin fragment 1-5, angiotensin II, neurotensin, ACTH clip (18-39) and Insulin B-chain oxidised was used as received from Laser Bio Labs (Sophia-Antipolis Cedex, France) following manufactures instructions. All other reagents and chemicals were obtained from Fisher (Loughborough, Leicestershire, UK). Human serum samples were obtained from healthy individuals after consent and with Nottingham Trent University human ethical approval.

4.2.2 BCA Assay

A BCA assay was used to calculate the protein concentration of the serum sample.

BSA solutions were prepared in the range 1 mg mL^{-1} to 0.1 mg mL^{-1} in d.d. H_2O and aliquots dispensed ($25 \mu\text{L}$) in triplicate into a 96 well microtitre plate. Serum samples were diluted 1:200 and 1:300 in d.d. H_2O and Tris buffer was used as a blank solution with data being collected for triplicate aliquots of $25 \mu\text{L}$. The BCA working reagent was prepared by mixing bicinchonic acid reagent (7.84 mL) with a 4 % copper sulphate solution ($160 \mu\text{L}$). Aliquots ($200 \mu\text{L}$) were added to each well and incubated at $37 \text{ }^\circ\text{C}$ for 30 minutes whilst shaking. The absorbance of all solutions was measured at 570 nm in a plate reader (BioRad, Herts, UK). The concentration of the serum sample was calculated by plotting a standard curve, Section 4.3.1.

4.2.3 Method Development for the Analysis of Native Serum Peptides

4.2.3.1 Dilutions

Neat serum was diluted 1/2, 1/5, 1/10, 1/20, 1/30 and 1/40 with 0.1 % TFA in d.d. H₂O and analysed by MALDI-TOF MS without further purification. The series of diluted samples were also subjected to Zip-Tip clean up prior to analysis using the method detailed in Section 4.2.3.2. The MALDI targets were spotted using the dried droplet method; sample (1 µL) and matrix, α -cyano-4-hydroxycinnamic acid (1 µL) prepared as described in Section 4.2.1, were mixed on plate and allowed to air dry.

4.2.3.2 ZipTip Clean-up

A C₁₈ ZipTip (Millipore, Bedford, MA, USA) was pre-wet with 5 cycles of 50 % acetonitrile/water (10 µL) and equilibrated with 5 cycles of 0.1% TFA in d.d. H₂O (10 µL). Sample solutions (10 µL) (diluted serum or tryptic serum peptides detailed in Section 4.2.4) acidified to 0.1 % TFA were aspirated through the ZipTip for 20 cycles to allow binding of the proteins/peptides. 0.1 % TFA in d.d. H₂O was then cycled through the tip to wash the adsorbent and dispensed to waste. This wash step was repeated 5 times followed by elution in 80 % acetonitrile/water + 0.1 % TFA (2-4 µL). The eluant was aspirated 20 times without introducing air into the sample and dispensed through the ZipTip pipette tip. This ZipTip procedure was used in all subsequent experiments.

4.2.3.3 Ultrafiltration

For initial experiments neat serum (200 µL) was loaded onto an ultrafiltration device (Microcon YM-10, Millipore, Bedford, MA, USA) with a 10 kDa cut-off and centrifuged at 9000 rpm for 30 minutes. An aliquot (30 µL) of the LMW eluate was acidified with 1% trifluoroacetic acid in d.d. H₂O (15 µL) and was subjected to ZipTip clean up following the procedure described in Section 4.2.3.2. Samples were applied to the MALDI target plate using the thin layer method; ³² matrix solution (1 µL) was spotted onto the plate and

allowed to air dry followed by the addition of the analyte (1 μL) which was also air dried. In subsequent investigations, the serum (20 μL) was diluted 1 in 5 with ammonium bicarbonate (25 mM), loaded onto the microcon and spun at 9000 rpm for 10 minutes prior to ZipTip clean-up or direct deposition onto the target plate. Application of the sample to the target plate was by the dried droplet method described in Section 4.2.3.1. For all subsequent experiments the dried droplet method of preparation was used because improved AP-MALDI spectra were obtained in comparison to the thin layer method. Blank samples were prepared using ammonium bicarbonate (25 mM) in place of the neat or diluted serum.

4.2.3.4 Acetonitrile Precipitation

Neat serum (100 μL) was mixed with HPLC grade acetonitrile (100 μL), vortexed for 30 seconds and allowed to stand at room temperature for 30 minutes. Centrifugation was carried out at 9000 rpm for 10 minutes and the samples were applied to the target plate with, or without, ZipTip clean-up using the procedure outlined in Section 4.2.3.2. The precipitation and centrifugation procedure was repeated using fresh samples and aliquots of the filtrate (150 μL) were speed vacuumed (SVC-100H, Thermo Electron Corporation, Waltham, MA, USA) to near dryness. Samples were reconstituted to their original volume by addition of 0.1 % TFA, before being deposited onto the target with, or without, ZipTip clean-up. Triplicate samples were prepared by dilution of the serum with 0.1 % TFA in the ratios 1/10, 1/20, 1/30 and 1/40. Following dilution, the samples were combined with an equal volume of acetonitrile (25 μL) and subjected to the procedure described above. After centrifugation an aliquot (30 μL) of the supernatant was removed and subjected to lyophilization, resuspended and analysed by MALDI MS as above.

4.2.4 Method Development for the Analysis of Tryptic Serum Peptides

4.2.4.1 Proteolytic Digestion

Lyophilized trypsin gold (Promega, Southampton, UK) was reconstituted in acetic acid (50 mM) to a concentration of $1.0 \mu\text{g } \mu\text{L}^{-1}$ and stored in aliquots of $20 \mu\text{L}$ at $-80 \text{ }^\circ\text{C}$ until required. Serum ($25 \mu\text{L}$ of $60 \text{ pmol } \mu\text{L}^{-1}$), ammonium bicarbonate ($50 \mu\text{L}$ of 100 mM), trypsin ($2 \mu\text{L}$) and d.d. H_2O ($23 \mu\text{L}$) were mixed in an eppendorf and incubated at $37 \text{ }^\circ\text{C}$ overnight. Neat serum samples were also diluted with d.d. H_2O in the range $1/2 - 1/40$ and aliquots ($25 \mu\text{L}$) of the diluted sample were digested in the same manner. The reaction was quenched by the addition of TFA to give a final concentration of 0.1% TFA. The resultant peptides were purified by ZipTip (Section 4.2.3.2) or stored at $-80 \text{ }^\circ\text{C}$ until required for further experimentation.

4.2.4.2 Acetonitrile Precipitation Prior to Digestion

Serum was diluted with 0.1% TFA to $60 \text{ pmol } \mu\text{L}^{-1}$ based on the concentration of neat serum determined from the BCA assay (Section 4.2.2). Protein precipitation was carried out by the addition of an equal volume ($100 \mu\text{L}$) of acetonitrile to the diluted serum. The sample was vortexed and allowed to stand at room temperature for 30 minutes followed by centrifugation at 9000 rpm for 10 minutes. A $150 \mu\text{L}$ aliquot of the supernatant was removed, speed vacuumed to dryness and reconstituted in ammonium bicarbonate ($50 \mu\text{L}$ of 100 mM), d.d. H_2O ($23 \mu\text{L}$) and trypsin ($2 \mu\text{g}$). Digestion was carried out at $37 \text{ }^\circ\text{C}$ overnight and the resulting peptides were concentrated and de-salted by ZipTip clean-up (Section 4.2.3.2) and analysed by MALDI-TOF and AP-MALDI MS using optimised conditions described in Sections 4.2.8.1 and 4.2.8.2 respectively. The acetonitrile precipitation was repeated using serum diluted $1/10$ with 0.1% TFA and an aliquot ($25 \mu\text{L}$) was combined with acetonitrile ($25 \mu\text{L}$). Following vortexing for 30 seconds and centrifugation at 9000 rpm for 10 minutes, the supernatant ($30 \mu\text{L}$) was speed vacuumed to near dryness and resuspended in digestion buffer [ammonium bicarbonate ($50 \mu\text{L}$ of 100

mM), trypsin (2 μL), and d.d. H_2O (23 μL)] as described in the Section 4.2.4.1. TFA (25 μL 0.1 %) was taken through the same procedure as a blank. After the reaction was quenched the samples were either concentrated using the ZipTip procedure (Section 4.2.3.2) or analysed without further clean-up by MALDI-TOF and AP-MALDI-QIT.

4.2.4.3 ZipTip Clean-up Prior to Digestion

Serum was diluted 1/10 with 0.1 % TFA and the ZipTip procedure (Section 4.2.3.2) was carried out on an aliquot (25 μL) with the sample eluted in 4 μl of 80 % acetonitrile + 0.1 % TFA. The eluate was combined with ammonium bicarbonate (16.6 μL of 100 mM), d.d. H_2O (7.6 μL), trypsin (1.3 μL of 0.5 $\mu\text{g } \mu\text{L}^{-1}$) and incubated at 37 °C overnight. The reaction was quenched by the addition of TFA to give a final concentration of 0.1 % TFA, and the sample was analysed by MALDI mass spectrometry with, and without, ZipTip purification.

Blank and control samples were prepared alongside the serum digests as follows,

- (i) 0.1% TFA (25 μL) was subjected to ZipTip clean-up before and after tryptic digestion
- (ii) serum was diluted 1/10 with 0.1 % TFA (25 μL) and incubated overnight at 37 °C in the digestion buffer without the incorporation of trypsin, (iii) the incubation was repeated after ZipTip clean-up of the diluted serum and (iv) a bovine serum albumin (BSA) solution was made up to the same concentration as neat serum (1 nmol μL^{-1}), diluted 1/10 with 0.1 % TFA and digested as above. In all cases ZipTip clean-up (Section 4.2.3.2) was carried out immediately prior to spotting the sample on the MALDI target plate.

4.2.5 Optimised Methodology for the Analysis of Native Peptides

The optimised method for the analysis of native peptides used the acetonitrile precipitation method followed by ZipTip clean-up, Figure 4.1. The preliminary method described in Section 4.2.3.4 was repeated with the following variations. Acetonitrile (12.5 μL) was added in a 1:1 ratio to diluted serum (1/20 with 0.1 % TFA) and vortexed for 30 seconds.

After centrifugation at 9000 rpm for 10 minutes, an aliquot of the supernatant (15 μ L) was carefully removed, speed vacuumed to near dryness and reconstituted to its original volume in 0.1 % TFA. The sample was stored at -80 °C. ZipTip clean-up was performed on the thawed sample using the procedure described in Section 4.2.3.2. Following ZipTip clean-up, samples were spotted onto the MALDI target plate using the dried droplet method (Section 4.2.3.3, mixing of equal volumes of CHCA and sample (1 μ L) on plate) and spectra acquired *via* autoquality mode with 5 shots per profile (see Section 4.2.8.1). A total of 200 profiles were averaged per sample and no smoothing or background subtraction was applied to the collected data.

4.2.6 Optimised Methodology for the Analysis of Tryptic Serum Peptides

Tryptic peptides were prepared following the ZipTip prior to digestion method, described in Section 4.2.4.3. After digestion the samples were quenched and the ZipTip procedure was repeated with resultant peptides eluted in 80 % ACN + 0.1 % TFA (4 μ L). Samples were deposited onto the MALDI-TOF target plate using the dried droplet method (Section 4.2.3.3) and data acquired using the optimised autoquality MALDI-TOF parameters used for native peptides (see Section 4.2.8.1). The procedure is summarised in Figure 4.1.

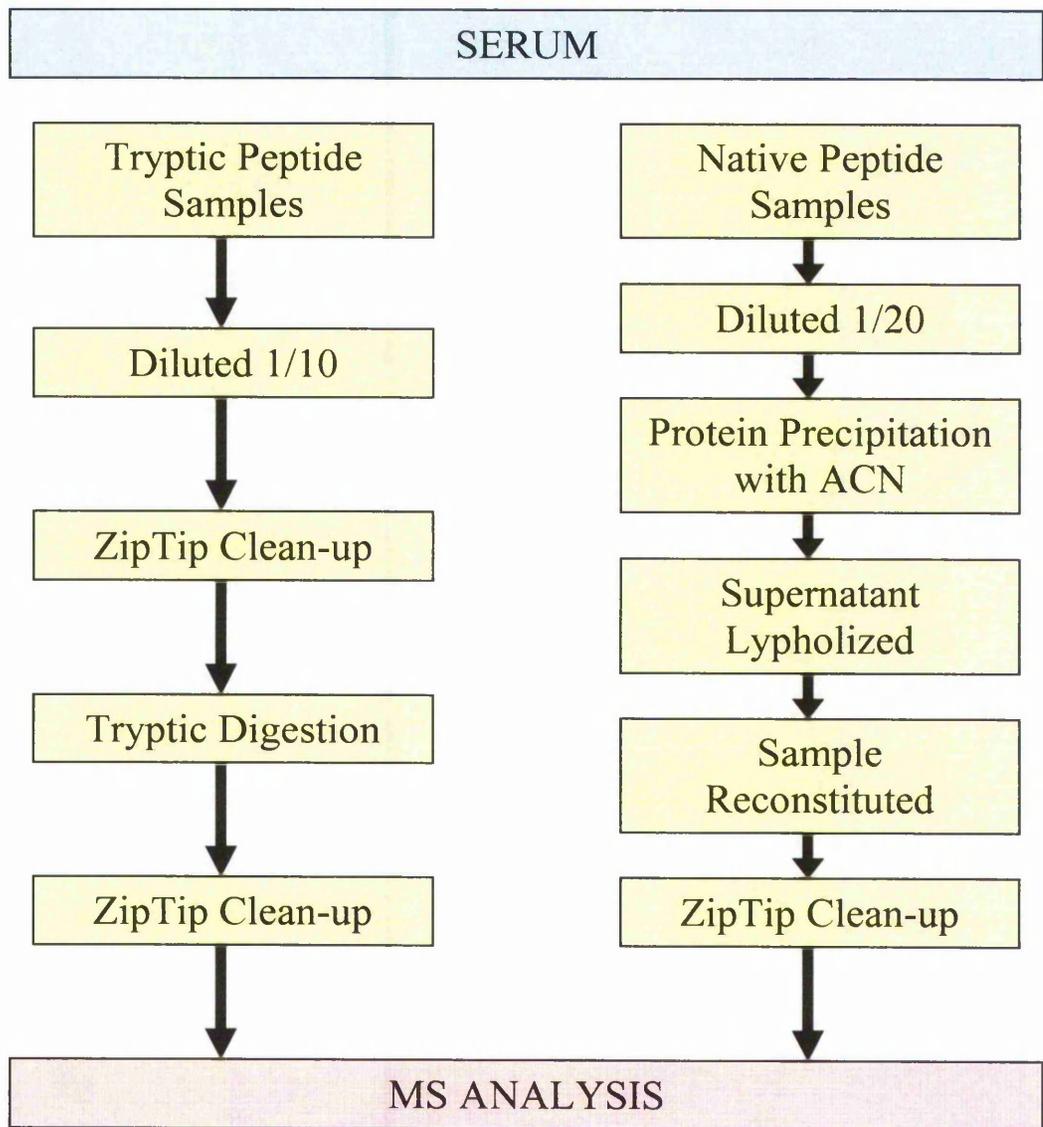


Figure 4.1: Developed methodology for the analysis of serum peptides

4.2.7 Serum Aging Study

Serum samples were collected from 2 male and 2 female donors, with aliquots being snap frozen in liquid nitrogen immediately following collection. Aliquots of fresh neat serum (1 mL) were exposed to the following treatments: (i) aging at room temperature for 3, 8, 24 and 96 hours, subsequently snap frozen in liquid nitrogen and then stored at -80°C until analysed, (ii) freeze-thaw cycling, where the samples were defrosted at room temperature and then refrozen to -80°C three times, and (iii) incubation at 4°C overnight in a fridge.

Great care was taken to prevent sample degradation whilst preparing the serum samples for analysis. The optimum native and tryptic sample methodologies described in Sections 4.2.5 and 4.2.6 respectively, were applied to the samples in duplicate and samples were deposited onto the target plate in a random manner. Serum samples were prepared in small batches and always kept on ice between steps. The data were submitted for artificial neural network (ANNs) analysis. This bioinformatics processing was kindly carried out by Graham Ball and Lee Lancashire at Loreus Limited, Nottingham, UK.

4.2.8 Mass Spectrometry

4.2.8.1 MALDI-TOF

Samples were analysed using an Axima CRF+ MALDI-TOF mass spectrometer (Kratos-Shimadzu, Manchester, UK) operated in reflectron mode (1-3500 Da). The instrument is equipped with a 337 nm N₂ laser with pulse duration of 3 ns. Calibration was carried out with peptide Calibration Mix 4 (Proteomix) (Laser Bio Labs, Sophia-Antipolis Cedex, France) using the close external calibration method, where there is 1 calibration spot to every 4 sample spots, Figure 4.2.

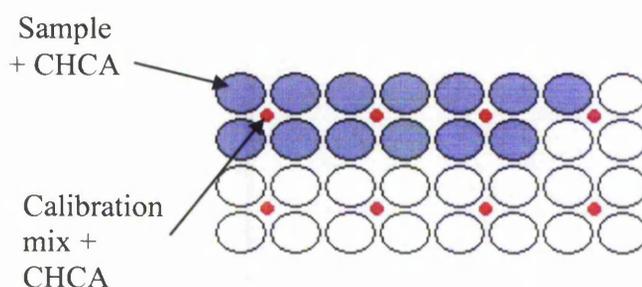


Figure 4.2: Schematic of the close external calibration method.

Optimisation of MALDI-TOF parameters were carried out by comparing data obtained for manual and automated acquisition of 200 shots per sample. Using the IntelliMarque software (Kratos, Manchester, UK) varying methods for automated analysis were set up. The main configurations tested included (i) raster-only mode, whereby the number of

sampling sites within a given spot was calculated based on the dimensions of the raster and the stage moved in a TV raster pattern allowing all of the sample spot to be scanned and (ii) autoquality-mode, where the sample spot was pre-scanned to select sampling sites based upon the signal intensity of the pre-scan. Data was collected from the positions in the sample which gave the highest signals (sweet spots). Data was compared for 1 and 5 laser shots per sample profile, for both raster only and autoquality methods. For autoquality, the parameters were set as follows; the criteria to accept a profile consisted of a minimum intensity of 4 mV with a minimum resolution of 400. Default values of 5 and 50 % for the minimum S/N ratio and S/N percentage respectively were used. When the S/N ratio for the base peak decreased to 50 % of the first accepted profile for a given raster point, the next ranked raster point was used.

4.2.8.2 AP-MALDI

AP-MALDI experiments were performed on a LCQ ion trap mass spectrometer (ThermoFinnigan, San Jose, CA, USA) fitted with an AP-MALDI ion source (MassTechnologies, Burtonsville, MA, USA). A UV nitrogen laser ($\lambda = 337$ nm, 300 μ J per pulse, 4 ns pulse width), (Spectra Physics, Berks, UK) was interfaced to the AP-MALDI source. The control software developed in-house, detailed in Chapter 2, was used to control the sample stage. For all experiments, the ion trap automated gain control (AGC) was deactivated, the heated capillary inlet was maintained at 350 °C and the gate time was set to 500 ms. No sheath or auxillary gases were used and the potential applied to the target plate was set at 2.7 kV. MS/MS was typically performed using a collision energy of 35 %.

4.2.8.3 Target plates

Experimentation focused upon optimising the sample preparation for both native and tryptically digested serum samples. Initial experiments used an integrated approach where

the same sample spot was analysed by both MALDI-TOF and AP-MALDI-QIT. A sample plate from the Axima CFR+ MALDI-TOF spectrometer was machined to take the smaller sample plate from the AP-MALDI source shown in Figure 4.3. This allowed sample transfer between the MALDI-TOF and the LCQ ion trap mass spectrometer (ThermoFinnigan, San Jose, CA, USA) resulting in accurate mass measurements on the TOF and sequencing data from the same sample using the QIT. Method development was also carried out using the MALDI-TOF instrument with standard 384 well target plates, allowing automated autoquality parameters to be set on the mass spectrometer, which was not possible using the machined plate. Since AP-MALDI was less sensitive than MALDI-TOF, sample preparation was evaluated in terms of attaining high quality and high signal intensity, which would allow AP-MALDI MS/MS analysis to be performed if potential biomarkers were identified.

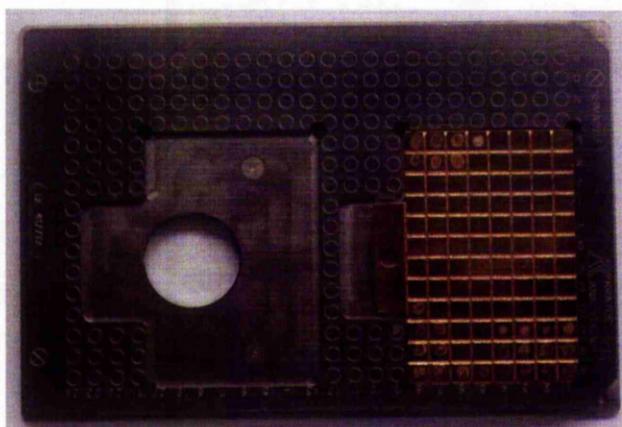


Figure 4.3: Adaptor plate to allow transfer of sample between MALDI-TOF (stainless steel target plate) and AP-MALDI-QIT (gold target plate).

4.2.8.4 Target Plate Cleaning Method

All target plates were wiped with a kimwipe soaked in methanol to remove surface material. The target was then sonicated for 10 minutes in each of the following solvents: methanol, d.d. H₂O, propan-2-ol, dichloromethane and hexane and then allowed to air dry prior to use.

4.3 Results and Discussion

The objective of the investigations described in this chapter was the development of a robust high throughput method for the analysis of LMW serum peptides. In order to achieve the sensitivity and reproducibility required, multi-stage clean-up procedures were developed for both native and tryptic peptides. This Section details the various methods explored, their efficiency and optimisation.

4.3.1 BCA Assay

The concentration of serum protein was ascertained by the BCA method which measures the reduction of Cu (II) to Cu (I) by bicinchoninic acid (BCA). A stable copper complex is formed by the reaction of two BCA molecules with one Cu (I), which is detectable at 570 nm. This method allows protein concentrations to be determined after plotting a calibration curve for BSA solutions of known concentration (described in Section 4.2.2), Figure 4.4. The concentration of serum used in the method development experimentations was calculated to be 66.7 mg mL^{-1} which is in agreement with literature values.³³ This concentration was used to prepare diluted samples of known protein concentrations for subsequent experiments.

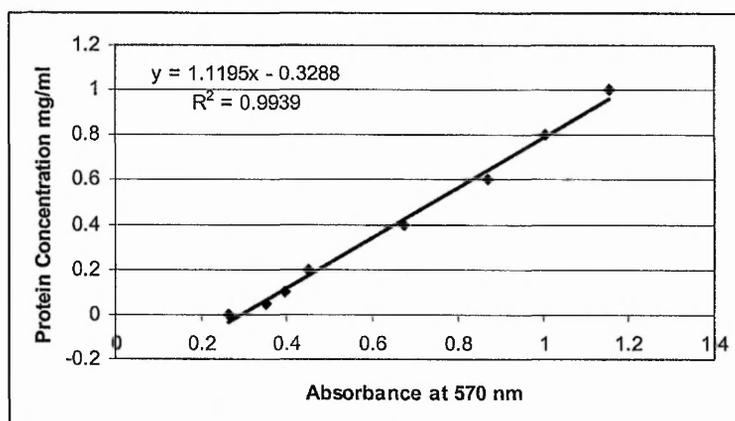


Figure 4.4: BCA protein assay standard curve

4.3.2 Analysis of Native Serum Peptides

4.3.2.1 Dilutions and ZipTip Clean-up

The effect of dilution and ZipTip clean-up on mass spectral quality was examined for native serum peptides. Serum was diluted 1/2, 1/5, 1/10, 1/20, 1/30 and 1/40 with 0.1% TFA and either subjected to ZipTip clean-up or directly spotted onto the sample plate. Analysis by MALDI-TOF showed that all samples benefited from the ZipTip procedure. ZipTip clean-up removes salts which can reduce the MALDI ionisation efficiency,³⁴ and concentrates the sample leading to improved spectral quality and S/N ratio. Samples that were directly spotted onto the plate produced poor quality spectra compared to samples subjected to ZipTip clean-up. Peptide peaks were undistinguishable from the background above m/z 1000 for samples directly spotted onto the plate, which can be seen for 1/20 dilution of serum in Figure 4.5.

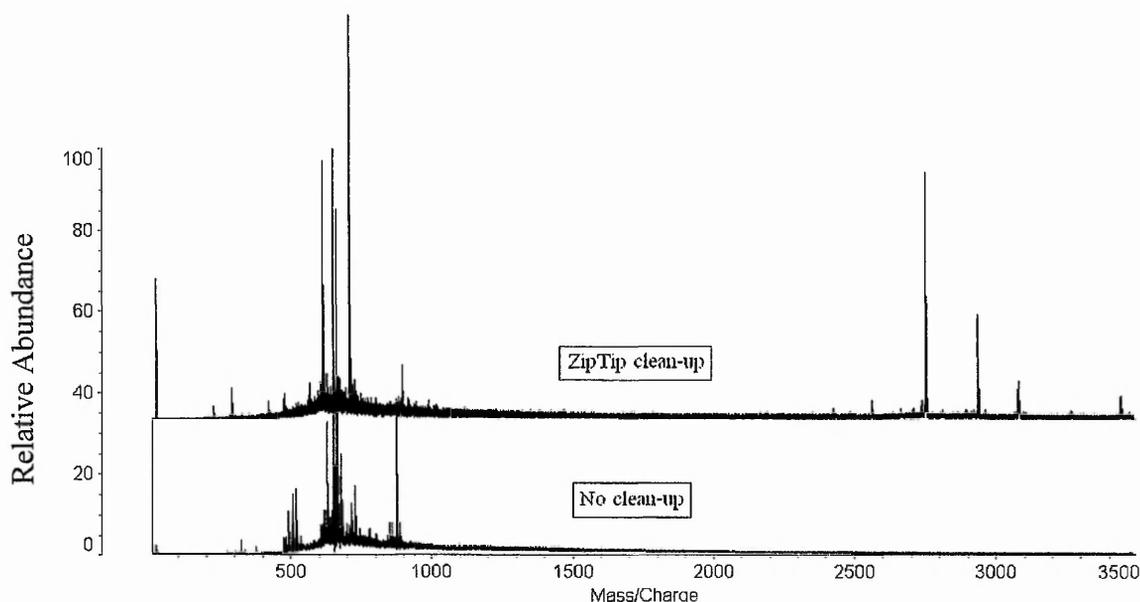


Figure 4.5: MALDI-TOF mass spectra obtained for serum samples diluted 1/20 with and without ZipTip clean-up prior to analysis.

However, following purification by ZipTip, the peptide peaks in the m/z region 1000-3500 were prominent for all serum dilutions in the range 1/10-1/40. Mass spectra presented in figure 4.6 show that no prominent peaks are observed in the m/z range 1000-3500 for the

1/2 and 1/5 serum dilutions. Higher concentrations may cause the ZipTip stationary phase to become saturated with bound proteins. The larger, more abundant serum proteins have a higher binding affinity and therefore would dominate the surface,³⁵ preventing smaller peptides from binding which would explain why few LMW species were seen in the mass spectra between m/z 1000-3500. Data were not acquired above 3500 Da because of the limited m/z range of the QIT (up to m/z 2000 or m/z 4000 in extended mass range), which was used for subsequent analysis.

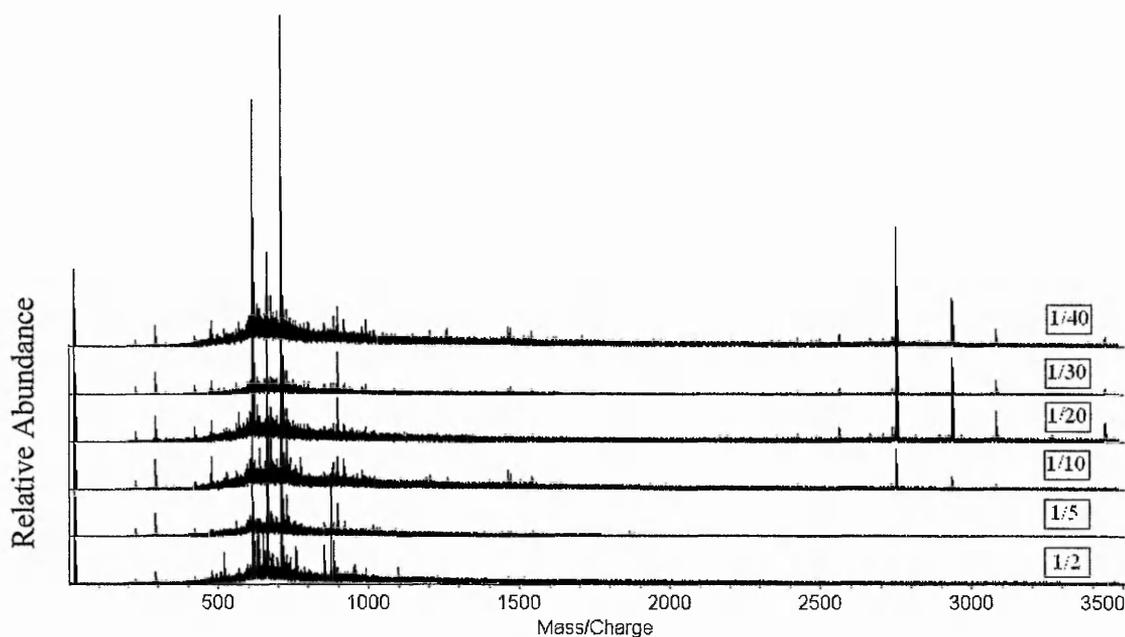


Figure 4.6: MALDI –TOF spectra obtained for diluted serum samples with subsequent ZipTip clean-up

4.3.2.2 Ultrafiltration

Ultrafiltration devices have been reported to reduce sample complexity by the removal of high molecular weight serum proteins allowing analysis of the lesser abundant serum species.^{28, 29, 36} As with the dilution experiments discussed previously in Section 4.3.2.1, ZipTip clean-up after ultrafiltration was necessary for all samples. Without ZipTip clean-

up it was not possible to obtain a mass spectrum. Others have also reported poor quality spectra when processing serum with ultrafiltration devices without ZipTip clean-up.³⁷

Initially, undiluted serum was loaded onto the microcon filter and this resulted in large volumes of serum being required to yield even small amounts of LMW filtrate. This was because of the high viscosity of the serum, which led to blockage of the molecular weight cut-off membrane. Dilution of the serum with ammonium bicarbonate prior to ultrafiltration led to a greater volume of LMW filtrate being obtained, Table 4.1. As the dilution factor was increased the majority of sample volume would pass through the membrane without blocking. Table 4.1 shows that a higher filtrate volume yield percentage was obtained with increasing dilution.

Table 4.1: Serum volumes before and after ultrafiltration

Sample Volume Loaded* / μl	Serum Dilution Factor	Filtrate (< 10 kDa) Volume / μl	Filtrate volume yield / %	Retentate (> 10 kDa) Volume / μl	Retentate Volume yield / %
40	1 / 2	2	5	38	95
50	1 / 2.5	7	14	43	86
100	1 / 5	24	24	76	76
200	1 / 10	106	53	94	47
400	1 / 20	252	63	148	37

* Constant volume of neat serum (20 μl) diluted with varying volumes of ammonium bicarbonate solution, as per dilution factor.

Figure 4.7 shows that there was a linear relationship between the volume loaded onto the microcon device and the LMW volume obtained. Therefore diluting serum with ammonium bicarbonate allowed a lower volume of neat sample to be used and a greater volume of LMW filtrate to be collected.

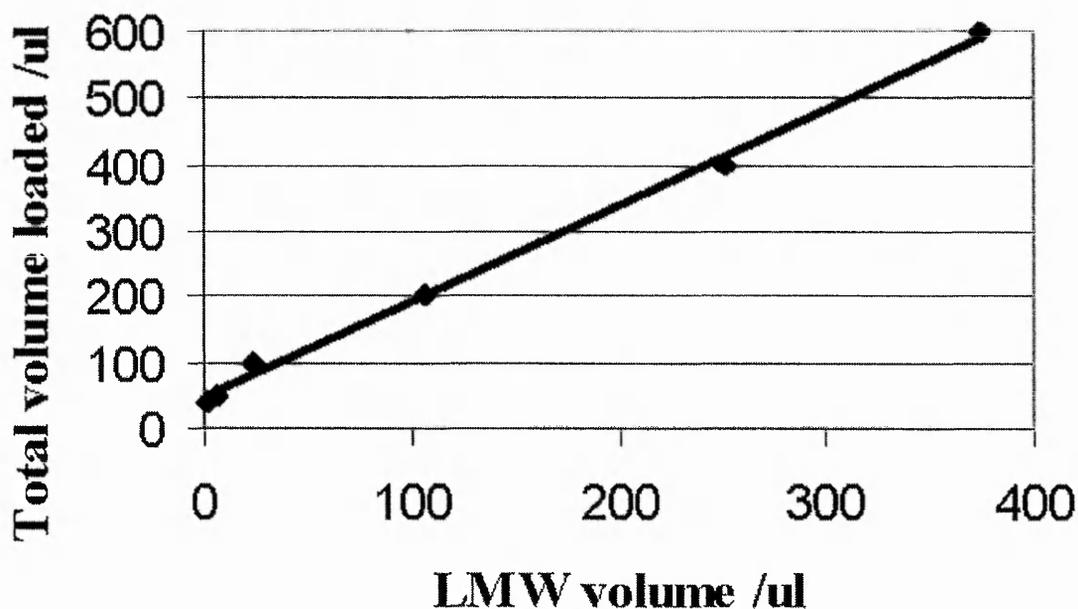


Figure 4.7: Volume of LMW filtrate obtained for varying total volumes loaded

Although dilution allows more LMW fraction to be obtained, which is particularly important for biological sample analysis (often only very low volumes of neat material are available) sampling times are increased due to the larger centrifugation cycle time required. A balance must be achieved between amount of sample obtained and sample throughput.

A further problem with ultrafiltration was the occurrence of contamination peaks in the resulting mass spectra when ammonium bicarbonate was passed through the cut-off membrane as a blank. An example of an AP-MALDI/QIT spectrum for a blank sample passed through the microcon device followed by ZipTip clean-up is shown in Figure 4.8. Contamination peaks were not observed unless the filtrate was subjected to ZipTip clean-up. However, examination of blanks passed through a ZipTip did not show this contamination, indicating that substances may have leached out of the microcon and were then concentrated by the ZipTip clean-up to a level where they could be detected by MS. Several attempts were made to remove the contamination, including spinning solvents through the filter prior to use and soaking the devices overnight in various concentrations of acetonitrile and TFA. Neither approach removed the problem. Due to the high intensity

of the unknown contaminants the method was deemed unsuitable and unreliable for biomarker identification.

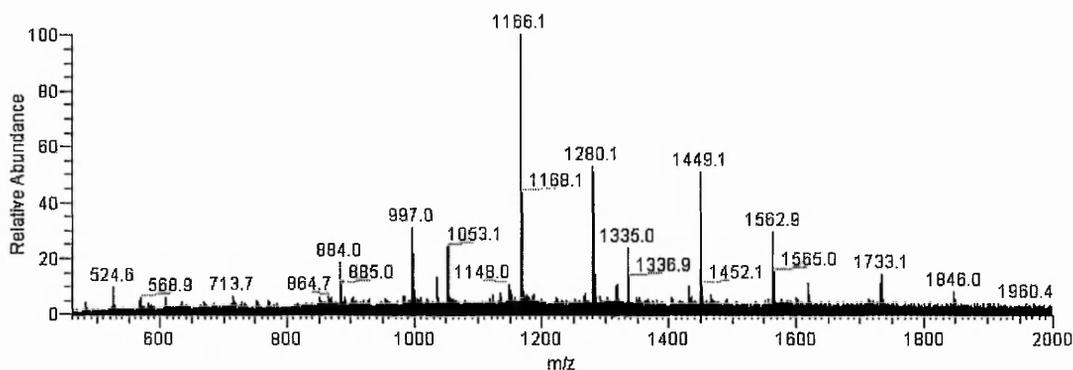


Figure 4.8: Solvent blank passed through the microcon followed by ZipTip clean-up and analysis by AP-MALDI/QIT

4.3.2.3 Acetonitrile Precipitation

Acetonitrile (ACN) is the commonly used precipitant for the removal of high molecular weight serum proteins.^{22, 38-40} The addition of organic solvent decreases the dielectric constant of the protein solution, which increases the attraction between charged molecules and aids electrostatic protein interactions. Organic solvent also displaces the water molecules surrounding the hydrophobic regions of the protein leading to a reduction in hydrophobic interactions between proteins. The electrostatic interactions become predominant, leading to protein aggregation. The effectiveness of protein removal from human serum and the reproducibility of varying precipitants has been reported.⁴¹ A 93 % protein precipitation efficiency with a relative standard deviation of 0.6 % was observed for ACN.

Preliminary experiments reported in this thesis followed literature methods, where equal volumes of neat serum and ACN were mixed. It has been suggested that albumin, the most abundant protein in serum, acts as a 'molecular sponge' interacting with LMW proteins and peptides and therefore retaining them in the precipitate. However, Merrell *et al*²⁹ demonstrated that on addition of ACN, serum proteins are denatured and the association

between the LMW species and larger species is disrupted, allowing the smaller proteins/peptides to remain in solution.

The effect of dilution prior to protein precipitation was also examined to reduce the volumes of serum required. Serum was diluted in 0.1 % TFA in the ratio 1/10, 1/20, 1/30, 1/40 and an aliquot (25 μ L) was mixed with ACN as described in Section 4.2.3.4. After precipitation, analysis of the supernatant by MALDI-TOF showed that ZipTip clean-up was essential (as with earlier findings in Sections 4.3.2.1 and 4.3.2.2) and significant peptide peaks were only observed in the m/z region 1000-3500 Da, for dilutions greater than 1/20, Figure 4.9. The 1/20 dilution gave higher signal intensities and increased S/N ratio compared with the other dilution factors

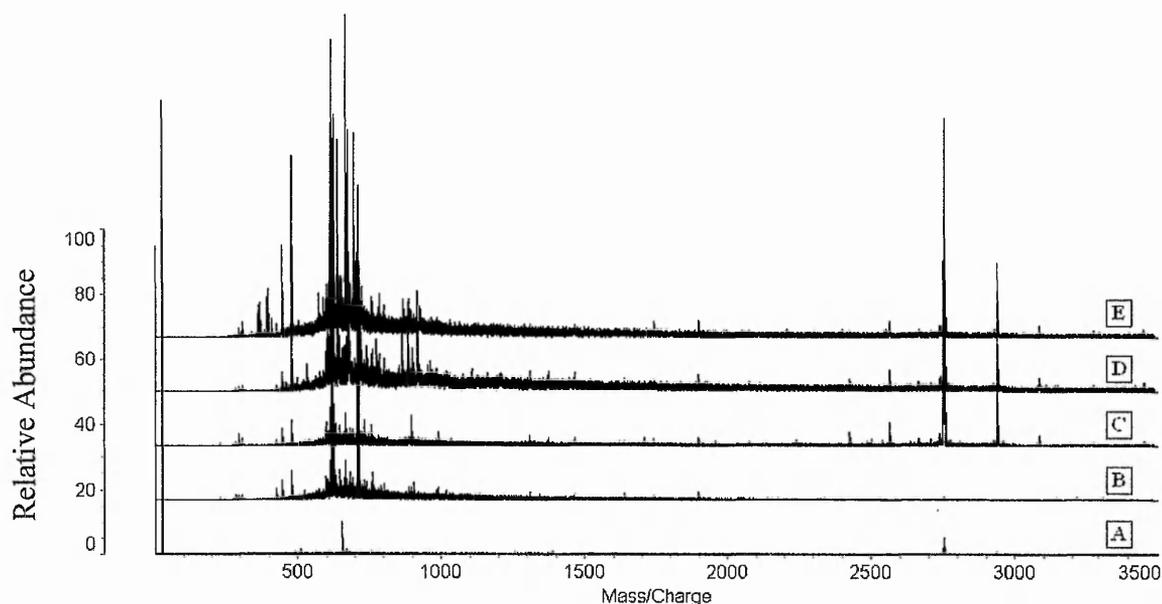


Figure 4.9: MALDI-TOF spectra of supernatant resulting from protein precipitation of diluted serum samples, where B, C, D and E correspond to 1/10, 1/20, 1/30, 1/40 fold dilution respectively followed by ZipTip clean-up. A represents 1/10 dilution with no ZipTip clean-up.

This data set was compared with results obtained following protein precipitation and the subsequent removal of the solvent from the supernatant by speed vacuum centrifugation and re-suspension of the sample in 0.1 % TFA, Figure 4.10. The resulting mass spectra again showed the need for ZipTip concentration and de-salting prior to analysis. It has been reported that peptides concentrated by vacuum centrifugation show poor recovery.⁴²

However, Figure 4.10, shows that the lyophilisation step greatly enhanced the number of peptide peaks in the resulting spectra. This is possibly because the speed vacuum centrifugation not only concentrated the sample, but removed the ACN allowing the sample to be resuspended in TFA prior to ZipTip clean-up improving peptide retention on the C₁₈ adsorbent. Decreasing the sample volume was also advantageous, because the time taken to remove the solvent from the sample was significantly reduced. No detrimental effects were observed by halving the sample volume from 25 μ L to 12.5 μ L, reducing the amount of supernatant which was lyophilized.

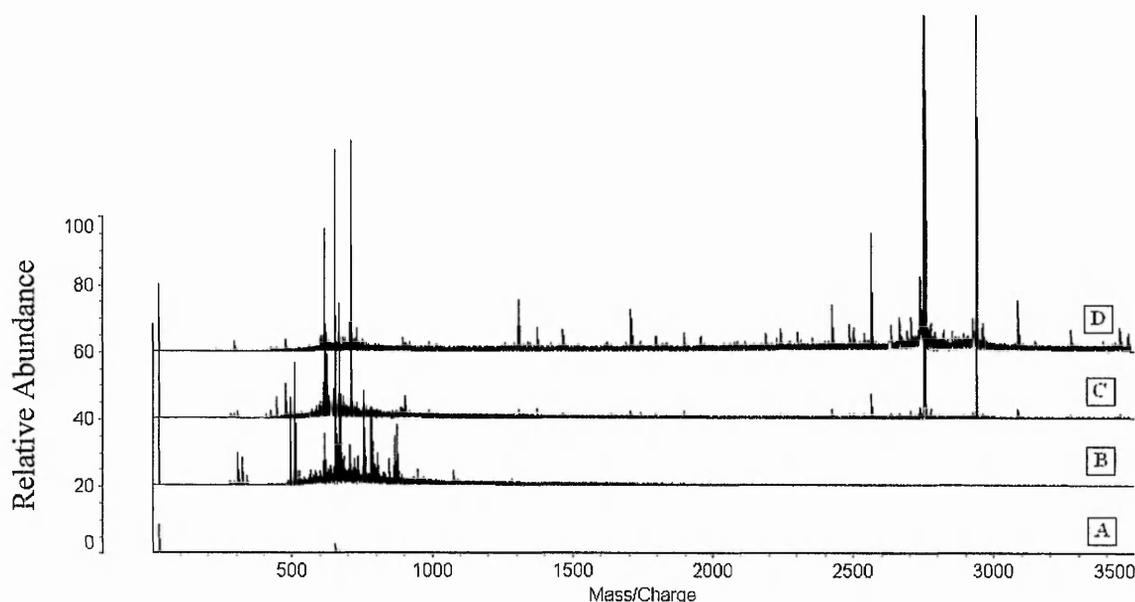


Figure 4.10: MALDI-TOF spectra obtained for (A) a 1/20 dilution of serum followed by protein precipitation with no ZipTip clean-up, (B) a 1/20 dilution of serum followed by protein precipitation and speed vacuumed to near dryness, reconstituted in 0.1% TFA and analysed with no ZipTip clean-up, (C) a 1/20 dilution of serum followed by protein precipitated and ZipTip clean-up (D) a 1/20 dilution of serum followed by protein precipitated, lyophilisation, resuspension and subjected to ZipTip clean-up.

The incorporation of a ZipTip clean-up after ACN precipitation in the developed methodology for native serum peptides described in this thesis meant that it was not possible to ascertain whether the ACN precipitated high mass proteins, whilst releasing LMW peptides as described by Merrell.²⁹ A direct comparison could not be made between pre- and post-precipitated samples, since the subsequent solid phase extraction depletes the

high abundant proteins. (discussed later in Section 4.3.3.3). However, precipitated samples without further ZipTip clean-up were analysed by MALDI-TOF in linear mode (1-70 kDa), and compared to un-treated serum, Figure 4.11. Results show that albumin (66 kDa) is present in both samples with similar intensities. However, the broad peak at m/z 50,000 is not observed in the precipitated sample indicating that the protein precipitation method used in this work is capable of removing some high mass proteins^{21, 22} and significantly enhancing the mass spectral quality in the peptide region below m/z 3500.

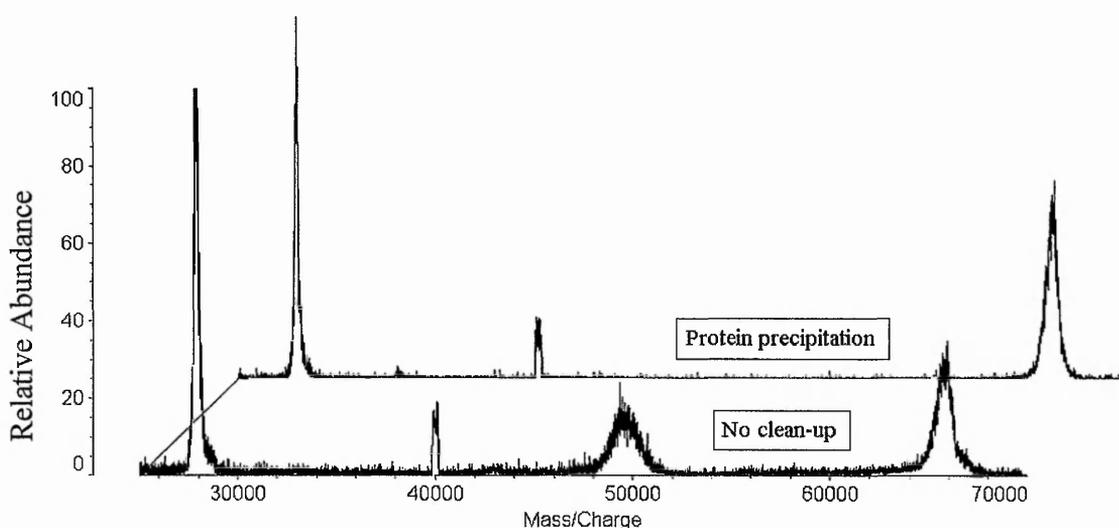


Figure 4.11: MALDI-TOF spectrum acquired in linear mode for serum diluted 1/20 with 0.1% TFA and analysed after protein precipitation and with no further clean-up. (analysis carried out by Dr Matharoo-Ball at Nottingham Trent University).

In the studies described above, various methods including dilution, ultrafiltration and acetonitrile precipitation have been investigated for the analysis of LMW native serum peptides. Results showed that all native serum peptides required ZipTip clean-up prior to mass spectrometric analysis (as discussed earlier in Sections 4.3.2.1, 4.3.2.2). The method which yielded the most consistent and optimal mass spectral data consisted of diluting serum 1/20 with 0.1 % TFA and mixing an aliquot with an equal volume of acetonitrile, lyophilization of the supernatant, reconstitution of the sample followed by ZipTip clean-up. Reproducibility of the methodology is discussed in Section 4.3.5. This optimised methodology (described in detail in Section 4.2.5) was used for all subsequent

investigations, including a serum aging study (Section 4.2.7) and analysis of melanoma serum samples detailed in Chapter 5.

4.3.3 Analysis of Tryptic Serum Peptides

As previously described in Chapter 1, Section 1.4.1 the 'top-down' proteomics approach deals with the analysis of intact proteins. The main limitation of this approach is that MS/MS data, and hence sequence information, cannot be obtained for large proteins. To overcome this limitation the 'bottom-up' method is widely used in combination with MALDI-TOF for protein analysis and identification.⁴³⁻⁴⁹ In this approach the proteome is usually separated by 2D gels and digested to give a complex peptide mixture, subsequent MS analysis of the resulting peptide mass fingerprint allows identification of individual proteins. Based on the prior success of protein identification from proteolytic digests of serum or plasma^{1, 30, 50, 51} this Section details the development and evaluation of a 'bottom-up' approach for serum protein analysis using MALDI-TOF MS without prior LC separation for high-throughput biomarker identification.

4.3.3.1 Proteolytic Digestion

Proteolytic digestion of the human serum proteome with trypsin yielded, as expected, a large number of peptide fragments and as with native serum analysis (Section 4.3.2) a ZipTip clean-up was required prior to MS analysis. This is consistent with the observation by Yates III *et al*,⁵² who reported the success rate of peptide mass fingerprinting for the analysis of proteins using MALDI-TOF was enhanced by sample clean-up by C₁₈ ZipTip, due to desalting and concentration of the sample.

The effect of diluting the serum in the range 1/2 to 1/40 with d.d. H₂O prior to proteolytic digestion on the MALDI mass spectral response was examined and the results are shown in Figure 4.12. A greater number of intense peptide ions were observed above m/z 850 in the MALDI-TOF spectra, when the serum was diluted by 1/10 or more, in comparison with the

lower dilution ratios. Ion abundance below m/z 850 changed little with dilution. The highest S/N ratio was observed at 1/20, but intensities of some higher mass peaks were greater at a 1/10 dilution. Therefore, the optimum dilution, based on a compromise between S/N ratio and the abundance of high mass ions was 1/10. All other concentrations resulted in poorer quality spectra with fewer visible peptides. In all subsequent analyses of tryptic peptides the serum was diluted 1/10 prior to digestion. To evaluate the efficiency of the tryptic digestion, a BSA control was prepared to the same concentration as the 1/10 diluted serum and analysed by MALDI-TOF. The resulting peptide mass fingerprint was submitted to Mascot and a significant score of 77 was obtained indicating that the digestion method was effective.

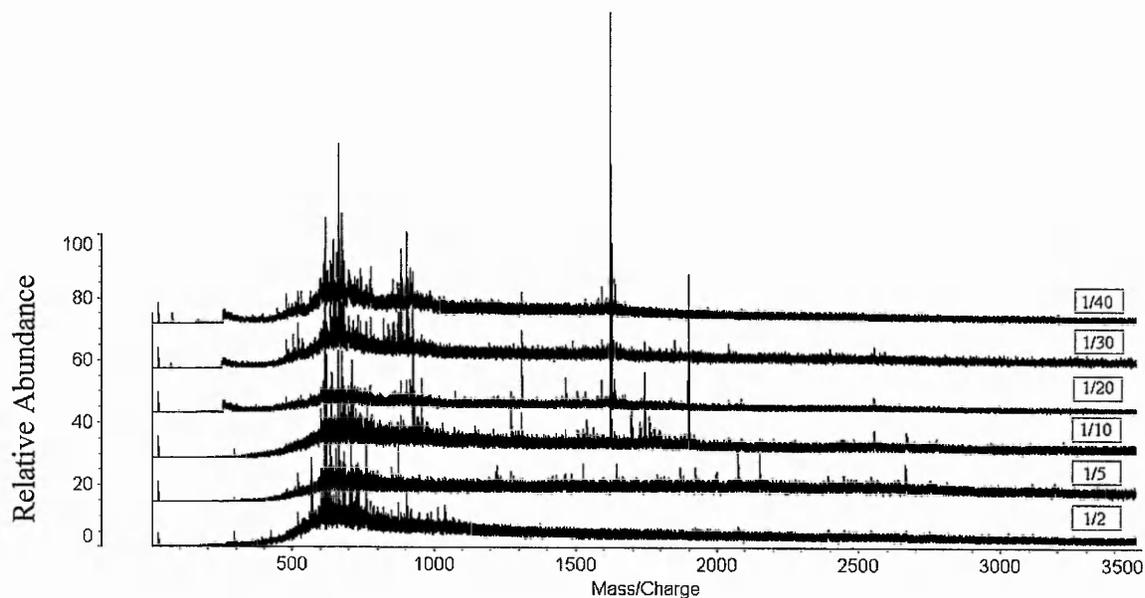


Figure 4.12: MALDI-TOF spectra for various serum dilutions subjected to proteolytic digestion followed by ZipTip clean-up.

4.3.3.2 Acetonitrile Precipitation Prior to Digestion

The aim of the investigation was to analyse the low abundant, LMW serum tryptic peptides. However, direct digestion of serum would lead to a peptide mixture dominated by the products of a small number of high abundant proteins such as albumin (Figure 1.19). Therefore it was desirable to remove the high abundant proteins prior to digestion.

Analysis of native serum proteins (discussed in Section 4.3.2.3) showed that although albumin levels were not affected other high molecular weight serum proteins were depleted by acetonitrile precipitation, Figure 4.11.

Methodology developed for the analysis of serum showed that a 1/10 dilution of serum prior to digestion, followed by ZipTip clean-up, resulted in the optimum mass spectral quality (discussed earlier in Section 4.3.3.1). A compromise was reached in terms of S/N and number of peptide fragments observed in the mass range m/z 850-3500. A method was developed similar to that used previously for tryptic peptides (Section 4.3.3.1) with the addition of an acetonitrile protein precipitation step as shown in Figure 4.13.

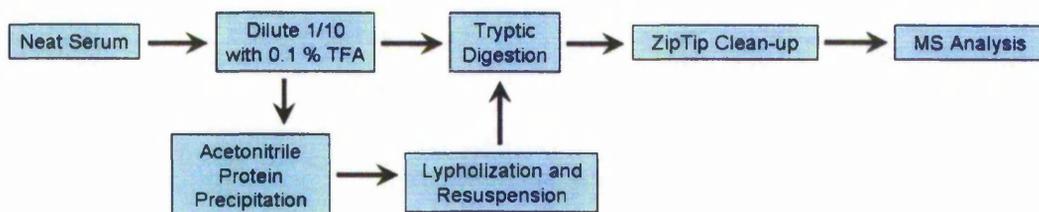


Figure 4.13: Schematic of experimental protocol developed for analysis of tryptic serum peptides.

The MALDI-TOF mass spectra obtained for tryptically digested serum, with and without preceding acetonitrile protein precipitation, are shown in Figures 4.14 and 4.15 respectively. These are comparable in terms of protein profiles, with all predominant peaks having high intensities in both spectra. This shows that the acetonitrile protein precipitation stage had little effect on the resulting tryptic peptidome mass spectra. This is in contrast to the results reported by others who showed that acetonitrile disrupted the association of LMW proteins/peptides with albumin, and a different mass spectral profile was observed for the protein precipitated samples in comparison with the un-precipitated samples.^{30, 53}

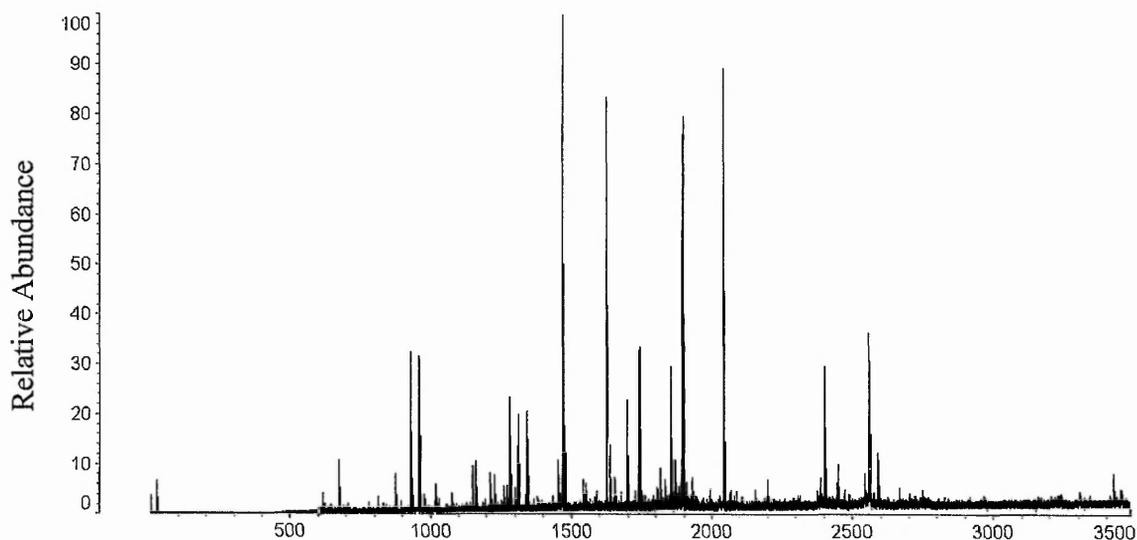


Figure 4.14: MALDI-TOF spectrum obtained for serum diluted (1/10) followed by acetonitrile protein precipitation, lyophilization, resuspension and subsequent digestion, followed by ZipTip clean-up.

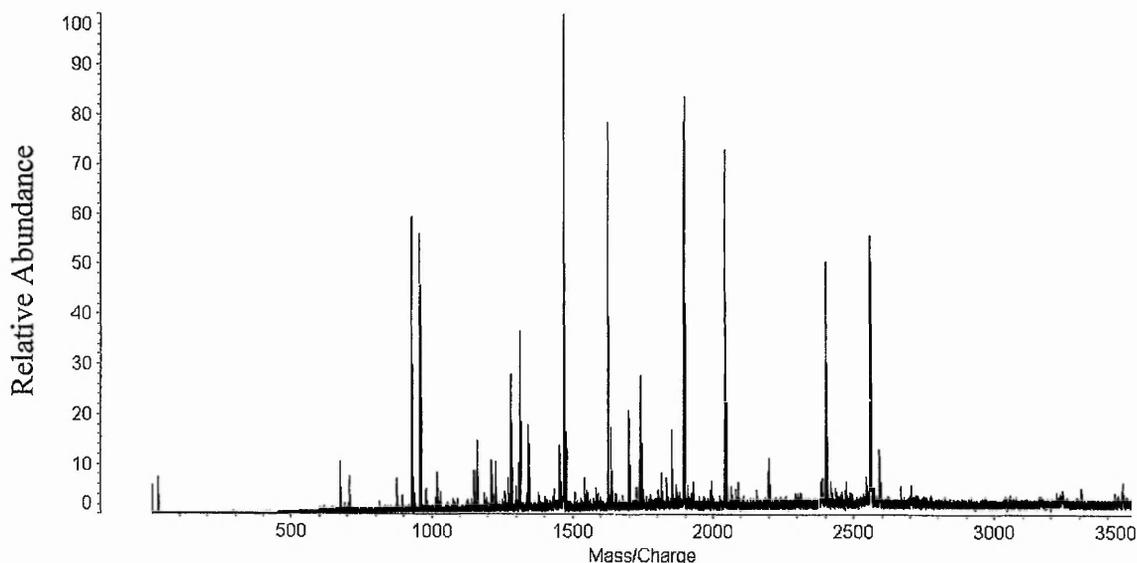


Figure 4.15: MALDI-TOF spectrum obtained for serum diluted (1/10) followed by tryptic digestion and ZipTip clean-up.

Following digestion, the peptides were analysed by AP-MALDI-QIT and MALDI-TOF with, and without ZipTip clean-up, to compare the spectra obtained by the two techniques. Figures 4.16 and 4.17 show the resulting mass spectra in the range $< m/z$ 2000, the normal mass limit for the LCQ mass spectrometer.

Similar tryptic peptide peaks, such as m/z 1899, 1623, 1467 and 960 are prominent in both AP-MALDI and MALDI-TOF spectra, although the intensities vary between the two techniques, Figures 4.16 and 4.17.

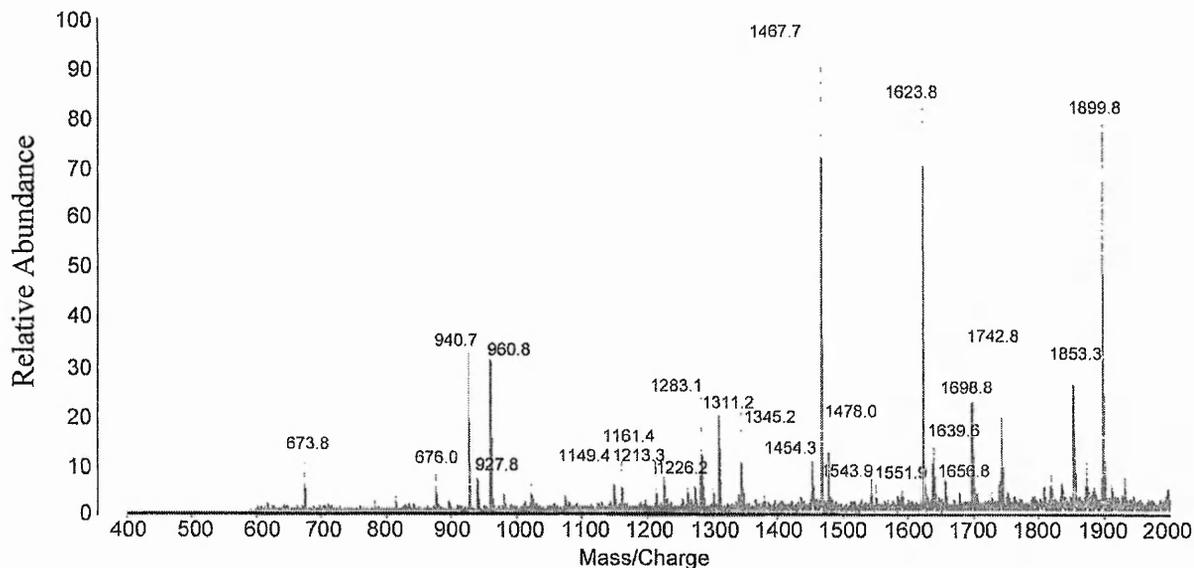


Figure 4.16: MALDI-TOF spectrum obtained for diluted serum (1/10) subjected to acetonitrile protein precipitation, lyophilization, resuspension followed by tryptic digestion and ZipTip clean-up. m

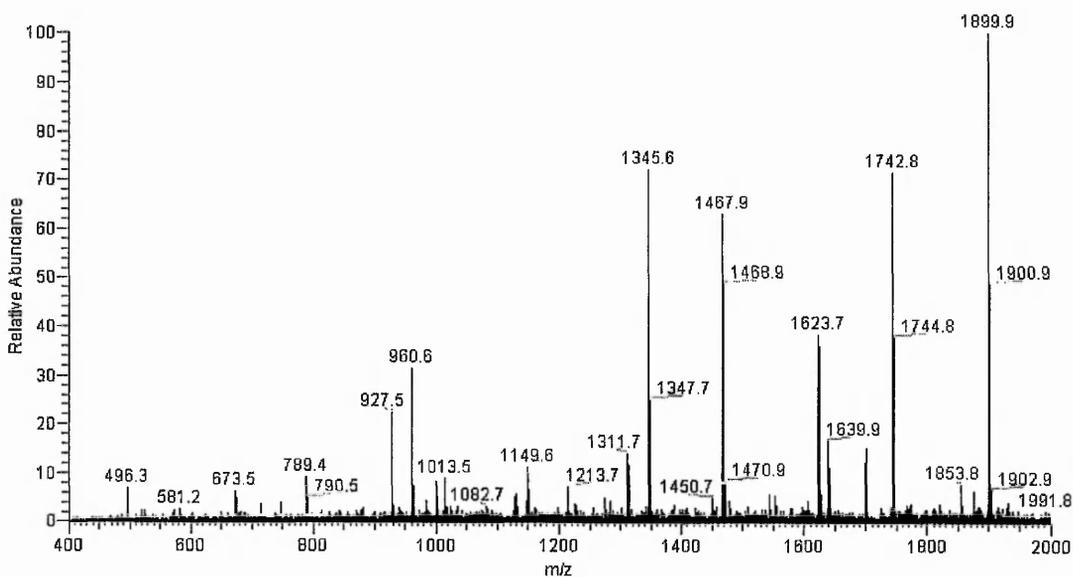


Figure 4.17: AP-MALDI-QIT spectrum obtained for diluted serum (1/10) subjected to acetonitrile protein precipitation, lyophilization, resuspension followed by tryptic digestion and ZipTip clean-up.

Peptide mass fingerprint data from both mass spectrometric techniques were searched against the MSBD database, using Mascot. A peptide mass tolerance of 0.4 Da was used. Mascot identified the main component of the PMF to be peptides derived from serum albumin chain A for both AP-MALDI-QIT and MALDI-TOF data. Mascot scores are shown in Table 4.2.

Table 4.2: Summary of Mascot results obtained for PMF data matched to serum albumin chain A.

	MALDI-TOF	AP-MALDI-QIT
Mascot Score	103	147
Number of peptides matched	10	11
Sequence Coverage / %	13	14

The peptides observed in the PMF were subjected to tandem mass spectrometry using AP-MALDI-QIT. A typical mass spectrum for the product ions of m/z 1467.9, a prominent ion in the PMF spectrum, is shown in Figure 4.18. MS/MS Mascot searching using a peptide tolerance of 500 ppm and an MS/MS tolerance of 0.3 Da, established that many of the tryptic peptide b_n and y_n fragment ions were derived from serum albumin chain A. These database searches show that despite the introduction of a protein precipitation stage using acetonitrile, the MALDI mass spectrum are still dominated by high abundant proteins.

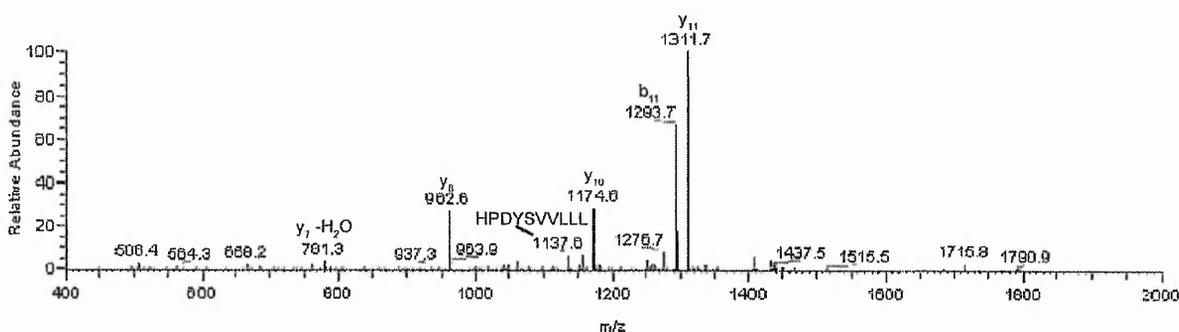


Figure 4.18: AP-MALDI-QIT MS/MS product ion spectrum of RHPDYSVLLLLR (m/z 1467.9) derived from serum albumin chain A.

4.3.3.3 ZipTip Clean-up Prior to Digestion

Due to the unsuitability of the acetonitrile protein precipitation method for the analysis of tryptic peptides, an alternative strategy for the removal of high abundant proteins was

investigated employing C₁₈ ZipTip clean-up before and after digestion. Serum was diluted 1/10 with 0.1 % TFA and then subjected to a ZipTip clean-up. Since both high and low molecular weight proteins bind to the C₁₈ stationary phase one might expect both to be eluted with ACN. However, a proportion of the larger species that generally bind with higher affinity to hydrophobic surfaces, may be retained by the tip. Elution of the low molecular weight peptides was achieved by cycling organic solvent (80 % ACN + 0.1 % TFA) through the tip and the eluant was combined with digestion buffer. The resulting tryptic peptides were then either deposited directly onto the MALDI-TOF target plate or desalted and concentrated by a further ZipTip clean-up stage. A ZipTip clean-up was required before samples were deposited onto the MALDI target in order to generate high quality spectra. As with acetonitrile precipitated samples, many peaks are observed in the resulting spectra for samples prepared using a ZipTip clean-up before and after proteolytic digestion as shown in Figure 4.19. Figure 4.20 compares the spectra obtained when acetonitrile precipitation and ZipTip clean-up were used prior to digestion. A large number of peaks associated with serum albumin are evident when using acetonitrile precipitation prior to digestion (Figure 4.20A). However, when ZipTip clean-up was used before and after digestion, many of these albumin peaks are not observed and those that are present are of lower intensity (Figure 4.20B). This shows that ZipTip clean-up depletes large, highly abundant serum proteins, allowing smaller protein/peptide ions to be observed at higher relative intensity.

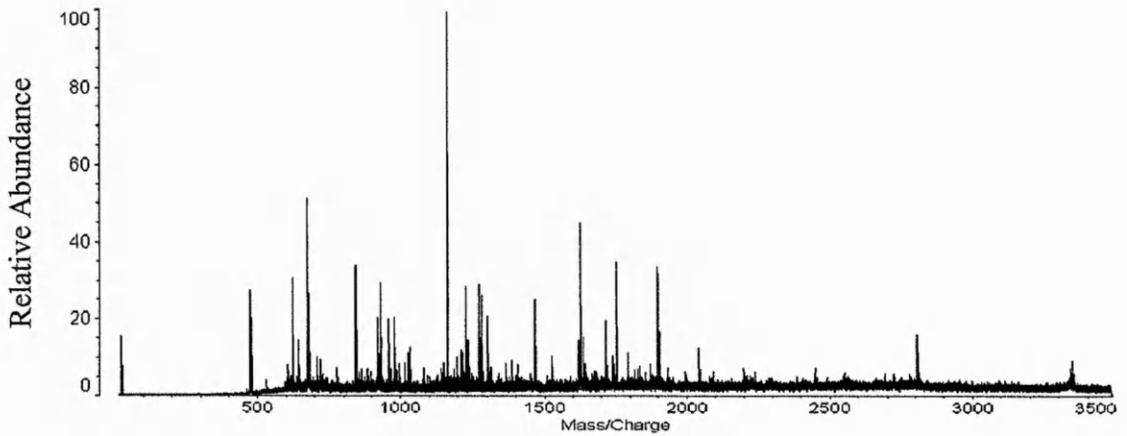


Figure 4.19: MALDI-TOF spectrum obtained for the analysis of serum samples prepared by ZipTip clean-up before digestion and subsequent ZipTip clean-up of resulting tryptic peptides.

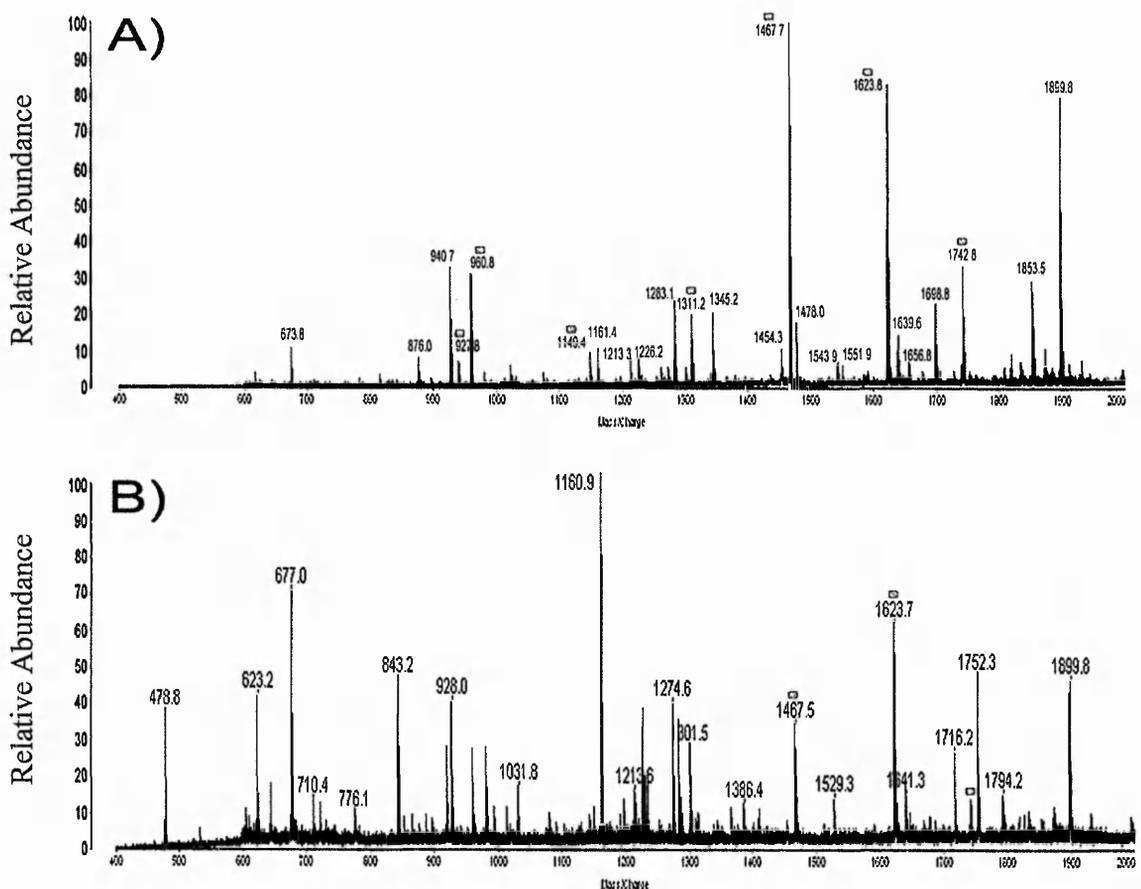


Figure 4.20: A) A MALDI-TOF mass spectrum for diluted serum (1/10) subjected to acetonitrile precipitation followed by proteolytic digestion and ZipTip clean-up and B) MALDI-TOF spectrum for diluted serum (1/10) subjected to ZipTip clean-up before and after digestion. Red markers indicate peptides associated with serum albumin chain A.

No autolysis of trypsin occurred when a 0.1 % TFA blank was subjected to ZipTip clean-up and the eluate was incubated overnight with digestion buffer, followed by ZipTip clean-

up prior to MALDI analysis. Samples were also prepared by serum dilution to 1/10 and incubation at 37 °C overnight in digestion buffer without the addition of trypsin. A second set of samples were prepared in the same manner but with the addition of a ZipTip clean-up step prior to digestion. The data obtained from these experiments showed that the ions observed in the spectra were generated from the serum samples and not from other sources, e.g. contaminants from trypsin or ZipTip procedures.

The depletion of high abundant proteins using a C₁₈ ZipTip prior to tryptic digestion with the resulting tryptic peptides concentrated and desalted by subsequent ZipTip clean-up was selected for all further analyses (Section 4.3.5, 4.3.6 and 5.3.2). This method was chosen because optimum results in terms of reproducibility (discussed in Section 4.3.5) and spectral quality were observed in comparison with the other methods evaluated.

4.3.4 Optimisation of Mass Spectrometry for the Analysis of Serum Peptides

The aim of the project was to develop a high-throughput MALDI-based method for the analysis of serum samples. If confidence is to be placed in the bioinformatics results for biomarker identification a large number of samples are required, making it unrealistic to rely on a manual acquisition method. Optimisation and automation of the mass spectrometric conditions was therefore carried out in parallel with the development of the sample preparation methodology. Comparison of results obtained for manual methods with automation, showed that high quality spectra could be generated by both approaches. The automated analysis was not only rapid but also very efficient and improved results were obtained, compared to manual acquisition once optimisation of the parameters had been carried out.

Optimisation of mass spectrometric conditions was evaluated by analysing native serum peptides prepared by acetonitrile precipitation, followed by lyophilisation of the supernatant

and subsequent ZipTip clean-up (Section 4.2.5). The effect of altering the raster size for automated acquisition was investigated first. Two different raster settings were examined, namely 2000 and 1750 microns. This value indicates the width of the raster sampling region, with the software calculating the positions for a number of equally distributed sampling sites within the raster area. For raster settings 1750 and 2000, this equated to 122 and 49 sample points respectively, spread over each sample spot. Comparable results were obtained for the analysis of native serum peptides using both methods, although the smaller raster (1750) resulted in a higher S/N ratio, Figure 4.21. It was observed that the larger raster acquired data outside of the sample spot and also fewer points are selected by default for acquisition. Therefore the 1750 raster setting was selected for subsequent analyses.

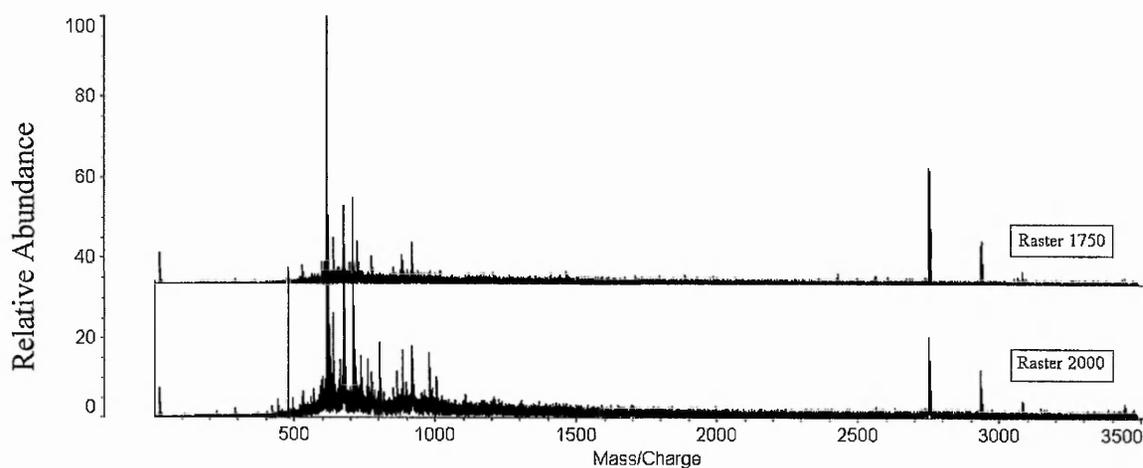


Figure 4.21: Comparison of raster 1750 and 2000 for MALDI-TOF analysis of native serum peptides

The number of shots acquired per profile was then investigated for native serum peptide samples. A total of 200 profiles were obtained for each sample with the number of shots acquired being adjusted, allowing the resulting spectra to be compared for 1 and 5 shots per profile. When the number of laser shots was set to 1 shot (i.e. only 1 laser pulse per profile), the peptide peaks were indistinguishable from background noise. With the shots increased to 5 results were more consistent and peptide ions were observable above inherent noise, Figure 4.22.

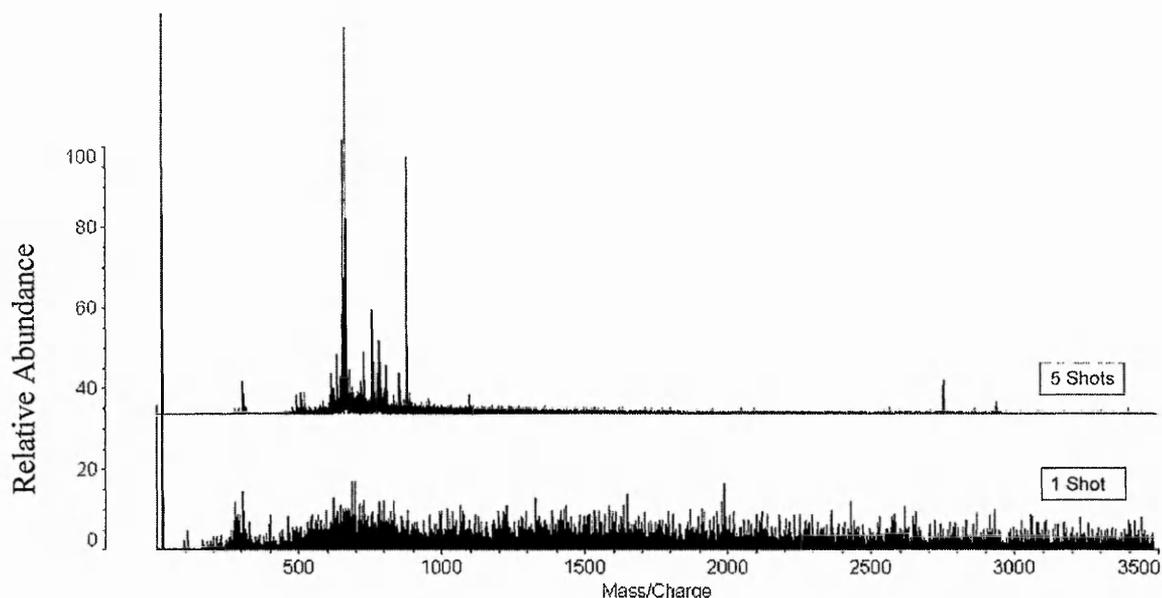


Figure 4.22: Comparison of raster 1shot and 5 shots per profile for MALDI-TOF analysis of native serum peptides

The use of the autoquality mode, where the sample is pre-scanned prior to the data acquisition and the areas producing high signal intensities are selected for subsequent acquisition was then investigated. Again, adjustment of the number of shots per profile from 1 to 5 showed an increase in the quantity of peptide ions observed and an improved S/N ratio, Figure 4.23.

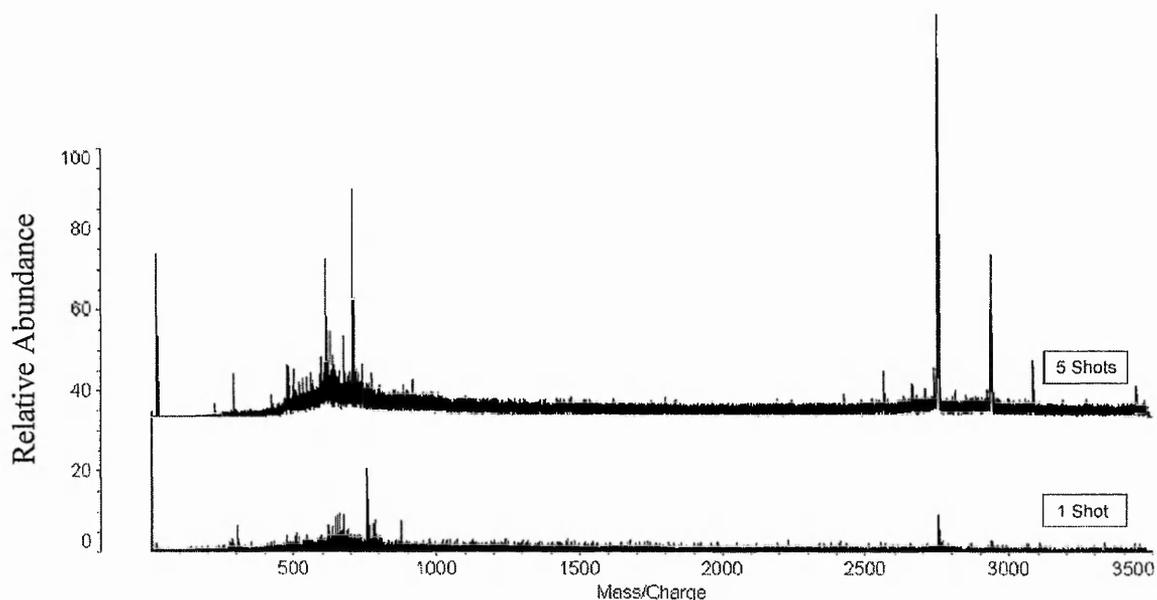


Figure 4.23: Data acquired with autoquality 1 shot and 5 shots for MALDI-TOF analysis

Comparing raster and autoquality mode with shots set to 5 for both, Figures 4.22 and 4.23, showed that the highest signal intensity was observed when the latter technique was used. One possible explanation for this is that there is greater ablation of the sweet spots when autoquality is used therefore more peptide ions will be generated. The optimised mass spectrometry parameters, of autoquality with a raster size of 1750 and shots set to 5 were used for all further studies.

4.3.5 Mass Spectral Reproducibility

Many studies have demonstrated the analysis of serum by SELDI/MALDI MS to distinguish between patients of specific disease state and control patients (see Section 1.4.4.1). Biomarker classification by these approaches relies upon spectral differences in terms of position and amplitude of the resolved peaks. However, little attention has been focused upon the reproducibility of the mass accuracies and intensities of the spectral peaks. The question of reproducibility using the methodology developed in this chapter was addressed for both native and tryptic serum peptides. Spectral patterns were consistent from sample to sample and also from day to day. Figure 4.24, shows spectra for samples of native serum prepared in triplicate using the optimised method and Table 4.3 shows the reproducibility data of peptide masses and intensities ($n = 5$) for the whole procedure (i.e. the samples were prepared five times). The percent relative standard deviations (% RSD) for m/z values and intensities, normalised to m/z 2753, for 5 selected m/z peaks were ≤ 0.03 % and ≤ 29.2 % respectively. The mass accuracy is in accordance with the instrument specification, but no specification exists for the amplitude of the peaks. de Noo *et al*⁵⁴ reported % RSD values of 20 % for the analysis of LMW serum peptides using MALDI-TOF analyses, which is inline with the work reported in this thesis. In contrast Petricoin III *et al*¹⁴ reported that the consistency of the intensities obtained by SELDI-TOF analysis was less than 10 %. The main difference for the variation in % RSD values reported by Petricoin III in comparison to the work reported in this thesis and by de

Noo is likely to be the difference in sample clean-up methods. De Noo used C₈ magnetic beads, Petricoin III used C₁₆ SELDI protein chips and the experiments detailed in this thesis used C₁₈ ZipTip clean-up. The technology is in place for automated sample preparation at Nottingham Trent University, which is expected to improve reproducibility, but development and evaluation of this equipment was outside the scope of this thesis. However, manual sample preparation and analysis reproducibility is in agreement with work reported by others as discussed above.

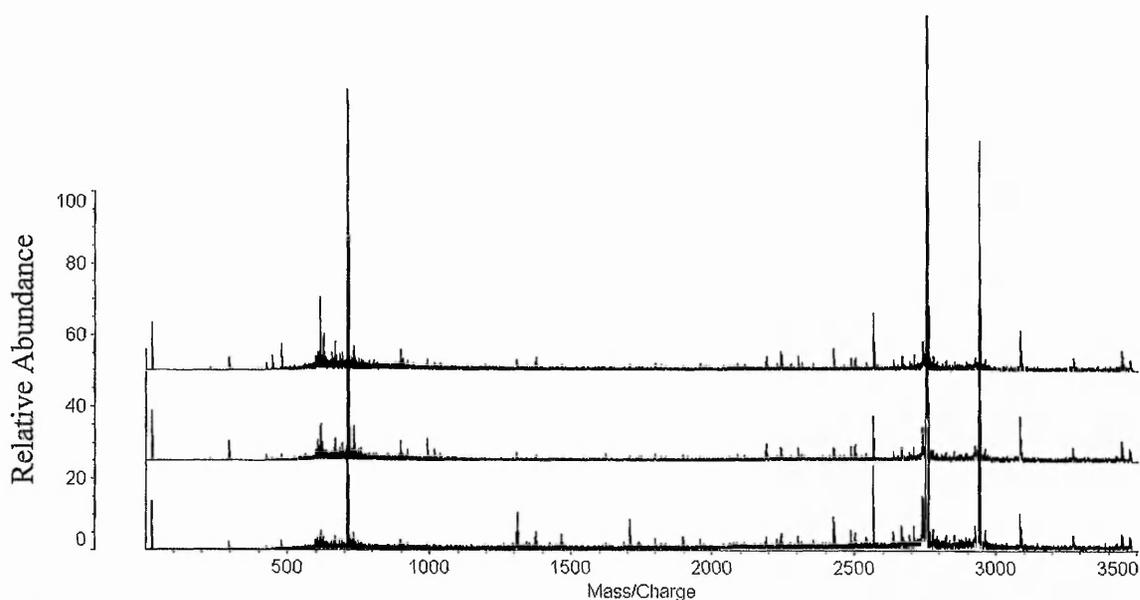


Figure 4.24: Replicate MALDI-TOF spectra of native serum peptides acquired in using the same sample preparation

Table 4.3: Summary of the reproducibility data of native serum peptide masses and relative intensities after normalisation to m/z 2753 (NIR = Normalised Intensity Ratio).

	m/z	NIR	m/z	NIR	m/z	NIR	m/z	NIR	m/z	NIR	m/z	NIR
	899.0	0.01	1377.2	0.02	2938.9	0.34	2246.2	0.02	1958.2	0.01	1799.9	0.01
	898.9	0.02	1377.7	0.03	2938.2	0.45	2245.4	0.03	1958.6	0.02	1799.1	0.02
	898.9	0.01	1377.1	0.03	2937.4	0.32	2246.0	0.03	1959.9	0.01	1798.8	0.01
	899.2	0.02	1377.2	0.04	2938.6	0.28	2246.1	0.03	1958.8	0.02	1799.4	0.02
	899.2	0.03	1377.4	0.04	2937.6	0.30	2247.1	0.04	1958.7	0.02	1799.3	0.02
Mean	899.0	0.02	1377.3	0.03	2938.1	0.34	2246.1	0.03	1958.8	0.02	1799.3	0.02
% RSD	0.02	28.78	0.02	24.81	0.02	19.86	0.03	22.47	0.03	29.19	0.02	22.73

Reproducibility was examined for tryptic digests prepared using the novel methodology developed (Section 4.2.4.3). Results were reproducible and signal intensities were fairly

consistent from sample to sample, Figure 4.25. Reproducibility data in terms of m/z and intensities for six tryptic peptides is presented in Table 4.4. Similar reproducibility values were observed for tryptic digest samples and native serum samples with % RSD for m/z and intensities normalised to m/z 1160 were $\leq 0.01\%$ and $\leq 36.5\%$ respectively.

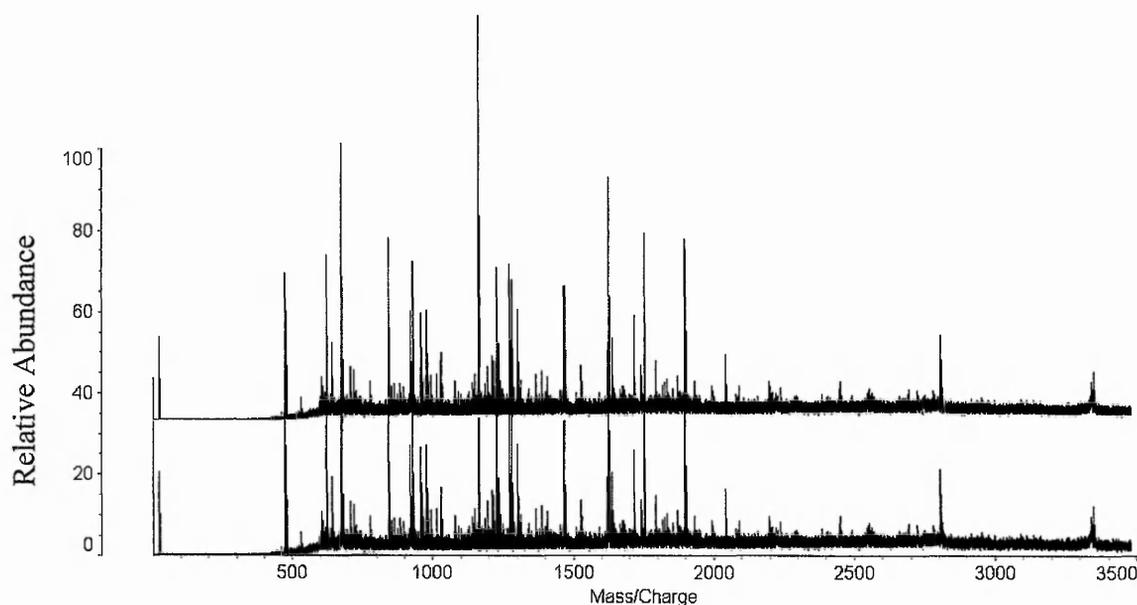


Figure 4.25: Replicate MALDI-TOF spectra of tryptic peptides prepared using the ZipTip before and after digestion method.

Table 4.4: Summary of the reproducibility data of tryptic serum peptide masses and intensities after normalisation to m/z 1160 (NIR = Normalised Intensity Ratio).

	m/z	NIR	m/z	NIR	m/z	NIR	m/z	NIR	m/z	NIR
	842.4	0.32	1160.7	1.00	1716.8	0.50	2451.6	0.16	2806.8	0.23
	842.4	0.34	1160.6	1.00	1716.8	0.32	2451.3	0.16	2806.6	0.20
	842.4	0.45	1160.7	1.00	1716.8	0.23	2451.7	0.18	2806.5	0.21
	842.3	0.29	1160.5	1.00	1716.7	0.20	2451.9	0.14	2806.5	0.24
	842.4	0.46	1160.6	1.00	1716.7	0.25	2451.5	0.24	2806.5	0.18
	842.3	0.22	1160.6	1.00	1716.7	0.27	2451.3	0.12	2807.3	0.19
Mean	842.4	0.35	1160.6	1.00	1716.8	0.29	2451.6	0.17	2806.7	0.21
% RSD	0.01	27.16	0.01	-	0.00	36.48	0.01	24.88	0.01	10.89

The reproducibility study showed that sample preparation and MALDI-TOF analysis for both native and tryptic serum peptide analyses are robust and reproducible, and indicates limits of confidence for protein expression profiling and the use of ANNs for biomarker identification.

4.3.6 Serum Aging Study

Protein profiles have been used to distinguish between healthy and disease state serum samples, for example ovarian cancer,¹⁴ and melanoma.⁵⁵ Samples from cancer patients may be obtained from various clinical centres and subject to different handling procedures prior to storage, transport and analysis (see Chapter 5). It was therefore necessary to validate the procedures shown in Figure 4.1 and also to ensure that any differences observed in the mass spectra are due to biomarkers and not a product of degradation of serum proteins or peptides as a result of handling in transit or storage.⁵⁴ A study of the effect of serum aging and freeze thaw cycles on mass spectrometric response was carried out, prior to MALDI analyses, using the optimised methodology for both native and tryptic serum peptides.

4.3.6.1 Native serum peptides

Fresh serum and serum aged for 3, 8, 24 and 96 hours at room temperature were analysed for native peptides by MALDI. The resulting spectra are shown in Figure 4.26. Significant changes in the intensities were observed in the sera samples over the 96 hour period. Storage for 3 hours at room temperature led to a significant increase in ion intensities in the range $> m/z$ 800, but a decrease in the intensity of ions below this value, notably the base peak at m/z 656, which does not recover in intensity within the 96 hours investigated. A cluster of prominent peaks arise in the m/z region 1000-1700 and two intense peaks appear at m/z 2752 and 2931. After 8 hours of sample aging at room temperature some ions $< m/z$ 1000 increased in intensity but those in the region m/z 1000-3500 were relatively depleted and not observed above baseline noise. After 24 hours incubation a number of peaks were apparent across the whole mass range m/z 500-3500, although fewer intense peaks are observed between m/z 300-1000. Again the peaks of high intensity at m/z 2752 and 2931 were observed along with a small number of ions between m/z 1000-1700. After 96 hours incubation the latter group of ions were not apparent, although those at m/z 2752 and 2931

were still present but observed with lower intensity, compared with the 24 hour sample. The mass region < 1000 shows little change between 24 and 96 hours.

A comparison of the spectra in an expanded region between m/z 1200-1800 obtained for 3, 24 and 96 hour aging (Figure 4.27), shows that many of the same peptides are present at 3 and 24 hours, but intensity variations are observed. The ions in this region (m/z 1200-1800) are not present in fresh serum or that aged for 8 hours, but appear after 3 hours and again at 24 hours. A possible explanation for the appearance, disappearance and recurrence of these peaks in this region of the spectrum is that the proteins initially degrade, giving peptide ions, which are observed after a 3 hour incubation. Further degradation of these peptides leads to the eventual disappearance of the peptides by 8 hours. Some ions are observed again at 24 hours because of further parent protein degradation and at 96 hours the intensity of these peptide products is minimal because of further degradation (Figure 4.27). These experiments were carried out in duplicate demonstrating similar intensity changes in both cases. The findings are supported by others, who have shown that up and down regulation of proteins/peptides in serum is due to proteolytic degradation with increasing incubation time.⁵⁴

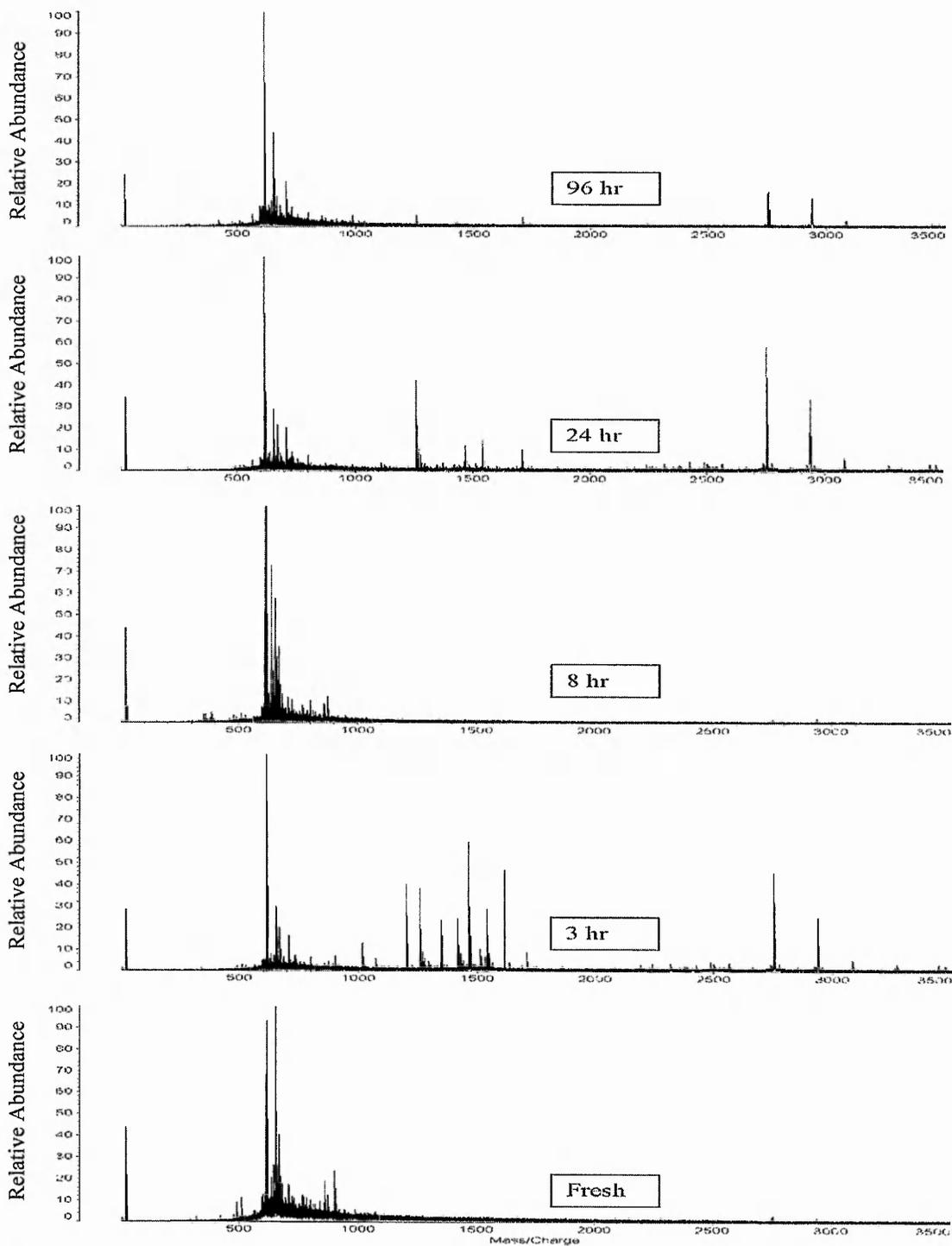


Figure 4.26: MALDI-TOF spectra of native serum peptides showing the spectral differences in samples after incubation at room temperature for 3, 8, 24 and 96 hours in comparison with freshly prepared samples.

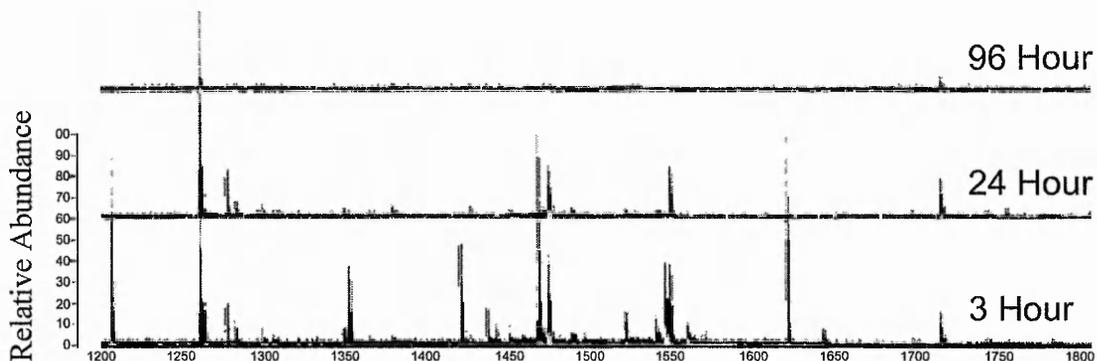


Figure 4.27: Comparison of spectra obtained for serum incubated for 3, 24 and 96 hour mass range region m/z 1200-1800.

The effects of incubation of the samples at 4 °C and freeze-thaw cycles were also examined. Some changes in ion intensities were observed for samples which had been incubated in the fridge for 24 hours or under gone 3 freeze thaw cycles, notably in the m/z 800-1000 and 1200-1500 regions, Figure 4.28. However, these intensity changes are small compared with the changes observed for room temperature incubation. Expansion of the region 2500-3500 Da shows ions of the same m/z are distinguishable and are slightly more intense in the fridge sample in comparison with the freshly prepared sample and the freeze thaw sample. The reason for this may be that the thaw times between re-freezing were insufficient for degradation to occur. However, slow degradation was proceeding for the samples incubated at 4 °C resulting in peaks of greater intensity in comparison with fresh samples. de Noo *et al*⁵⁴ examined freeze thaw cycles for native serum peptides and also reported that room temperature incubation has a greater effect on serum samples than freeze thawing. These findings are important in the clinical environment due to the necessity to develop standard sample handling protocols.

Serum aging data could be used to eliminate differences related to sample history which is very useful when looking for differences in samples due to specific biomarkers. A standardised procedure can be developed to reduce unwanted serum components (such as

degradation products) that may interfere with biomarker identification.

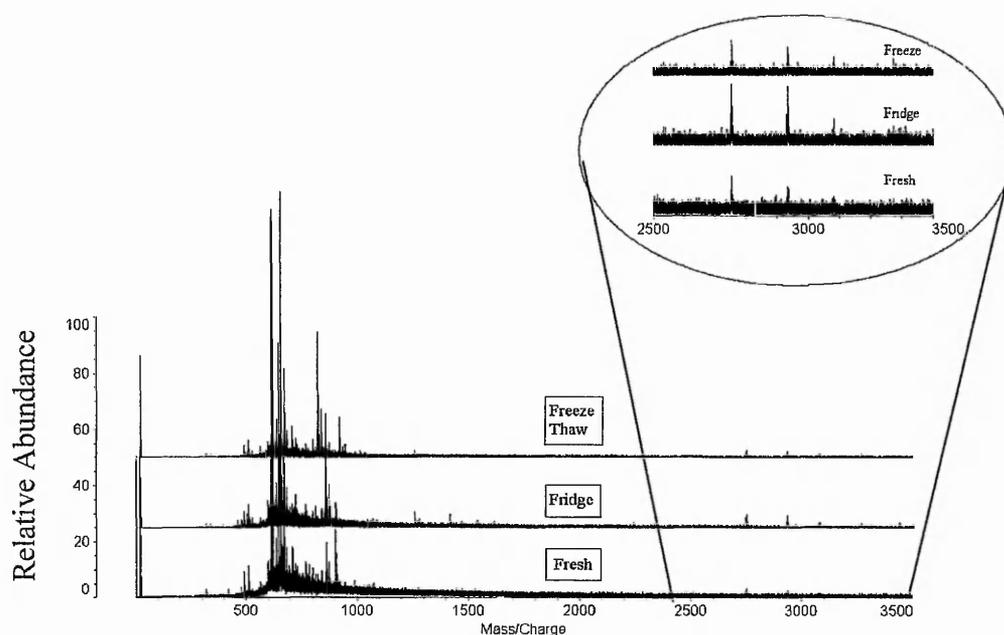


Figure 4.28: MALDI-TOF spectra obtained for freshly prepared samples, samples incubated in the fridge and samples subjected to 3 freeze-thaw cycles.

4.3.6.2 Sensitivity Analysis of Native Serum Peptide Data

The serum aging spectral data were submitted for artificial neural networks (ANNs), as detailed in Chapter 1.4.4.1. using a sensitivity assay. A sensitivity assay uses a basic neural network model that can be used to highlight ions of importance in distinguishing between fresh and aged serum.

The basic principle of an ANNs sensitivity assay initially involves running a set of test samples to determine the overall error on the predictions achieved. The network is then run again using the same samples, but this time the observed value of a particular input (in this case an m/z and intensity) is replaced by random noise, and again the error is calculated. If any important information is removed from the model at this stage, deterioration in the error would be observed. The importance of an input is determined by the ratio of the error observed with the random noise substitution to the original error. Thus the more important an input is to the model, the greater the ratio. If the ratio is one or lower, eliminating that variable either has no effect on the performance of the network, or even improves it. Any input with a ratio of one or less can essentially be removed without any detrimental effects

to the predictive capabilities of the model. Sensitivity analysis was carried out on the data for the m/z region 1000-1500 Da, chosen because it shows the most visual mass spectral differences. Although the statistical model can be used over the entire mass range, the observed sensitivity ratios for the native aged serum samples were much lower than anticipated because noise was incorporated when more data points were used. By limiting the data sets analysed to specific regions of interest (determined by visual examination of spectra) greater accuracy of sensitivity ratios were found. Sensitivity analysis plots of native serum incubated for 3, 8, 24 and 96 hours are shown in Figure 4.29. Sensitivity ratios >1 are deemed as important in distinguishing between a specific sample compared to a standard, in this case each aged serum sample was compared to fresh serum.

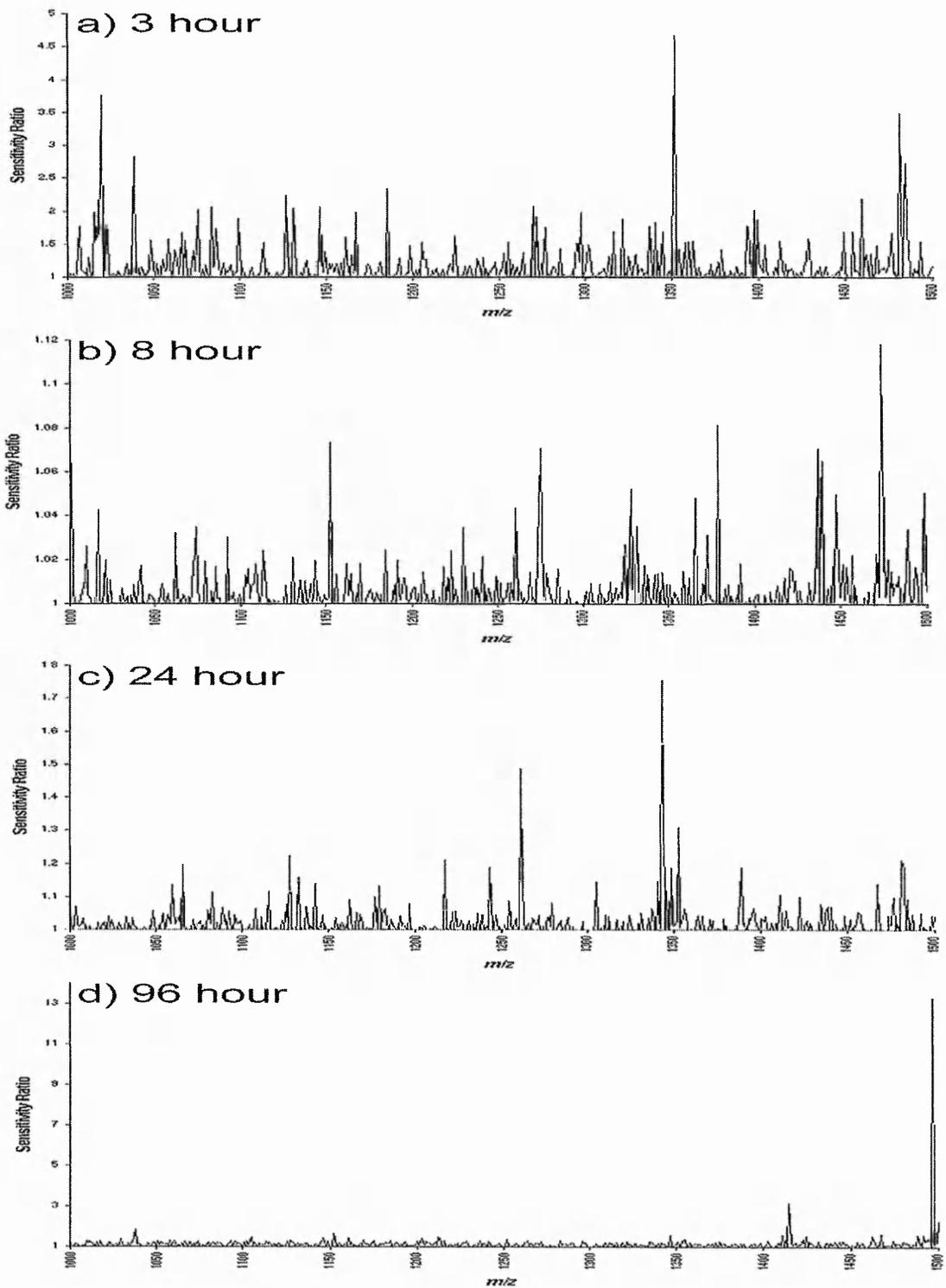


Figure 4.29: Sensitivity analysis derived from ANNs analysis of mass spectral data plots for native serum peptides, sensitivity ratios >1 indicate importance of the ions in the model for discriminating between fresh sera and serum incubated for a) 3 hours b) 8 hours c) 24 hours and d) 96 hours.

Visual inspection of the mass spectral data showed that the sample most easily distinguished from fresh serum was that incubated for 3 hours prior to analysis (containing a large number of degradation products, Figure 4.26). As discussed previously a higher sensitivity ratio suggests higher importance of the ion. For example, m/z 1351 was the most prominent peak in the m/z range 1000-1500, shown in the sensitivity plot for serum aged for 3 hours. The sensitivity ratios for this ion are shown in Table 4.5, which correspond well with the spectra, shown in Figures 4.26 and 4.27. This ion had a sensitivity ratio of 4.67 demonstrating that the ion was very significant in distinguishing between serum aged for 3 hours and a fresh serum sample. This particular ion is prominent in the expanded mass spectrum of serum aged for 3 hours (m/z 1200-1800, Figure 4.27) but is reduced in intensity from fresh and serum aged for 8 and 96 hours. The MALDI-TOF spectrum obtained for the analysis of serum aged for 24 hours had an ion present at m/z 1351, however its intensity was very low (almost within background noise) and therefore could be considered to be similar to fresh serum. Again this was reflected by a low sensitivity ratio.

Table 4.5: Sensitivity Ratios for m/z 1351 at 3, 8, 24 and 96 hour aged serum.

Aging Period / Hours	Sensitivity Ratio
3	4.67
8	0.99
24	1.03
96	1.00

MALDI analysis of native serum samples can therefore be used to distinguish between matured and fresh serum. Samples aged for different times prior to analysis contain degradation products which can be seen by visual inspection of the mass spectra, and are also recognized by statistical analysis. The differentiation of aged samples is important

when considering clinical samples as the history of the sample handling and storage is usually not well known.

4.3.6.3 Tryptic serum peptides

An aging study was also carried out for tryptic peptides derived from serum aged for 3, 8, 24 and 96 hours, as with the native samples. Although sample digestion was carried out over several hours, serum changes with time were investigated taking the incubation time as that prior to digestion. Experimental methodology developed as described previously (Section 4.2.6) was used, wherein the high abundant proteins were removed *via* ZipTip clean-up prior, and following, enzymatic digestion.

Notable spectral differences were observed between the mass spectra corresponding to the aging time points, Figure 4.30. Many high intensity peptide ions were observed in fresh serum samples, some of which remained in all samples incubated up to 96 hours. Changes in the peak intensities were noticeable between samples, with a number of ions in the range m/z 500-2500 decreasing in intensity for serum aged for 3, 8 and 96 hours, compared to fresh serum, but increasing for 24 hour aged serum.

For example, the ion at m/z 1478 was observed to decrease in intensity 5 fold from fresh to 3 hours, but increased in intensity at 24 hours. Some larger ions, however, were found to increase in intensity for serum aged for 3 and 8 hours compared with fresh serum. The ion at m/z 2807 increased approximately 4-fold at 3 and 8 hours, but was observed to decrease in intensity from 8 to 96 hours.

The mass spectra obtained for serum incubated for 24 hours resembled that of fresh serum. As with native peptides, this trend of disappearing and reappearing ions is probably due to the degradation of parent proteins that yielded the observed tryptic peptide ions. Further

degradation results in a decrease of the ion intensities for the samples incubated for 96 hours.

Slight differences were observed for samples which had been freeze thawed or incubated in the fridge in comparison with fresh serum samples, Figure 4.31. For example mass spectral peaks observed at m/z 3347 and 1366 for freeze-thaw samples are not as prominent in fresh samples. Other differences include reduced intensity of peaks observed in the spectrum obtained for samples incubated in the fridge in comparison with fresh serum in the m/z range 850-1200.

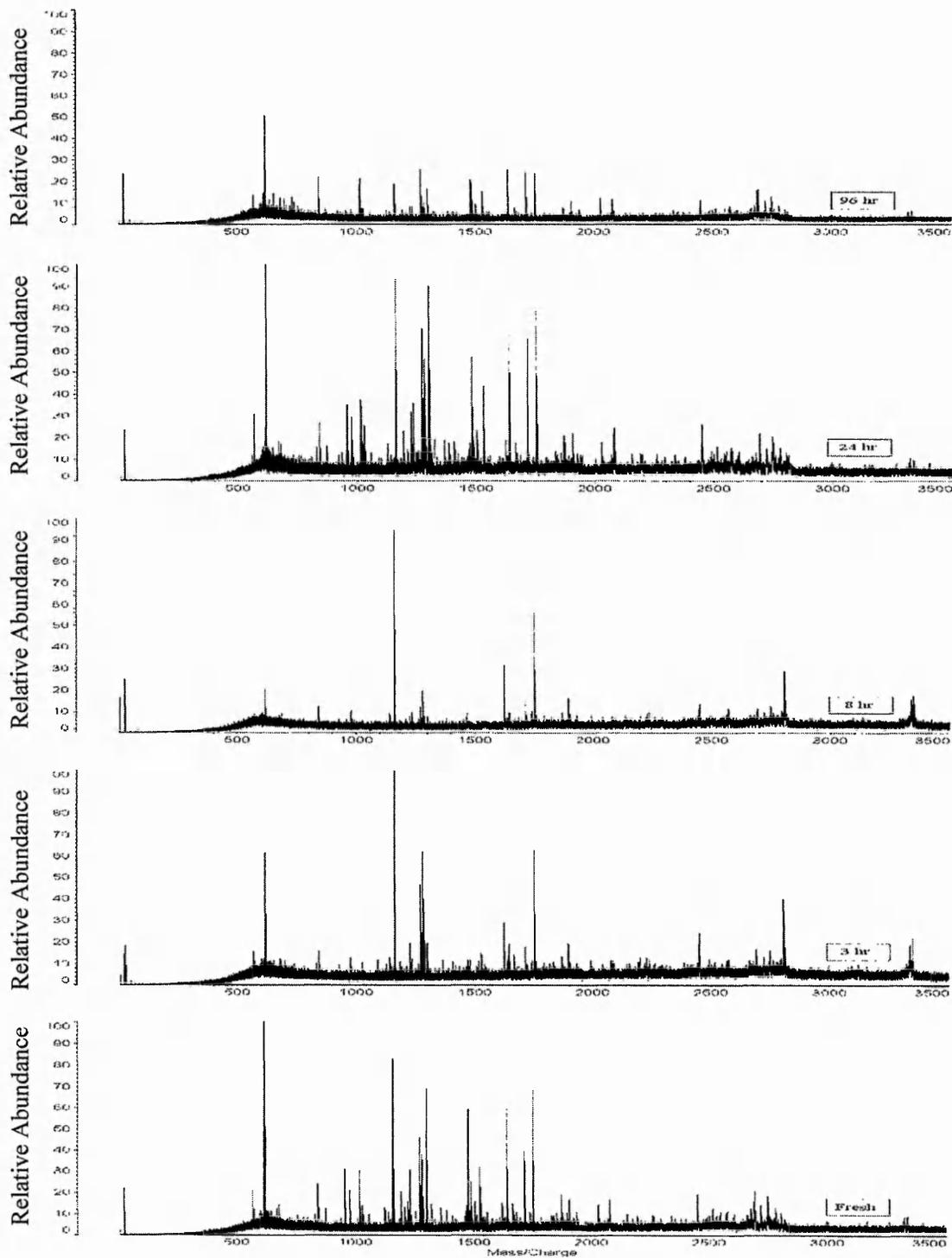


Figure 4.30: MALDI –TOF spectra of tryptic peptides showing spectral differences in samples after incubation at room temperature for 3, 8, 24 and 96 hours in comparison with freshly prepared samples.

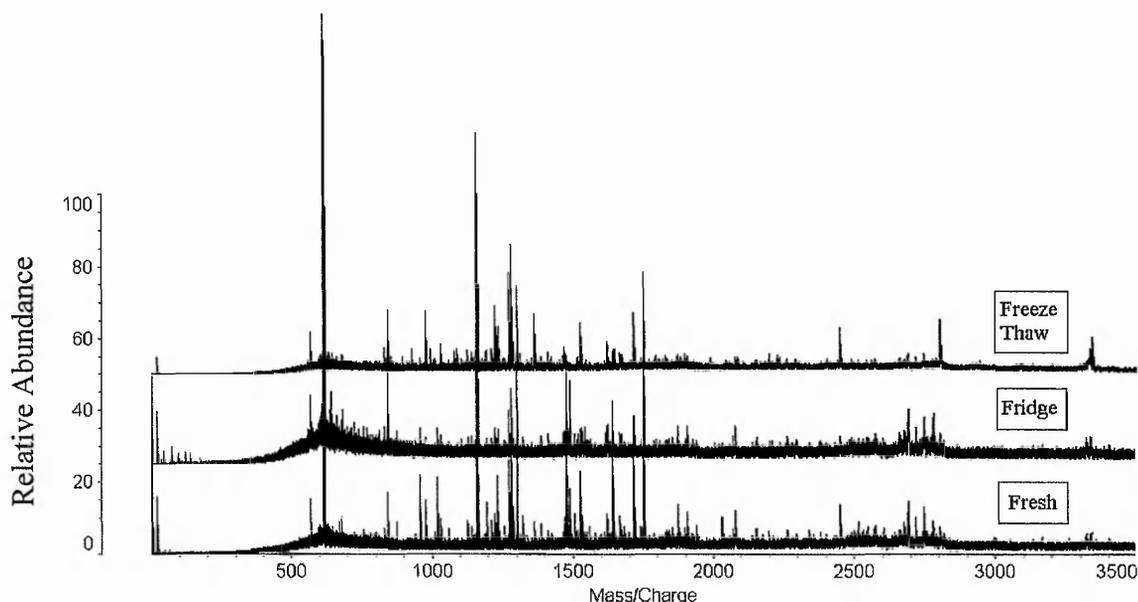


Figure 4.31: MALDI-TOF spectra obtained for tryptic peptides, comparing freshly prepared samples, samples incubated in the fridge and analyzed after 3 freeze-thaw cycles.

4.3.6.4 Sensitivity Analysis of Tryptic Serum Peptide Data

The data obtained for the analysis of fresh serum and serum incubated for 3, 8, 24 and 96 hours was submitted for a sensitivity assay. Plots of the sensitivity analysis for tryptic serum peptides are shown in Figure 4.32. As previously described in Section 4.3.6.2 only ions with a sensitivity ratio >1 are deemed significant in discriminating between fresh serum and aged serum. However, the greater the sensitivity ratio the more important the ion is in discriminating between aged and fresh samples.

The sensitivity analysis for serum incubated for 3 hours prior to digestion shows many significant peaks, although these were less important than those observed in the 8 and 24 hour sample plots, Figure 4.32. A number of peaks with high sensitivity ratios were apparent in the 8 hour sample sensitivity plot reflecting the contrast in mass spectral peak intensities observed compared to fresh serum, Figure 4.30. The peptide ion at m/z 1301 shows the highest sensitivity ratio (13.0) of all ions in the ANNs analysis. Mass spectra obtained for serum aged for 24 hours before digestion more closely resembled that of

digested fresh serum than any of the other aged samples. These differences are difficult to visualize in the mass spectra. However, sensitivity analysis showed ions present with high ratios distinguishing between the two states. For example, ions at m/z 1027, 1045, 1478 and 1489, which have sensitivity ratios of 1.73, 1.65, 1.67 and 1.86 respectively. Statistical analysis of the data acquired for 96 hours showed ions in the m/z region 1000-1500 to have much lower significance when compared to those obtained for all other time points analysed.

The results reported in this Section show that tryptic peptides can be used to distinguish between aged and fresh serum. As was observed with native serum (Section 4.3.6.1) degradation products gave rise to many observable peaks in the mass spectra. Sensitivity plots show significant ions are present that also classify differences between samples.

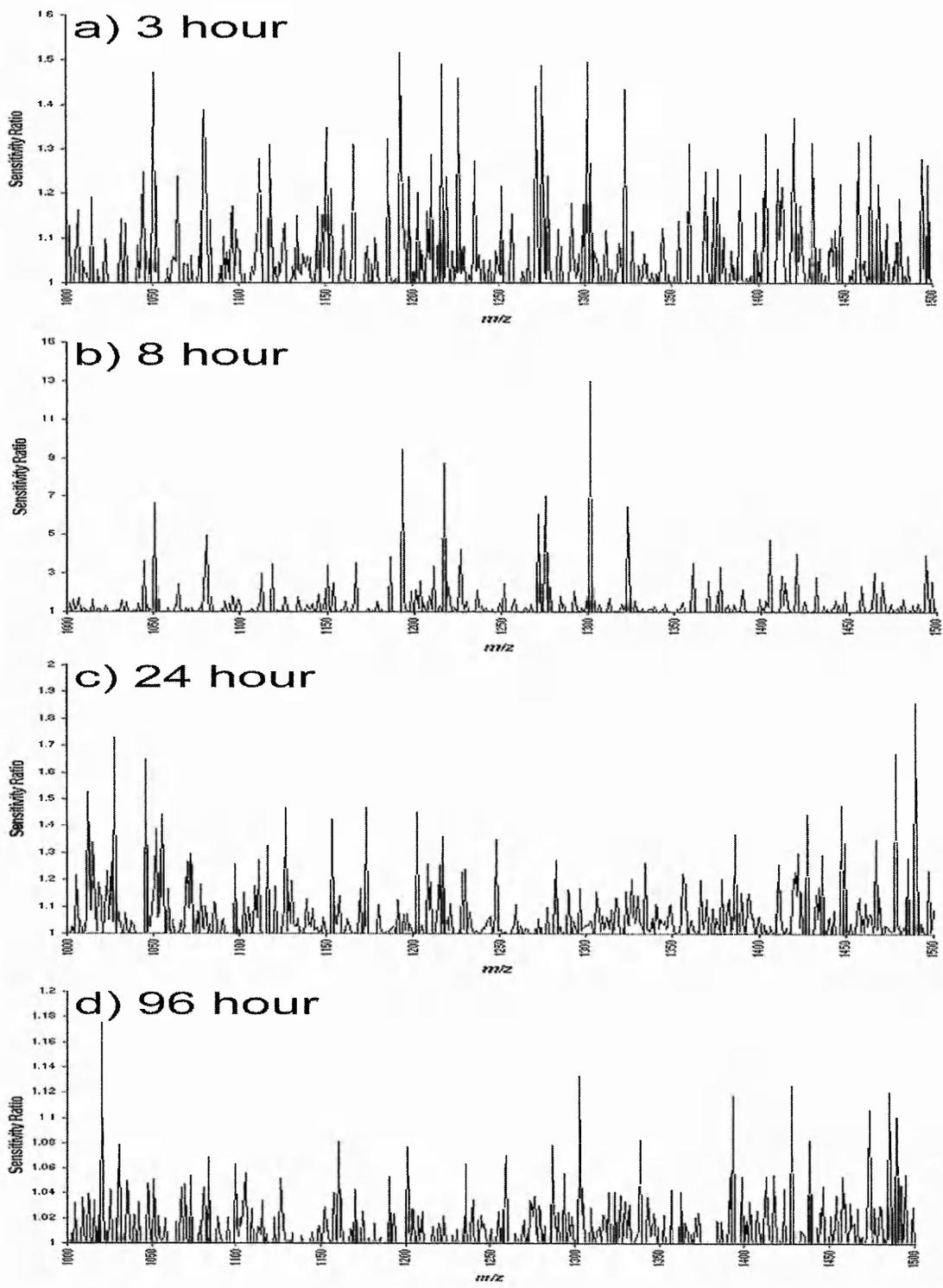


Figure 4.32: Sensitivity analysis plots for m/z 1000-2000 for serum samples subjected to 3, 8, 24 and 96 hour room temperature incubation prior to digestion.

4.4 Conclusion

Methods have been developed for the MALDI analysis of native and tryptic serum peptides for biomarker identification. Ultrafiltration, protein precipitation and ZipTip clean-up techniques were examined for the removal of high molecular weight proteins. This is the first reported method to use ZipTip clean-up prior to tryptic digestion to deplete high abundant proteins and further clean-up following digestion, in order to analyse LMW peptides. Results have demonstrated the reproducibility and validation in terms of m/z and signal intensity for the developed methods.

Changes in the proteome arising from room temperature incubation have been investigated and results show that spectral differences can be observed for both native and tryptic serum peptides. Future clinical samples can therefore be assessed for degradation, ensuring that any changes observed are due to disease states rather than sample aging. The results of this study will help future sample collection and storage conditions of serum to be addressed.

The serum aging data reported in this chapter emphasises the necessity for clearly defined sampling handling protocols. Many of these variables were already controlled as a matter of course during experimentation detailed in this thesis, such as keeping samples on ice, but the results reported here highlight their importance. The techniques developed within this chapter were evaluated for the analysis of serum samples taken from stage IV melanoma patients and control patients in Chapter 5. Defined protocols were adhered to, applying the optimised methods for native and tryptic serum peptide analysis as described in Sections 4.2.5 and 4.2.6 respectively. For native peptide analysis the protocol included serum dilution, acetonitrile protein precipitation, lyophilization, resuspension and ZipTip clean-up. For tryptic serum peptides samples were prepared by ZipTip clean-up before and after proteolytic digestion.

4. 5 References

- 1) Adkins J.N., Varnum S.M., Auberry K.J., Moore R.J., Angell N.H., Smith R.J., Springer D.L., Pounds J.G., *Mol. Cell. Proteomics*, **1**, 947-952 (2002)
- 2) Ball G., Mian S., Holding F., Alliborne R.O., Lowe J., Ali S., Li G., McCardle S., Ellis S.I.O., Creaser C., Rees R.C., *Bioinformatics* **18**, 395-404 (2002)
- 3) Bons J.A.P., Wodzig W.K., van Dieijen-Visser M.P., *Clin. Chem. Lab. Med.*, **43**, 1281-1290 (2005)
- 4) Clarke C.H., Buckley J.A., Fung E.T., *Clin. Chem. Lab. Med.*, **43**, 1314-1320 (2005)
- 5) Fung E.T., Wright G.L., Dalmaso E.A., *Curr. Opin. Mol. Ther.*, **2**, 643-650 (2000)
- 6) Stanley B.A., Gundry R.L., Cotter R.J., Van Eyk J.E., *Dis. Markers*, **20**, 167-178 (2004)
- 7) Fenselau C., Demirev P.A., *Mass Spectrom. Reviews*, **20**, 157-171 (2001)
- 8) Eberini I., Miller I., Zancan V., Bolego C., Puglisi L., Gemeiner M., Gianazza E., *Electrophoresis*, **20**, 846-853 (1999)
- 9) Anderson L., Anderson N.G., *Proc. Natl. Acad. Sci USA.*, **74**, 5421-5425 (1977)
- 10) Eberini I., Agnello D., Miller I., Villa P., Fratelli M., Ghezzi P., Gemeiner M., Chan J., Aebersold R., Gianazza E., *Electrophoresis*, **21**, 2170-2179 (2000)
- 11) Sasaki K., Sato K., Akiyama Y., Yanagihara K., Oka M., and Yamaguchi K., *Cancer Res.*, **62**, 4894-4898 (2002);
- 12) Grossklaus, D.J., Smith J.A., Shappell, S.B., Coffey, C.S., Chang S.S., Cookson, M.S., *Urol. Oncol.*, **7**, 195-198 (2002)
- 13) Petricoin E.F., Paweletz C.P., Liotta L.A., *J. Mammary Gland Biol. Neoplasia*, **7** 433-440, (2002)
- 14) Petricoin E.F., Ardekani A.M., Hitt B.A., Levine P.J., Fusaro V.A., Steinberg S.M., Mills G.B., Simone C., Fishman D.A., Kohn E.C., Liotta L.A., *Lancet* **359**, 572-577 (2002)

- 15) Basso D., Valerio A., Seraglia R., Mazza S., Piva M., *Pancreas* **24**, 8-14 (2002)
- 16) Rubin R.B., Merchant M., *Am. Clin. Lab.* **19**, 28-29 (2000)
- 17) Anderson N.L., Anderson N.G., *Mol. Cell. Proteomics*, **1**, 845-867 (2002)
- 18) Govorukhina N., Keizer-Gunnink A., van der Zee A.G., de Jong S., de Bruijn H.W., Bischoff R., *J. Chromatogr. A.*, **1009**, 171-178 (2003)
- 19) Sato, A.K., Sexton, D.J., Morganelli, L.A., Cohen, E.H., Wu, Q.L., Conley, G.P., Streltsova Z., Lee S.W., Devlin M., DeOliveira, D.B., Enright, J., Kent, R.B., Westcott C.R., Ransohoff T.C., Ley A.C., Ladner R.C., *Biotechnol. Prog.*, **18**, 182-192 (2002)
- 20) Zhou M., Lucas D.A., Chan K.C., Issaq H.J., Petricoin III, E.F., Liotta L.A., Veenstra T.D., Conrads T.P., *Electrophoresis.*, **25**, 1289-1298 (2004)
- 21) Viberg A., Sandstorm M., Jansson B., *Rapid Commun. Mass Spectrom.*, **18**, 707-710 (2004)
- 22) Chertov O., Biragyn A., Kwak L.W., Simpson J.T., Boronina T., Hoang V.M., Prieto D.A., Conrads T., Veenstra T.D., Fisher R.J., *Proteomics*, **4**, 1195-1203 (2004)
- 23) Caputo E., Lombardi M.L., Luongo V., Moharram R., Tornatore P., Pirozzi G., Guardiola J., Martin B.M., *J. Chromatogr. B.*, **819**, 59-66 (2005)
- 24) Siegel M.M., Tabei K., Bebernitz G.A., Baum E.Z., *J. Mass Spectrom.*, **33**, 264-273 (1998)
- 25) Wu S.L., Amato H., Shieh P., Hancock W.S., *J. Proteome Res.*, **1**, 459-465 (2002)
- 26) Ardekani A.M., Liotta L.A., Petricoin E.F. III., *Expert Rev. Mol. Diagn.*, **2**, 312-320 (2002)
- 27) Burtis C.A., Ashwood E.R., *Tietz fundamentals of Clinical Chemistry*, 5th Ed., W.B Saunders Company, Philadelphia, PA (2001)
- 28) Seraglia R., Vogliardi S., Allegri G., Comai S., Lise M., Rossi C.R., Mocellin S., Scalerta R., Raggazi E., Traldi P., *Eur. J. Mass Spectrom.*, **11**, 353-360 (2005)

- 29) Merrell K., Southwick K., Graves S.W., Esplin M.S., Lewis N.E., Thulin C.D., *J. Biomol. Tech.* **15**, 238-248 (2004)
- 30) Tirumalai R.S., Chan K.C., Prieto D-R A., Issaq H.J., Conrads T.P., Veenstra T.D., *Mol. Cell. Proteomics*, **2**, 1096-1103 (2003)
- 31) Georgiou H.M., Rice C.E., Baker M.S., *Proteomics*, **1**, 1503-1506 (2001)
- 32) Vorm O., Roepstorff P., Mann, M., *Anal. Chem.*, **66**, 3281-3287 (1994)
- 33) Whitby L.G., Smith A.F, Beckett G.J., Walker S.W., Lecture notes on clinical chemistry 5th Ed., Blackwell scientific publications, (1993)
- 34) Hsu J.L., Chou M.K., Liang S.S., Huang S.Y., Wu C.J., Shi F.K., Chen S.H., *Electrophoresis.*, **25**, 3840-3847 (2004)
- 35) Wilson C.J., Clegg R.E., Leavesley D.I., Percy M.J., *Tissue Engineering*, **11**, 1-18 (2005)
- 36) Ferrari L, Seraglia R., Rossi C.R., Bertazzo A., Lise M., Allegri G., Traldi P., *Rapid. Commun. Mass Spectrom.*, **14**, 1149-1154 (2000)
- 37) Chernokalskaya E., Kavonian H., Glazebrook S., Gutierrez S., Pitt A.M., Leonard J., *Electrophoresis*, **25**, 2461-2468 (2004)
- 38) Taylor R.L., Grebe S.K., Singh R.J., *Clinical Chemistry*, **50**, 2345-2352 (2004)
- 39) Daykin C.A., Foxall P.J.D., Connor S.C., Lindon J.C., Nicholson J.K., *Anal. Biochem.*, **304**, 220-230 (2002)
- 40) Kos T., Moser P., Yilmatz N., Mayer G., Pacher R., Hallstrom S., *J. Chromatogr. B*, **740**, 81-85 (2000)
- 41) Polson C., Sarkar P., Incedon B., Raguvanan V., Grant R., *J. Chromatogr. B.*, **785**, 263-275 (2003)
- 42) Speicher K.D., Kolbas O., Harper S., Speicher D.W., *J. Biomol. Tech.* **11**, 74-86 (2000)
- 43) Zhang Y-T., Geng Y-P., Zhou L., Lai B-C., Si L-S., Wang Y-L., *World J. Gastroentero/.*, **11**, 4679-4684 (2005)

- 44) Yoon S.W., Kim T.Y., Sung M.H., Kim C.J., Poo H., *Proteomics*, **5**, 1987-1995 (2005)
- 45) Tien C-C., Chen J-B., Wang C-C., Lee W-C., *Enzyme and Microbial Technology*, **33**, 488-491 (2003)
- 46) Flensburg J., Belew M., *J. Chromatography A.*, **1009**, 111-117, (2003)
- 47) Eklund P., Andersson H.O., Kamali-Moghaddam M., Sundstrom L., Flensburg J., *J. Chromatography A.*, **1009**, 179-188, (2003)
- 48) Ishino M., Aoto H., Sasaki H., *Tumour Res.*, **37**, 23-31 (2002)
- 49) Marvin L.F., Roberts M.A., Fay L.B., *Clinica Chimica Acta*, **337**, 11-21 (2003)
- 50) Harper R.G., Workman S.R., Schuetzner S., Timperman A.T., Sutton J.N., *Electrophoresis*, **25**, 1299-1306 (2004)
- 51) Koomen J.M., Zhao H., Li D., Abbruzzese J., Baggerly K., Kobayashi R., *Rapid Commun. Mass Spectrom.*, **18**, 2537-2548 (2004)
- 52) Lim H., Eng J., Yates III J.R., Tollaksen S.L., Giometti C.S., Holden J.F., Adams M.W.W., Reich C.I., Olsen G.J., *J Am. Soc. Mass Spectrom.*, **14**, 957-970 (2004)
- 53) Huang H-L., Stasyk T., Morandell S., Mogg M., Schreiber M., Feuerstein L., Huck CW., Stecher G., Bonn GK., Huber G., *Electrophoresis*, **26**, 2843-2849 (2005)
- 54) de Noo, M.E.; Tollenaar R.A.E.M., Oezalp A., Kuppen P.J.K., Bladergroen M.R., Eilers P.H.C., Deelder A.M., *Anal. Chem.*, **77**, 7232-7241 (2005)
- 55) Mian S., Ugurel S., Parkinson E., Schlenzka I., Dryden I., Lancashire L., Ball G., Creaser C., Rees R., Schadendorf D., *J. Clin. Oncol.* **23**, 5088-5093 (2005)

Chapter 5

Identification of Diagnostic Biomarkers Associated with Stage IV Melanoma Using MALDI-TOF Mass Spectrometry and Bioinformatics

5.1 Melanoma

Melanoma of the skin, also known as cutaneous melanoma or malignant melanoma, is the least common but most deadly type of skin cancer resulting in 79 % of skin cancer deaths.¹

The term melanoma is derived from the Greek words '*melas*' and '*oma*', meaning black and tumour respectively. Such malignant tumours occur in melanocytes, the cells responsible for the production of melanin which give hair and skin their pigmentation protecting them against UV damage. Melanoma is associated with darkened skin deformities such as moles due to the high localisation of pigments found in these areas. Although melanoma is most common in skin it can also develop in the eye and the lining of the nose or mouth. If melanoma is detected and treated early enough the chances of total recovery and disease free survival is excellent.²

The first indication of melanoma is a change in size, shape or colour of an existing mole or the development of a highly pigmented area. Tumours develop in the epidermis, the outermost layer of skin although can become rooted in the dermis, the inner skin layer, as the disease progresses, Figure 5.1. Melanoma can metastasize quickly to other body parts through the lymph system or through blood. Since the majority of primary melanomas are visible on the skin, there is a good chance of detecting the disease in its early stages.

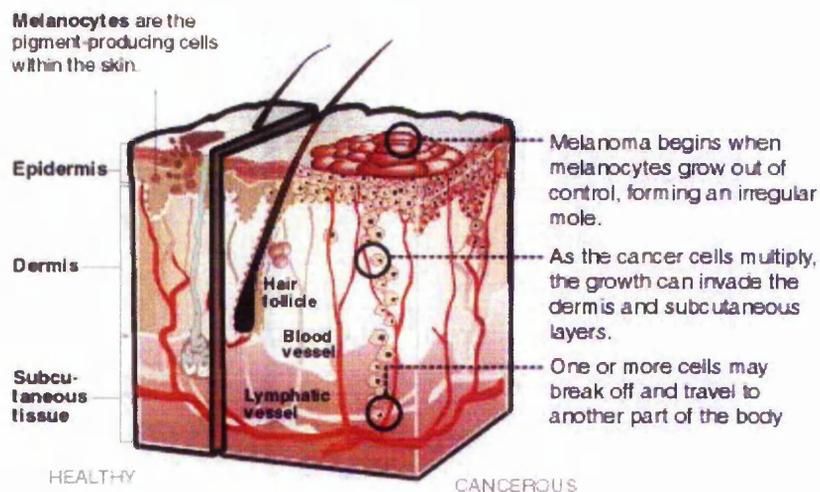


Figure 5.1: Diagram showing the skin.³

5.1.1 Stages of Melanoma

Stages describe the extent of a cancer within the body, especially with respect to whether the cancer has spread from the original tumour site to other parts of the body.

There are five main categories of melanoma, stages 0, I and II relate to melanoma confined to the skin, stage III for lymph node involvement and stage IV for the spread of the cancer to other organs. These are summarised in Table 5.1. The prognosis usually correlates with the disease stage. ^{4, 5} Therefore early detection of melanoma is crucial, a five year survival rate drops dramatically from 90-95 % for stage I patients to less than 5 % for stage IV. ⁶ Poor survival rate is partly due to the lack of effective treatment for metastatic disease. ⁷

Table 5.1: Stages of Melanoma ⁸

Stages	Description
0	Abnormal cells are found only in the outer layer of skin cells and do not invade deeper tissue
I	Cancer is found in the epidermis and/or the upper part of the dermis but it has not spread to the nearby lymph nodes. The tumour is 0-2.0 mm thick without ulceration or 0-1.0 mm thick with ulceration. Ulceration is when the epidermis covering the primary melanoma is not intact.
II	The tumours are greater than 1.0 mm, are ulcerated or 2.0 mm or more and non-ulcerated. Cancer has spread to the lower part of the dermis but not into the tissue below the skin or into nearby lymph nodes.
III Regional disease	The tumour maybe larger or smaller than 4 mm thick, may have spread to the lower layers of the skin, have additional tumour growths within 1 inch of the original tumour (satellite tumours), have spread to surrounding lymph nodes and be actively spreading to nearby areas of the body.
IV Distant Metastasis	The tumour has spread beyond the lymph nodes, usually to the lungs, liver or brain.

The tumour depth and the tumour-node-metastasis (TNM) staging system, defined in Table 5.2, are normally used to help stage melanoma. ⁹ Breslow thickness is a method of measuring in mm how deeply the tumour has penetrated into the skin. Generally, the thickness of the primary tumour correlates with prognosis. ¹⁰ TNM staging is common to all cancers, the T represents the primary tumour thickness and N describes whether the

lymph nodes contain cancer cells and may be further divided into a, b, and c. If cancer cells in the lymph nodes are only visible with a microscope it is termed stage a. Obvious signs of cancer in the nodes are termed stage b and stage c applies to in transit metastases where nodules are further than 5 cm from the primary tumour. The M term describes whether the cancer has spread to other parts of the body. Using table 5.2, it can be seen that a cancer described as T1-4aN3M0 means that the tumour is not ulcerated and ranges in size from less than 1 mm to 4 mm thick. The cancer has spread to four or more lymph nodes and the tumour has not spread to other parts of the body.

Table 5.2: TNM classification ⁹

T	1	Melanoma is <1 mm thick
	2	Melanoma between 1-2 mm thick
	3	Melanoma between 2-4 mm thick
	4	Melanoma > 4 mm thick or there are clusters of melanoma cells in the surrounding skin less than 5 cm from the primary tumour
	A	Not ulcerated
	B	Ulcerated
N	0	No positive lymph nodes
	1	One positive lymph node
	2	2-3 positive lymph nodes
	3	4 or more positive lymph nodes
M	0	No sign of cancer spread anywhere else
	1	Melanoma in another part of the body

5.1.2 Development and Progression of Melanoma

Melanoma develops in a similar manner to other types of cancers. DNA genes contained within a cell responsible for controlling cell division and reproduction become damaged. In the case of melanoma this is usually caused by UV radiation, which damages DNA causing mutations that lead to malignant transformation.¹¹ Once this damage has occurred the cells divide and grow without control eventually becoming a malignant tumour. The effected cells in the case of melanoma are melanocytes, pigment producing cells located in the basal layer of the epidermis. Therefore the first tumours usually develop in the skin. If

treatment is not early enough the melanoma grows and spreads horizontally along the epidermis before penetrating into deeper layers of skin.¹² Once the tumour has reached the dermis it comes into contact with circulation pathways, such as blood vessels, nerve endings, oil and sweat glands, muscle fibres and lymph vessels. The lymphatic system consists of tissues, organs and a network of vessels which produce, store and transport lymph around the body. This fluid is drained to lymph nodes where it is filtered to remove foreign bodies including cancer cells and then recirculated *via* the lymph vessels throughout the body. Some cancer cells may break away from the primary tumour and enter the lymph vessels spreading to the lymph nodes. Melanomas that are spread by the blood stream or the lymphatic system are known as metastatic melanoma.¹³

Melanoma may be categorised into four different types.² The most common is superficial spreading melanoma which spreads along the epidermis for a long period of time before penetrating deeper into the skin. Nodular melanoma tends to occur more frequently in men than women, penetrates deeply into the skin and spreads much faster than other types. Melanoma arising from a flat brown spot associated with aging or sun damage (liver spots or aging spots) is classified as lentigo maligna melanoma. On a worldwide scale acral lentiginous melanoma is the least common but accounts for the majority of African, Asian and American melanomas. This type typically develops on palms, soles and under the nails and it is common to mistake early signs as bruises or injuries to these extremities delaying diagnosis.

5.1.3 Melanoma Biomarkers

The diagnosis of malignant melanoma remains one of the most difficult to render in surgical pathology, partially because of its extreme histologic variability. While early stages of melanoma are recognized by changes in size, shape or colour of black nevi (moles), most melanomas grow downward from the skin therefore eluding detection. They invade neighbouring tissues before they are detected as highly metastatic tumors in the

lymph nodes or other organs.¹⁴ The problem with most routine radiological investigations is that they are of low sensitivity and specificity. As are blood tests such as full blood count, liver function tests and Lactate dehydrogenase (LDH). A study by Jillella *et al*¹⁵ found that of 49/279 patients who developed tumour recurrences, 47 were identified by history or examination and none by blood tests. Serum LDH has been found to be a sensitive indicator of tumour load (amount of cancer in the body), but is only of value in Stage IV melanoma. Several studies have reported links between elevated levels of LDH and shortened survival.¹⁶⁻¹⁸

A low molecular weight protein, S100, has been shown to be up-regulated in the serum of patients with melanoma¹⁹⁻²² and correlations between S100 concentration and disease progression have been reported.^{19, 20, 23} However, it has been suggested that S100 has little value as a diagnostic melanoma marker because of the low percentage of stage I and stage II patients testing positive for the protein.^{21, 22, 24} Limits in the sensitivity and/or specificity of the currently available melanocytic markers such as anti-S100, HMB45, and anti-MelanA complicate the diagnosis and prognosis of melanoma. Other prognostic factors for stage IV melanoma include sites of metastasis, tumour mass and number of metastases at distant sites.¹⁰ Despite the use of these markers in cutaneous melanoma, the disease progression is complex, with the later stages indicative of aggressive disease and poor survival. Clinical proteomics aims to scan the realm of expressed proteins to identify biomarkers that can answer specific clinical questions.

Recent advances in mass spectrometry based proteomics have yielded novel and promising techniques to aid in biomarker identification.²⁵ One such advancement is the development of protein mass spectrometry and the ability to analyze complex samples using this technique. SELDI-TOF and MALDI-TOF, discussed in Section 1.4.4.1, are the two most popular mass spectrometry approaches presently employed for detecting quantitative or qualitative changes in circulating serum or plasma proteins in relation to a pathological

state such as the presence of a tumour. Both represent high throughput proteomic approaches that allow protein expression profiling of large sample sets^{26, 27} SELDI and MALDI have been applied, with reported high sensitivity and specificity using a variety of statistical analyses, to the analysis of many types of cancer including prostate^{28, 29} and ovarian.^{30, 31} Many reports of the detection of potential melanoma biomarkers have used SELDI or MALDI mass spectrometry to identify serum protein biomarker ions.³²⁻³⁴

These reports along with many other biomarker studies provide only *m/z* values for intact proteins or native protein fragments which distinguish between disease state and healthy controls, but do not allow protein identification.^{25, 28, 35-41} The experimentation reported in this Chapter describes the application of the standardised, validated methods developed in Chapter 4 for the MALDI analysis of native and tryptic serum peptides for control and melanoma stage IV patients, to search for biomarkers associated with malignant melanoma. The prognostic melanoma peptides/proteins deemed to be important by ANNs were identified using a combination of MALDI-TOF-MS PSD, AP-MALDI-MS/MS and nano-ESI-MS/MS techniques followed by database searches on MASCOT and SEQUEST.

5.2 Methods and Materials

5.2.1 Serum samples

Sera samples from stage IV melanoma patients (30 male and 20 female) and sera samples from healthy controls were supplied by DKFZ University of Mannheim after patient consent. Ethics approval was obtained from the institutional review board of the medical faculty Mannheim.

5.2.2 Sample Preparation

Matrix, α -cyano-4-hydroxycinnamic acid (LaserBio Labs, Sophia-Antipolis Cedex, France) was prepared as a 10 mg mL^{-1} solution in 50 % acetonitrile/water + 0.1% TFA. For both native and tryptic serum peptides all samples (50 control and 50 stage IV melanoma) were prepared using the optimised methodology described in Section 4.2.5 and 4.2.6 respectively and briefly described below.

Native serum peptides were prepared by an acetonitrile precipitation method with subsequent C_{18} ZipTip clean-up. Serum was diluted 1/20 with 0.1 % TFA and equal volumes ($12.5 \text{ }\mu\text{L}$) of the diluted serum and acetonitrile were combined and allowed to stand at room temperature for 10 minutes, followed by centrifuged for 10 minutes. A $15 \text{ }\mu\text{L}$ aliquot of the supernatant was removed and speed vacuumed to near dryness and resuspended in 0.1 % TFA followed by C_{18} ZipTip clean-up.

For the preparation of tryptic peptides, aliquots ($25 \text{ }\mu\text{L}$) of diluted serum (1/10 with 0.1 % TFA) were subjected to C_{18} ZipTip clean-up with peptides eluted in $4 \text{ }\mu\text{L}$ of 80 % ACN. The digestion buffer consisting of ammonium bicarbonate ($16.6 \text{ }\mu\text{L}$ of 100 mM), water ($7.6 \text{ }\mu\text{L}$) and trypsin ($1.3 \text{ }\mu\text{L}$ of $0.5 \text{ }\mu\text{g }\mu\text{L}^{-1}$) was added to the eluate. Digestion was carried out at $37 \text{ }^\circ\text{C}$ overnight and the resulting peptides were concentrated and de-salted using a ZipTip Clean-up.

Processed samples were applied to the MALDI-TOF target plate using the dried droplet method; equal volumes ($1 \text{ }\mu\text{L}$) of sample and matrix were mixed on plate and allowed to

air dry. The samples were analysed in duplicate and assigned to positions on the target plate in a randomised manner to avoid instrumental bias. Samples were assigned a number and then randomised using a Microsoft Excel™ function.

5.2.3 Mass Spectrometric Analysis

5.2.3.1 MALDI-TOF MS

MALDI-TOF experiments were performed on a Axima CFR+ MALDI-TOF (Shimadzu, Manchester, UK) operated in reflectron mode (1-3500 Da) using the optimised parameters as described in Chapter 4, Section 4.2.8.1. Close external calibration was performed using peptide calibration mix 4 (proteomix) supplied by Laser Bio Labs (Sophia-Antipolis Cedex, France), based on the monoisotopic values of $[M+H]^+$ of bradykinnin fragment 1-5, angiotensin II, neurotensin, ACTH clip (18-39) and insulin B-chain oxidised at m/z 573.31, 1046.54, 1672.91, 2465.19, 3494.65, respectively. PSD analysis was carried out by gating the targeted ion and increasing the laser irradiance to optimise the production of fragments.

5.2.3.2 AP-MALDI MS/MS

AP-MALDI-QIT MS/MS was carried out on tryptic peptides of m/z identified as being discriminatory between melanoma stage IV and control samples by ANNS. AP-MALDI experiments were carried out on a Finnigan LCQ ion trap (ThermoFinnigan, San Jose, CA, USA) fitted with a MassTechnologies (Burtonsville, MD, USA) AP-MALDI ion source. The heated capillary was maintained at 350 °C, and the ion trap AGC function was deactivated. The trapping time was set manually to 500 ms and no sheath or auxiliary gas was used. Typical collision energy used for MS/MS was between 35-38 %.

5.2.3.3 LC-ESI MS/MS

Tryptic serum samples were prepared as detailed in Section 5.2.2. The resulting 4 μ l elute from the ZipTip clean-up was diluted with 0.1 % TFA (100 μ L). The samples were placed

in glass vials with low-volume inserts (Chromacol, UK) and transferred to an LC Packings Famos autosampler (Dionex Ltd, UK). Samples were loaded (100 μL) onto a LC Packings precolumn (C_{18} -PepMap, 100 \AA , 3 μm particle size, 300 μm ID x 15 mm, Dionex UK) connected to a six-port automated switching valve (Figure 5.2), using a loading mobile-phase (0.1 % TFA, 30 $\mu\text{L min}^{-1}$), supplied by a loading pump (Dionex Ltd, UK), to effect sample pre-concentration and desalting (total of 14.0 min.). The path of the six-port valve was switched to introduce a counter-current solvent flow (180 nL min^{-1} , split flow) to the pre-column *via* an LC Packings UltiMate gradient pump (Dionex Ltd, UK), to transfer the sample onto an LC Packings reverse-phase capillary LC column (75 μm ID x 150 mm, C_{18} -PepMap, 100 \AA , 3 μm particle size, Dionex Ltd, UK) connected to the mass spectrometer ESI interface by a fused silica transfer line (280 μm ID, 300 mm).

On-line sample separation prior to mass spectrometric detection was carried out using a linear gradient (Solvent A: 0.1 % formic acid in water; Solvent B: 80 % acetonitrile in 0.1 % formic acid in water) from 5% B (at time of six-port valve switch) to 75% B (over 60 min., hold for 2 min.) then to 5% B (over 2 min.) and re-equilibrate for the next analytical run.

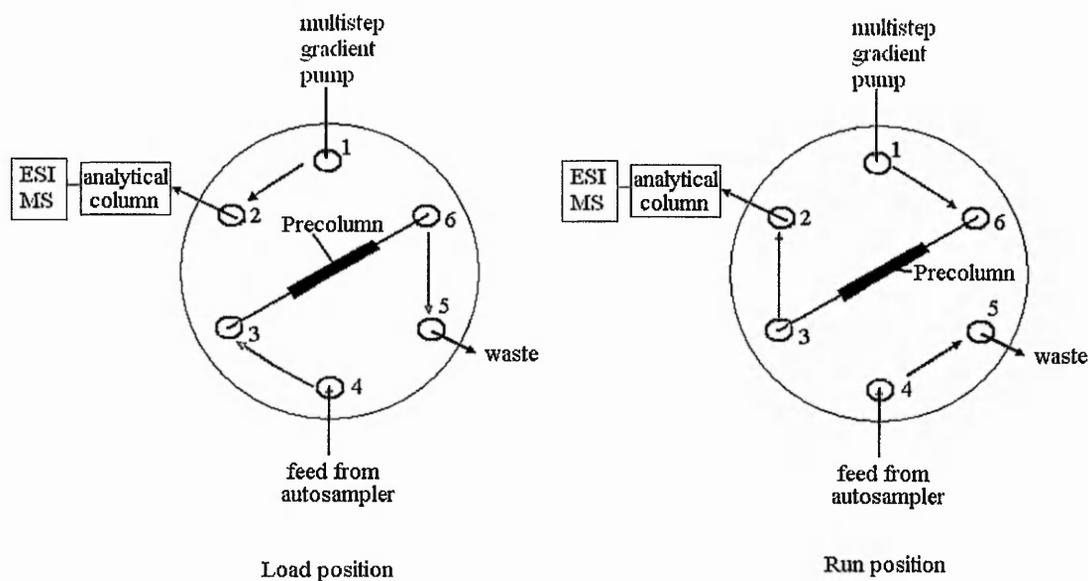


Figure 5.2: Schematic of a six-port switching valve

LC-ESI-MS/MS was carried out using an LCQ Classic ion trap mass spectrometer (ThermoFinnigan, San Jose, CA, USA) operated in positive ion mode equipped with a dedicated nano-electrospray ion source. Analytical performance of the hyphenated LC/MS/MS system was assessed by analysis of a BSA tryptic digest standard, bracketing replicate serum samples, to provide validation for human serum analysis. Data were acquired for human serum (both control and cancer patient samples) following LC introduction using full scan mode (m/z 300-2000, 3 microscans) and targeted tandem mass spectrometry (MS/MS) mode (200 ms activation time, isolation m/z of 3.0, collision energy 38 %). The ions identified as being discriminatory between cancer and control by ANNs were selected for MS/MS analyses, both the singly charged and doubly charged species were targeted. Automatic Gain Control (AGC) was applied in all data acquisition modes.

5.2.4 Database Searching

An automated database search of fragment ion spectra was carried out using the SEQUEST algorithm (ThermoFinnigan, San Jose, CA, USA). Parameters were set to assess tryptic digest peptides with m/z tolerances of ± 0.9 Da for the parent ion and ± 0.2 Da for fragment ions. A total m/z range of 200-3500 was specified, with a minimum total ion chromatogram count of 5×10^4 counts. Confident peptide assignment from SEQUEST output satisfied a confidence level (xC) value > 2.0 and probability score (Sp) > 200 , matching a minimum of 4 product ions above m/z 400 upon manual inspection. LC-MS/MS and SEQUEST searches were carried out by Dr A. Miles, Nottingham Trent University.

All data were also searched against the NCBIInr database, using MASCOT (Version 2.0, MatrixScience, London). Data were submitted to MASCOT with the following peptide mass tolerances: MS/MS ion search, peptide tolerance ± 0.8 Da and MS/MS tolerance ± 0.6 Da.

5.2.5 ANNs Analysis

Raw data was exported from the mass spectrometer as ASCII text files. To reduce the number of data points the data were binned across the 800-3500 Da mass range. The median intensity was determined across each 1 Da range to give an intensity value for each mass. Data were normalised (the highest intensity having a value of 1), where raw values were scaled linearly so that the smallest value in the dataset was the minimum, and the largest the maximum. This scaling method ensured that all potential relationships amongst variables were kept identical; therefore no bias was introduced into the data.

A three layer ANNs model was used, containing an input, output and hidden layer. The input layer consisted of the m/z values between 800-3500 Da with their corresponding intensity value. The hidden layer performed the statistical analysis and the output layer consisted of a single node, where control samples were coded as "1", and Stage IV melanoma samples were coded as "2".

During ANN modelling the data were randomly split into three sets, training, test and validation. The training set consisting of 60 % of the data was used to train the ANNs model to distinguish between stage IV melanoma samples and control samples. The test set (20 % of the data) was used to monitor the ANNs performance with regards to predictive performance. The validation set (remaining 20 % of the data) was then analysed.

The modelling process involved a novel stepwise approach.⁴¹ Initially, each variable from the dataset (in this case m/z value and corresponding intensity) was used as an individual input in a network, thus creating n (2700) individual models. The inputs (m/z) were ranked in ascending order based on the mean squared error values and the model which performed with the lowest error selected for further training. The input which performed with the lowest error was selected for the next step. Remaining inputs were then sequentially added to the previous best input, creating $n-1$ models each containing two inputs. Training was repeated and performance evaluated. The model which showed the best capabilities to model the data was then selected and the process repeated, creating $n-2$ models each

containing three inputs. This stepwise process is shown in Figure 5.3. This process was repeated until no significant improvement was gained from the addition of further inputs resulting in a final model containing the proteomic pattern which most accurately predicted between the two outcomes. ANNs analysis was carried out by Graham Ball and Lee Lancashire at Loreus Ltd, Nottingham, UK.

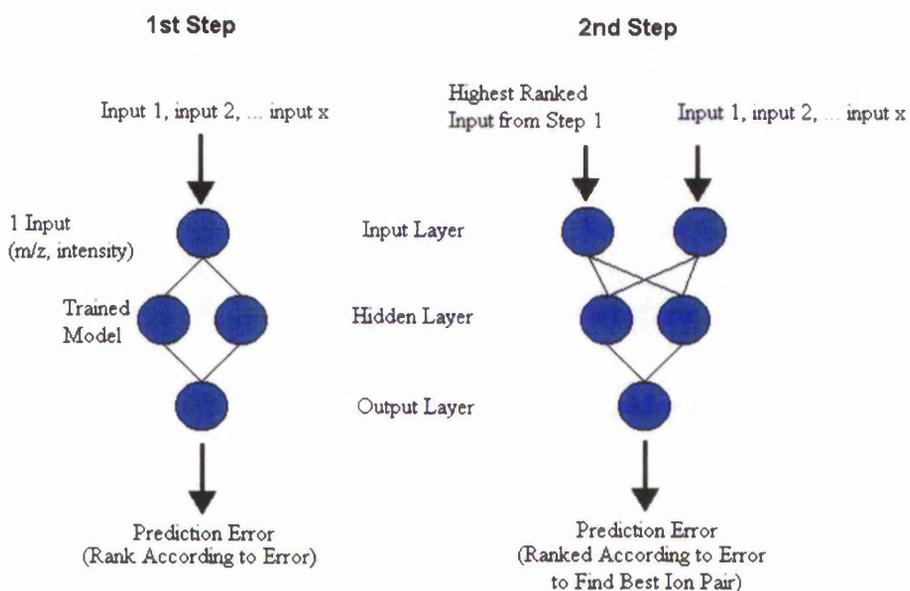


Figure 5.3: Diagram showing the ANNs stepwise model

5.3 Results

5.3.1 Identification of Melanoma Stage IV Biomarkers: Native Serum Peptides

Sera samples from 50 stage IV melanoma patients and 50 healthy control patients, analysed using the methodology developed in Chapter 4, showed reproducible variations between the two states. Examples of MALDI-TOF mass spectra obtained for six representative native sera samples (3 cancer stage IV and 3 controls) processed as described in Section 5.2.2, without tryptic digestion, are shown in Figure 5.4. Spectral differences between melanoma stage IV and control sera were observed. The most prominent of these was the peptide ion at m/z 2755 which has a higher intensity in stage IV melanoma patient samples in comparison with healthy control patient sample indicating up-regulation of the parent protein. These peaks are highlighted in Figure 5.4 and an expanded mass region is shown in Figure 5.5.

The data were subjected to ANNs analysis to prioritise peptides/proteins, which were differentially expressed in the melanoma samples in comparison with the control samples. The top nine m/z values identified by the ANNs as discriminatory biomarkers between serum from stage IV melanoma patients and healthy controls are shown in Table 5.3, along with test performance values which indicate the ability of the model to correctly classify the blind data. The table lists the biomarker ions in order of significance in distinguishing between the two states, with the first ion at m/z 2755 being the only peptide of high significance which can be identified by visual inspection of the spectra. This ion alone was able to predict stage IV melanoma with 67 % accuracy using the ANNs model, although inclusion of further ions (> 6 identified ions in total) led to 99 % accuracy in the ANNs model. A peak at m/z 1206 which was observed to show intensity variations between the two states in some samples by visual inspection (Figures 5.4 and 5.6), was identified by ANNs analysis but was not ranked in the top 40 significant ions. A possible reason for this low significance rating may be because the ion was observed in 37 % of control samples and 54 % of melanoma samples, whereas other more significant biomarker ions were only

observed in stage IV melanoma samples or control samples, for example m/z 2755 which was only present in melanoma samples. Therefore caution needs to be exercised when interpreting spectral variations or ANNs outputs between states in the search for biomarkers. As these observed differences may not translate into predictive differences between cancerous and control samples. Bioinformatics looks at the interaction between a set of ions and correlates this with disease states.

Mass spectra also show that the ion of highest significance at m/z 2755 was observed to be up-regulated in stage IV melanoma samples, the identity of this ion could not be achieved by AP-MALDI-QIT MS/MS or LC-ESI-MS/MS because it lies outside the normal operating range (1-2000 Da) of the LCQ instrument. Further work would be required to analyse the peptide by another MS/MS analyser, for example a qTOF mass spectrometer.

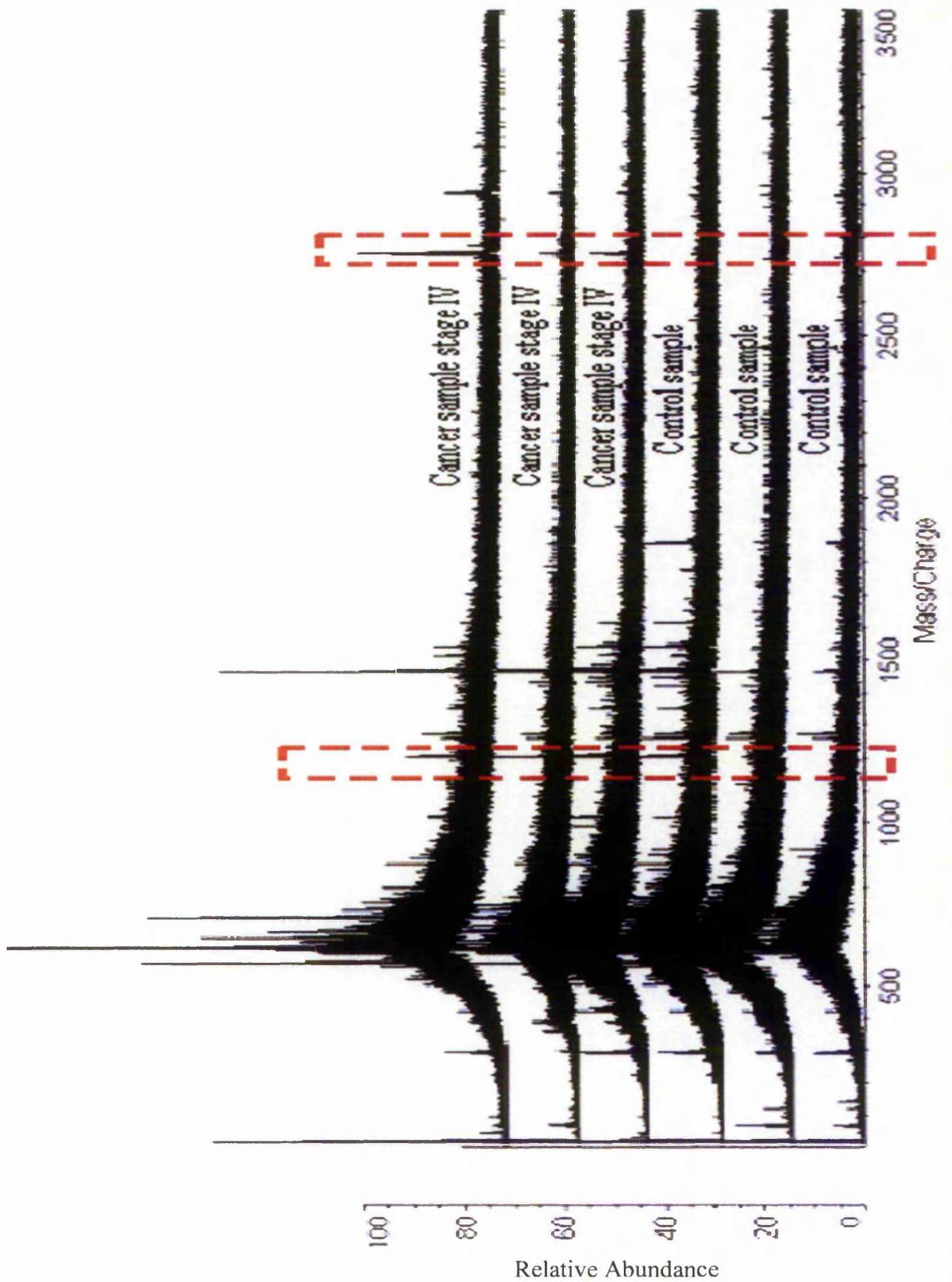


Figure 5.4: MALDI-TOF mass spectra (mass range 1-3500 Da) obtained for the analysis of stage IV melanoma native serum peptides and control serum.

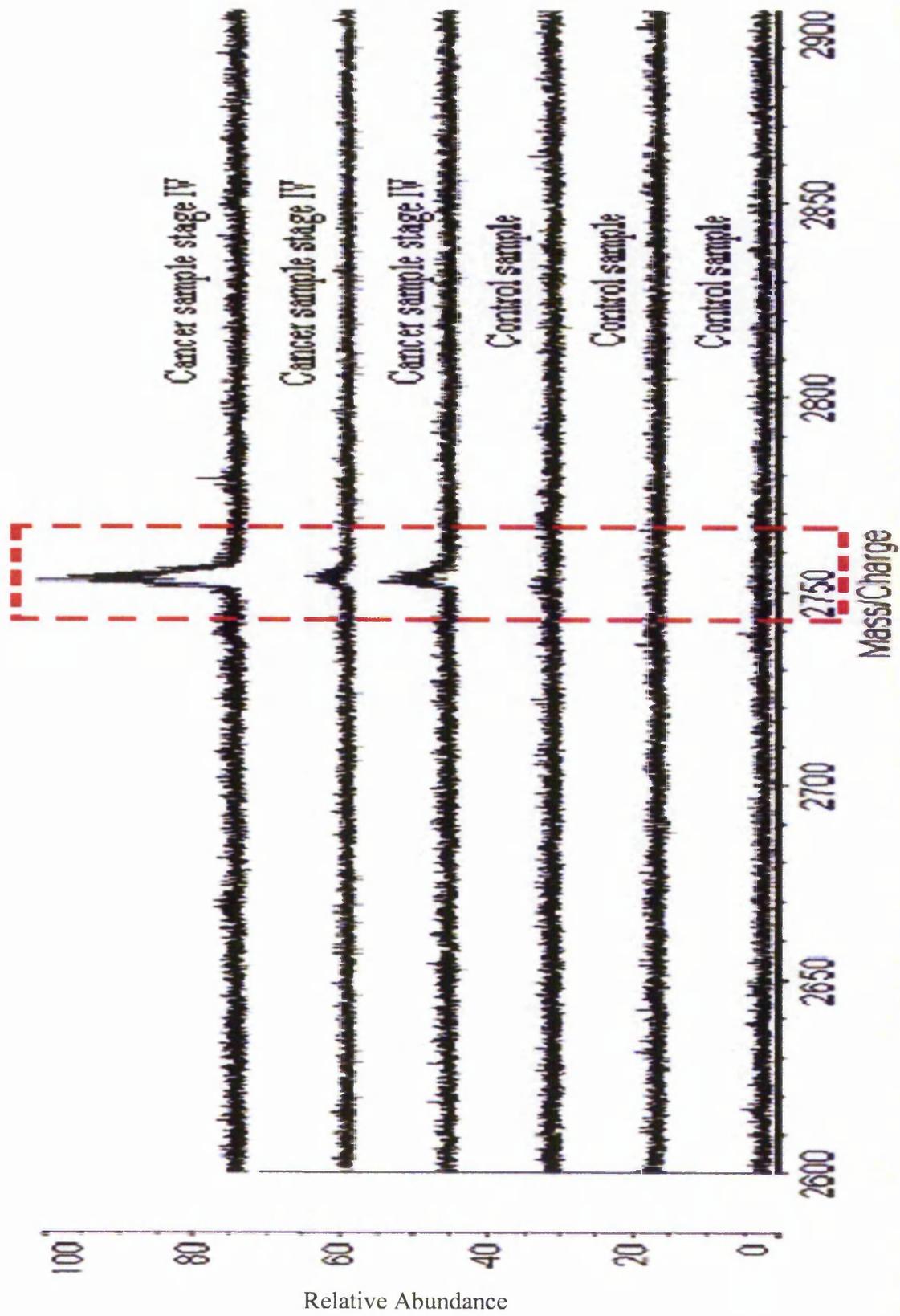


Figure 5.5: MALDI-TOF mass spectra (mass range 2600-2900 Da) obtained for the analysis of stage IV melanoma native serum peptides and control serum.

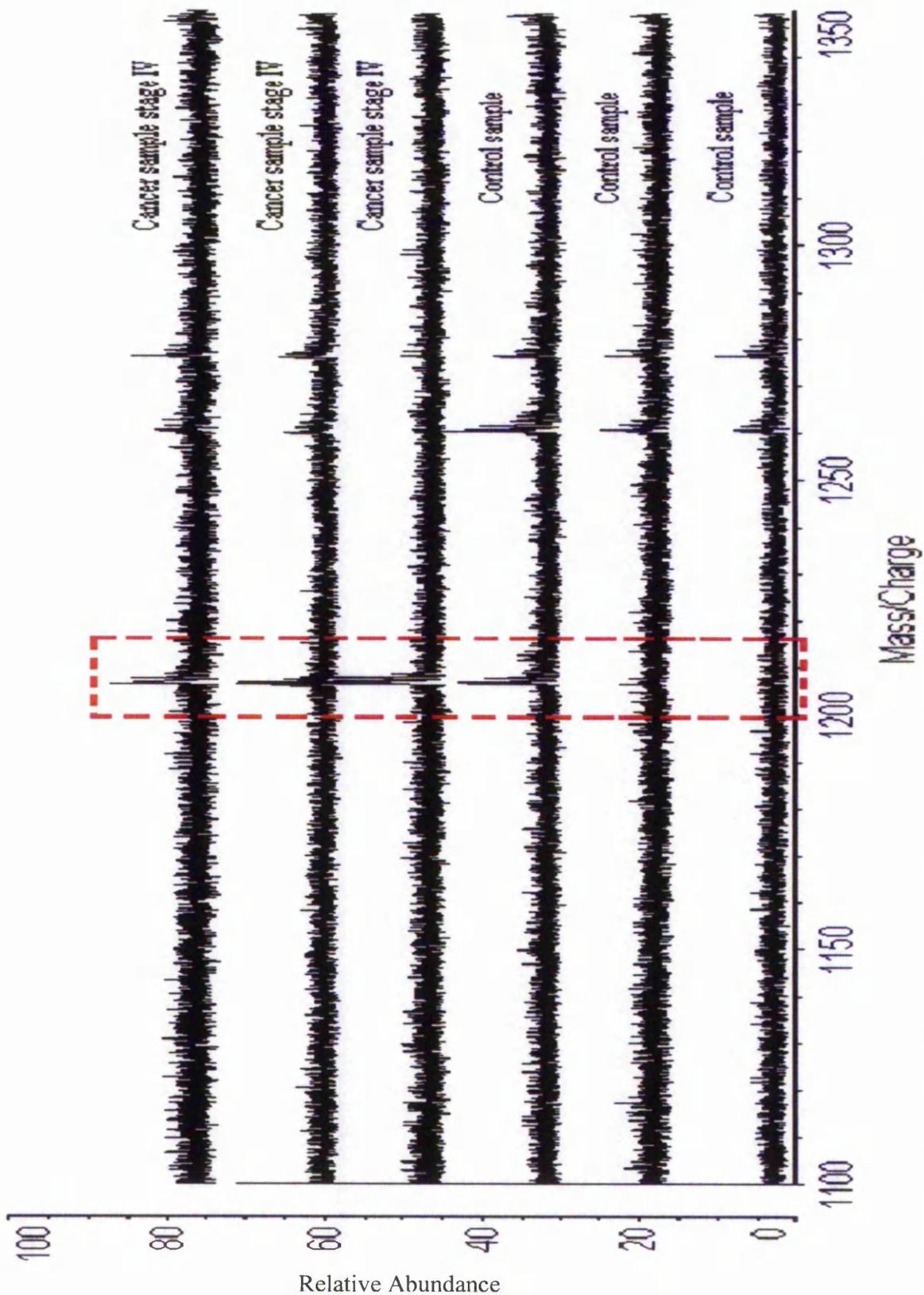


Figure 5.6: MALDI-TOF mass spectra (mass range 1100-3500 Da) obtained for the analysis of stage IV melanoma native serum peptides and control serum.

Table 5.3: Native peptide biomarker ions predicted by ANNs as discriminating between cancer and control sera

<i>m/z</i> values	Test Performance /%
2755	66.999
2998	92.684
1433	95.966
2740	96.484
906	98.428
1154	97.889
3259	98.874
1432	99.021
1737	99.347

The ANNs model performance is shown in Figure 5.7. As the number of ions added to the ANNs model was increased the median accuracy increased and the mean squared error decreased. The addition of ions to the model was stopped once an accuracy of 99.3 % was achieved. This level was reached with the incorporation of the nine biomarker ions identified in Table 5.3.

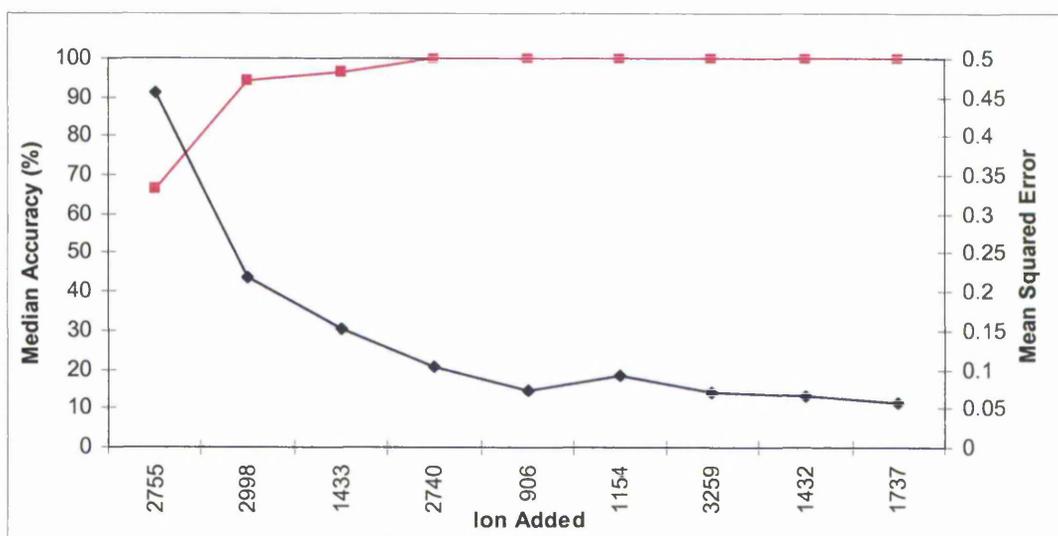


Figure 5.7: Performance of the ANNs model for native serum peptides showing variation of median accuracy (■) and mean squared error (●)

The methodology developed for the mass spectral profiling of native serum peptides, reported in Chapter 4, was used in combination with ANNs to successfully distinguish between stage IV melanoma and control sera. This demonstrated the potential of mass

spectrometry for the rapid screening of patients using a simple blood test and the biomarker ions identified in this work. Although a number of ions were observed as being significant in classifying disease samples, the limited mass range of the LCQ quadrupole ion trap prevented identification of the most significant native peptide (m/z 2755). Therefore tryptic serum peptides were also analysed because the peptide fragments are expected to fall within the mass range of the LCQ, allowing tandem mass spectrometry to be carried out yielding peptide and protein identification.

5.3.2 Identification of Melanoma Stage IV Biomarkers: Tryptic Serum Peptides

Reproducible differences were observed in the MALDI-TOF mass spectra for the tryptic peptides derived from melanoma stage IV patient serum samples and those derived from control serum samples (prepared using the procedure described in Section 5.2.2.) Some of these differences can be visualized in the spectra, shown in Figure 5.8 and the expanded spectral regions shown in Figure 5.9. For example, the most notable visual difference being the presence of ions at m/z 1160 and 1753, which were significantly up-regulated in the stage IV melanoma patient samples in comparison with the control samples. These two biomarker ions were classified by ANNs as significant with 92.4 % predictability between blinded melanoma stage IV and control patients. Other biomarker ions were only detected through the use of ANNs. Table 5.4 lists the top six m/z values, in priority order, as deemed significant by the ANNs in distinguishing between stage IV melanoma and control samples. Incorporation of these six ions into the model lead to a 98 % predictability. Figure 5.10 shows the performance for the models at each step of the ANNs analysis. As with the native peptides, the addition of ions led to an overall reduction in the mean squared error associated with the predictive capabilities of the model for blind data and an increase in the median accuracy for samples correctly classified to melanoma stage IV or control sera. These data show that tryptic peptide biomarker ions, identified by MALDI-TOF mass spectral analysis and ANNs processing, may also be used with confidence for screening

patients with suspected stage IV melanoma. However, the larger processing time required for enzymatic digestion makes this approach less amenable to high throughput screening than the analysis of native serum biomarkers.

The advantage of this bottom-up proteomics approach is that it opens up the potential for sequence analysis of peptide biomarker ions identified as important discriminatory factors by the ANNs. This was attempted initially by MALDI-TOF using PSD to give partial sequence information. The PSD spectrum for m/z 1753 is shown in Figure 5.11. Searching the data *via* MASCOT using the NCBI nr database with a peptide mass tolerance of 0.8 Da and MS/MS tolerance of 0.6 Da yielded a match to the peptide YVGGQEHFALLILR derived from α -1-acid glycoprotein 1 precursor protein with a score of 48. Significant values (>36) are 95% certain to be peptides derived from the protein. Although PSD analysis resulted in an identification of the precursor ion at m/z 1753 the spectrum was of poor quality and therefore confirmation of the identification was carried out using alternative methods.

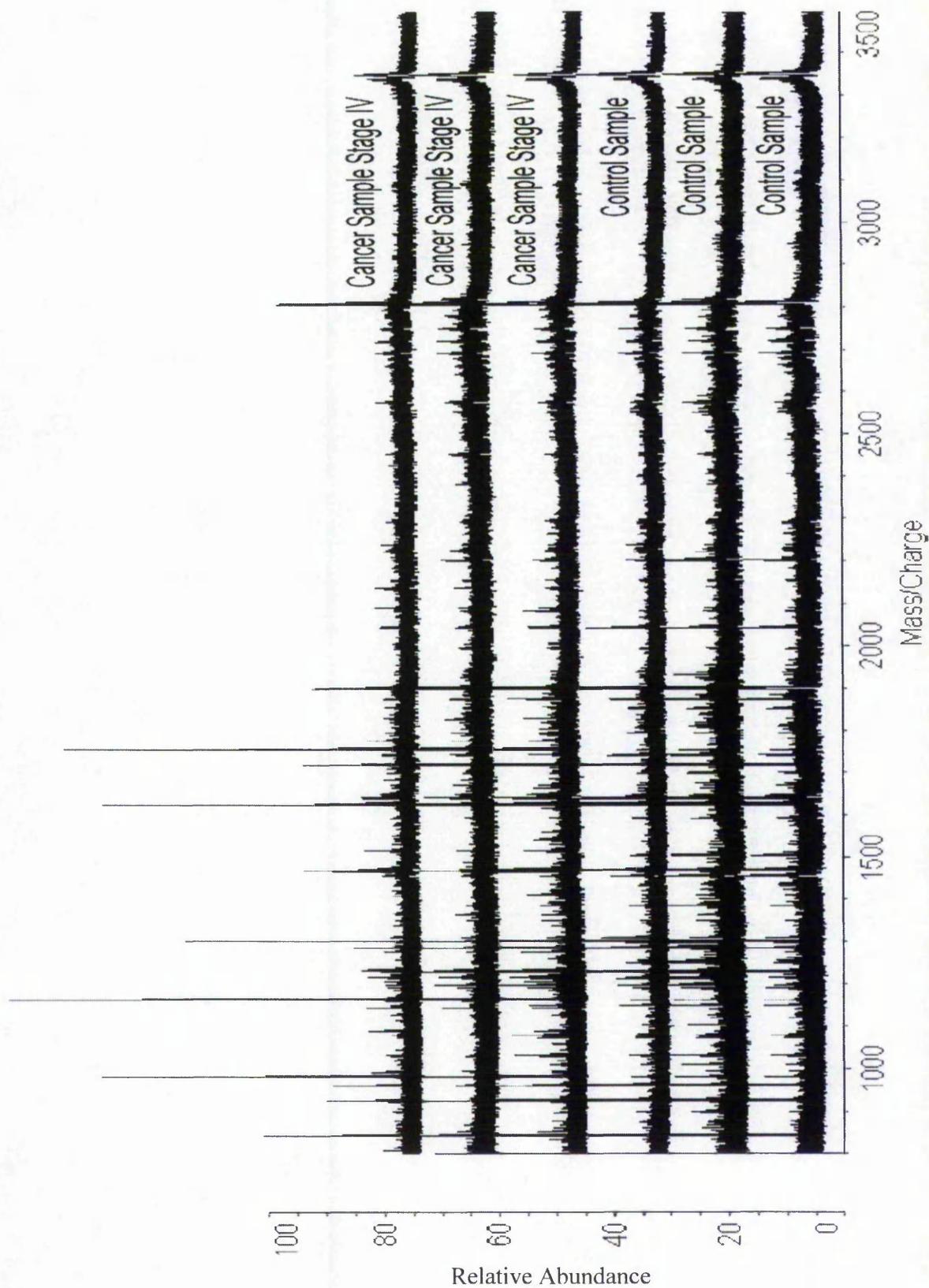


Figure 5.8: MALDI-TOF spectra (mass range 800-3500 Da) obtained for tryptic peptides derived from stage IV melanoma serum samples and controls.

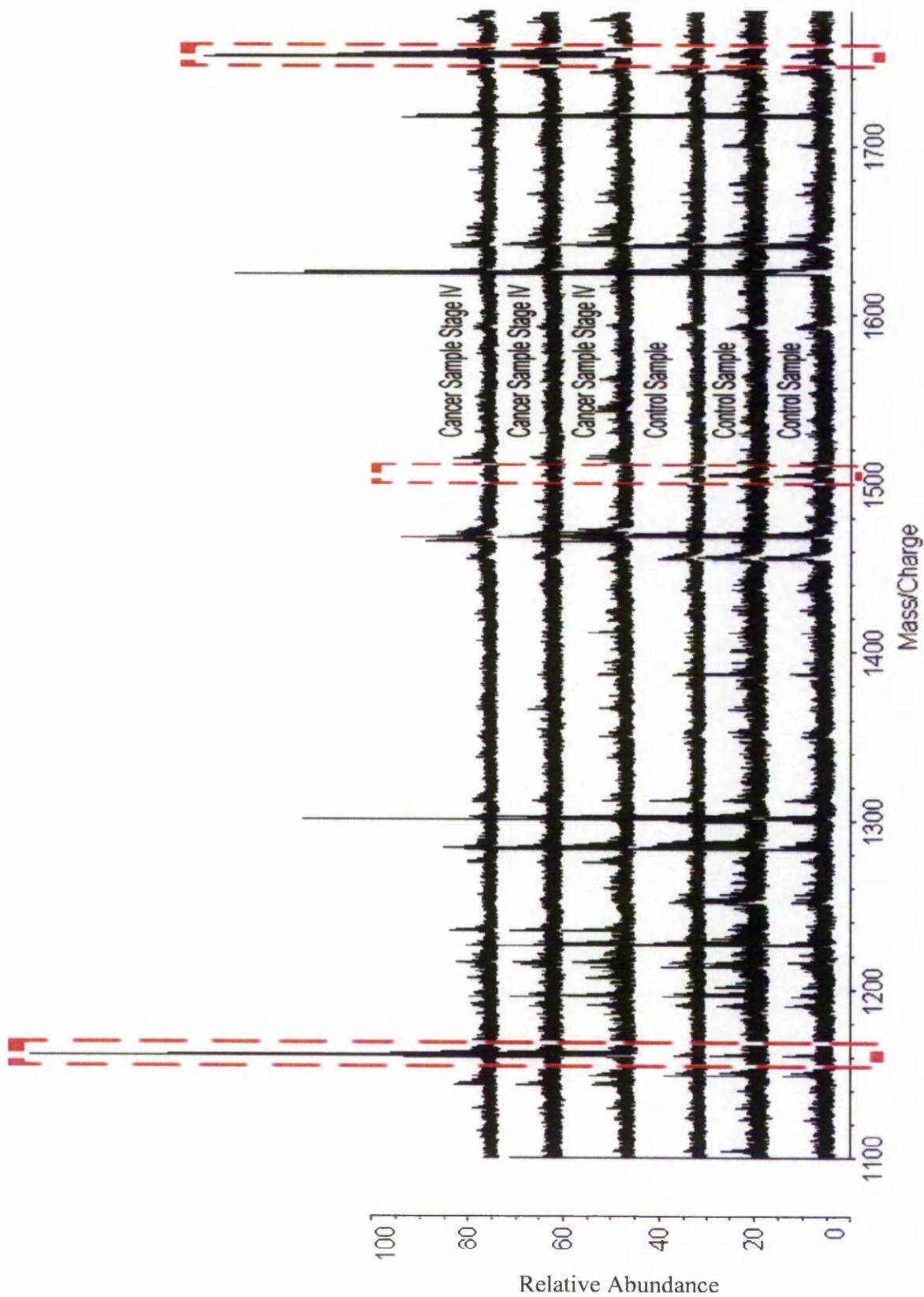


Figure 5.9: MALDI-TOF spectra (mass range 1100-1780 Da) obtained for tryptic peptides derived from stage IV melanoma serum samples and controls.

Table 5.4: Tryptic peptide ions predicted by ANNs as discriminating between cancer and control

<i>m/z</i> value	Test Performance /%
1753	78.077
1161	92.353
1505	94.747
854	97.104
1444	98.387
1093	98.492

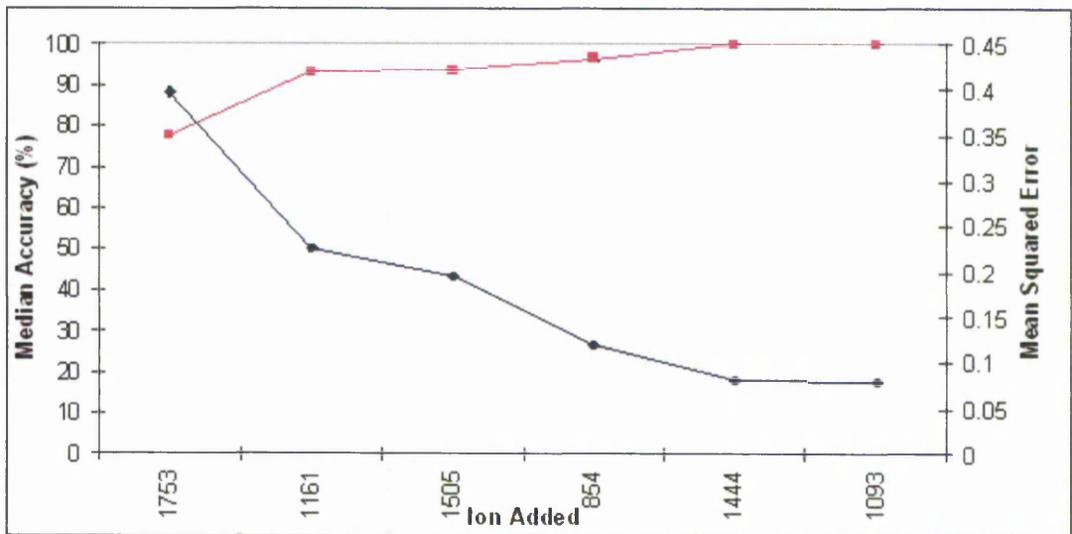


Figure 5.10: Performance of the ANNs model for tryptic serum peptides showing variation of median accuracy (■) and mean squared error (◆)

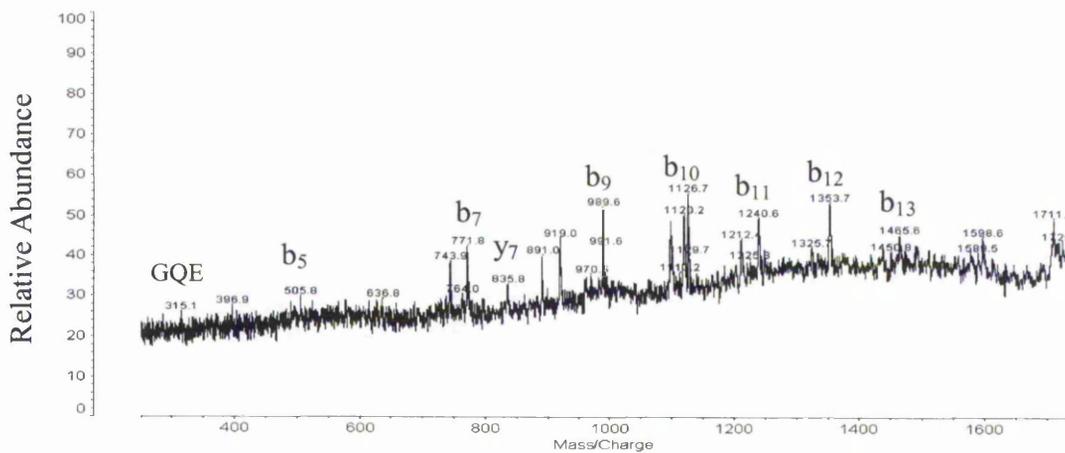


Figure 5.11: MALDI-TOF PSD of serum peptide *m/z* 1753, identified by MASCOT as peptide YVGGQE^HFAHLLIL^R derived from α -1-acid glycoprotein

MS/MS data obtained for the m/z 1753 precursor ion using AP-MALDI-QIT, Figure 5.12, were submitted to MASCOT. This resulted in a higher confidence of assignment (score 50, significance level > 36) for the identification of the precursor peptide ion at m/z 1753 as the α -1-acid glycoprotein 1 precursor protein derived tryptic peptide YVGGQEHFAHLLILR. The presence of this protein was further confirmed by the identification of precursor ion m/z 1160.5 as peptide WFYIASAFR derived from α -1-acid glycoprotein 1 precursor by AP-MALDI-QIT MS/MS. However, none of the other targeted masses (Table 5.5) led to protein identification due to the low sensitivity or quality of the AP-MALDI product ion spectra or loss of water without further fragmentation. Due to this limitation, which is frequently observed for singly charged peptide ions, the samples were analysed using capillary LC-ESI-MS/MS.

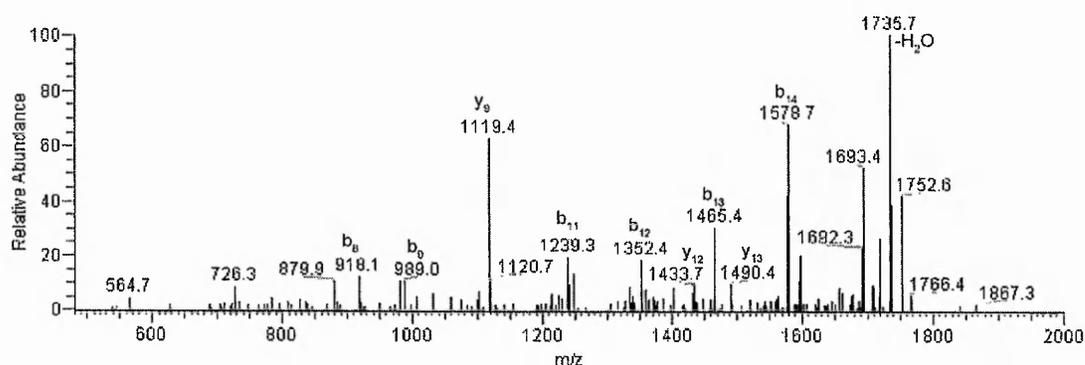


Figure 5.12: AP-MALDI/QIT MS/MS of serum peptide m/z 1753, identified by MASCOT as peptide YVGGQEHFAHLLILR derived from α -1-acid glycoprotein

Table 5.5: SEQUEST and MASCOT results for MALDI PSD, AP-MALDI MS/MS and LC-MS/MS analysis of tryptic serum peptides (Results for $[M+2H]^{2+}$ shown in brackets)

m/z	Sequence	MALDI PSD	AP-MALDI MS/MS	LC-MS/MS			ID
		MASCOT score	MASCOT score	MASCOT score	SEQUEST Xcorr	Sp	
1753 (877)	YVGGQEHFAHLLILR	48	50	48 (61)	3.29 (4.47)	726 (1820)	α -1-acid glycoprotein 1 precursor
1160 (580)	WFYIASAFR	33	14	41 (43)	2.1 (2.92)	450 (999)	α -1-acid glycoprotein 1 precursor
1093 (547)	NTLIYLDK	-	-	41 -	2.44 (2.08)	664 (556)	Complement C3 precursor

The LC-ESI-MS/MS selected ion chromatogram (SIC) and mass spectrum for the precursor ion m/z 1753 is shown in Figure 5.13. The SIC shows a single strong peak with a retention time of 41.25 minutes eluted off the column. The associated MS/MS spectrum (Figure 5.13) shows good sequence coverage for the VGGQE^HFAHLLILR peptide of the α -1-acid glycoprotein.

Using data-dependent LC runs on the melanoma stage IV and control sera the identification of three out of the six highlighted biomarker ions from the ANNS analysis was achieved using singly and doubly charged precursor ions (Table 5.5). SEQUEST Xcorr values >2.0 with a probability score (Sp) >200 indicate significant homology. The LC-ESI-MS analysis was validated using a bovine serum albumin (BSA) tryptic digest as a control. PMF data yielded a MASCOT score of 73 (indicating significant homology) with sequence coverage of 18 % for the 10 matched peptides.

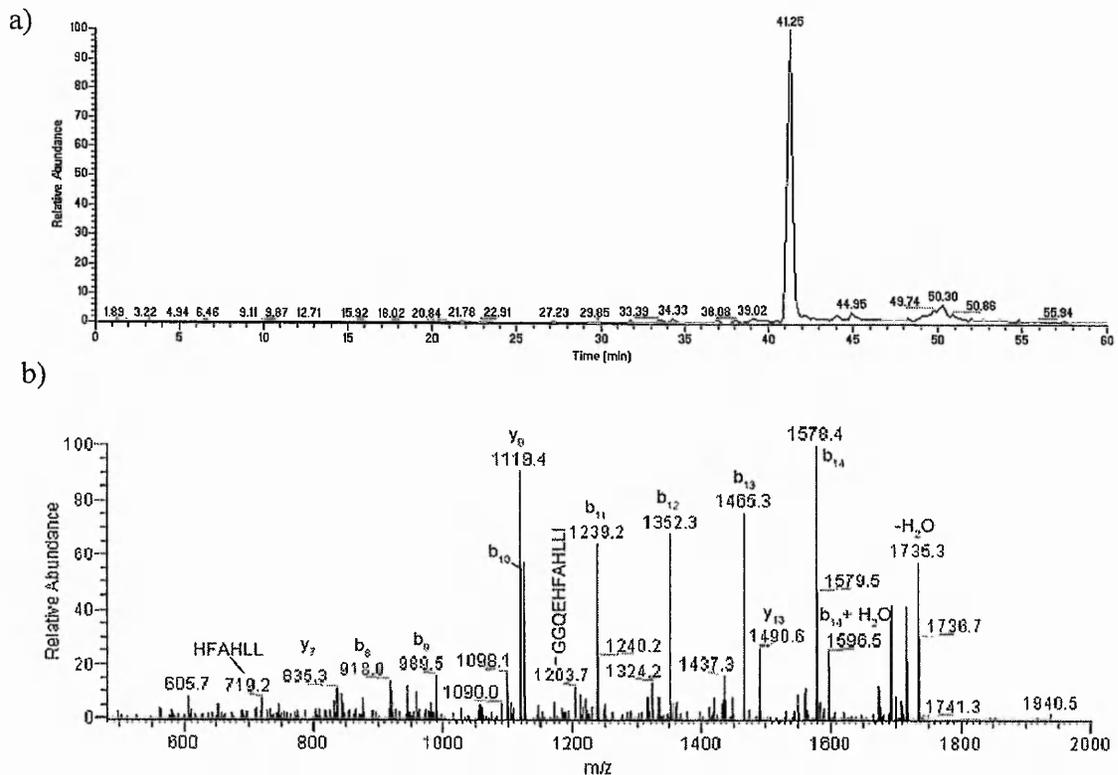


Figure 5.13: a) SIC LC-ESI-MS/MS for the precursor ion at m/z 1753 b) LC-ESI-MS/MS of serum peptide 1753, identified by SEQUEST and MASCOT as peptide YVGGQE^HFAHLLILR derived from α -1-acid glycoprotein

The LC-ESI-MS/MS spectra for the precursor ions at m/z 1753 and 1160 were identified by MASCOT as peptides YVGGQEHLFAHLLILR and WFYIASAFR respectively, derived from α -1-acid glycoprotein 1 precursor. The same protein identification was obtained *via* SEQUEST for both the singly and doubly charged species targeted in the LC-ESI-MS/MS acquisition. This confirmed the protein identification and validated the MALDI-TOF PSD and AP-MALDI MS/MS results. Using LC-MS/MS a further biomarker ion, m/z 1093, was identified by both MASCOT and SEQUEST as the tryptic peptide NTLIIYLDK derived from the complement C3 protein. The MS/MS spectra obtained for precursor ions at m/z 1505.4, 854.53 and 1444.50 did not give sufficient information for identification.

The method described in Chapter 4 has been applied to tryptic peptides, showing that mass spectrometry can be used in combination with ANNs to identify stage IV melanoma. Tryptic peptides gave high performance values from 78.0% with a single ion, up to 98.5% for a six ion model.

5.4 Discussion

The results presented in this Chapter show the direct MALDI-TOF analysis of native and tryptic serum peptides combined with advanced bioinformatic strategies to yield cancer biomarker ions characteristic of disease state. The importance of profiling the serum peptidome rather than proteome is that sequence identification of the biomarker peptides is possible through PSD or tandem mass spectrometry yielding the identification of the parent proteins. This approach contrasts to the MALDI or SELDI analysis of intact proteins used in the direct proteomic profiling which, as previously mentioned, yields protein biomarker ions, but not protein identities for high mass proteins.^{25, 28, 35-41} A further advantage of the direct analysis of native or tryptic serum peptides without lengthy gel electrophoretic or HPLC separation is the potential for high throughput clinical screening.

The application of the novel peptide methodology reported in this work to profile native and tryptic serum peptides enabled visual differences in ion intensities to be identified in the region 800-3500 Da, in MALDI-TOF mass spectra obtained for stage IV melanoma patient serum samples compared to control samples. Any differences observed between diseased state and healthy patient sera would be expected to be seen in the stage IV and be more pronounced than in stages I, II and III.

Results reported in Section 5.3.1, showed that native peptides can be analysed by MALDI-TOF MS and used to distinguish between melanoma and control samples. Data showed reproducible variations between the two states, for example, the native peptide observed at m/z 2755 was only apparent in melanoma stage IV sera samples and absent in controls (shown in Figure 5.5). Data were subjected to stepwise ANNs analysis and nine biomarker ions were identified by the bioinformatics as significant ions in discriminating between melanoma stage IV and healthy control sera (shown in Table 5.3).

As with the work reported in Section 5.3.1, previous groups have reported profiling of native melanoma serum peptides, without tryptic digestion using MALDI-TOF MS.^{24, 33, 42} For example Ferrari *et al*³³ reported a three peak pattern consisting of m/z 2761, 3089 and 3455 only observed in melanoma sera samples compared with healthy controls. However, the cohort of samples was small, with only seventeen sera from various stages of malignant melanoma and fourteen control sera.

Caputo *et al*³⁴ studied native plasma peptides (without tryptic digestion) from melanoma and breast cancer patients. Plasma samples were pre-fractionated using HPLC to remove the high salt content from the anticoagulants used, with resulting eluates being collected manually and deposited onto SELDI chips. Analysis was carried out using TOF MS. Ions at m/z 2023, 2039 and 2053 were primarily observed in melanoma plasma samples.

Interestingly there is no overlap with the m/z values reported by Caputo, Ferrari and the data presented in this thesis for native serum peptides (Section 5.3.1). The closest biomarker ion is the m/z 2761 ion identified by Ferrari, which is close to the m/z 2755 ion observed in this work. Although, the mass tolerance of the work reported in this thesis is ± 1.6 at the 95 % confidence level, due to the lower mass accuracy of linear TOF, these ions may be associated with the same peptide. This variation in the discriminatory ions could be due to differing sample preparation techniques. As previously reported in Chapter 4 of this thesis, the use of ultrafiltration devices was studied and deemed unsuitable for serum analysis due to contaminant leaching from the microcon devices. However, Ferrari *et al*³³ used ultrafiltration to prepare samples prior to mass spectrometric analysis and this is likely to result in differences in the observed mass spectra. Different peptides are detected with different matrices, for all MALDI-TOF experiments reported in this thesis, analyses were carried out using the reflectron mode and α -cyano-4-hydroxy-cinnamic acid (CHCA) matrix. In contrast Ferrari *et al*³³ performed analyses using the linear mode and Sinapinic acid (SA) matrix. This may be another contributing factor to the variation in discriminatory ions reported.

However, as with the work reported in this thesis, Seraglia *et al*²⁴ analysed low molecular weight native peptides in serum using MALDI-TOF MS operated in reflectron mode with CHCA matrix. Samples obtained from patients with malignant melanoma stage III were analysed, (stages of melanoma are outlined in section 5.1.1) along with samples from healthy volunteers. Mass spectral variations were observed between healthy serum and stage III melanoma samples, with 18 ionic species of possible diagnostic value only detected in cancerous samples and absent in controls. Again, there was no correlation with ions highlighted in this thesis as discriminatory between stage IV melanoma and healthy control serum samples.

Many studies have reported only visual mass spectral differences between disease states rather than full identification (sequence information) of such discriminatory ions. In contrast to the work reported in this thesis, small sample numbers are commonly analysed and no bioinformatic tools have been used to discriminate between cancer and controls. A complex data mining algorithm was applied in the study reported here to deconvolute the data, therefore it is not surprising that the studies detailed above failed to identify the same biomarkers, particularly ions which are not easily identified visually. Some studies have used statistical analysis of the data to distinguish between disease states with high sensitivity and specificity on the basis of protein mass spectral profiles.²⁵ These discriminatory patterns have the potential to be used as diagnostic markers in cancer.

The analysis of native serum peptides reported in this Chapter, resulted in 99 % accuracy in discriminating between sera from melanoma stage IV patients and control sera. However, tryptic serum peptides were also investigated because of prior success of protein identification from proteolytic digests reported in the literature⁴³⁻⁴⁵ and peptide fragments resulting from digestion are within the normal mass range of the LCQ ion trap (1-2000 Da) allowing tandem mass spectrometry to be performed for protein identification, which has not been included in previous studies. As reported by Koomen *et al*⁴⁵ data were improved,

in terms of the S/N ratio for the mass spectra acquired for tryptic serum peptides in comparison with native serum peptides.

ANNs analysis of the mass spectral data resulted in the identification of three tryptic peptides capable of distinguishing melanoma stage IV sera from control sera with high confidence (> 92.3 %). Sequence analysis of these biomarker ions was achieved using MALDI-TOF PSD, AP-MALDI MS/MS and capillary LC-ESI-MS/MS. The top two tryptic peptide ions at m/z 1753 and 1160 were found to be sequences YVGGQEHFAHLLILR and WFYIASAFR respectively derived from α -1-acid glycoprotein precursor 1 or 2 (AGP). The third precursor ion identified (m/z 1093) was classified as the tryptic peptide NTLIIYLDK from the complement component 3 protein. The significance of these markers is discussed below.

α -1-acid glycoprotein (AGP, orosomucoid) is a highly glycosylated, acidic acute phase protein, very soluble in aqueous media. Changes in the extent of glycosylation and the carbohydrate structure of proteins has been shown to correlate with cancer and other disease states highlighting the clinical importance of this modification as an indicator of pathologic mechanisms.^{46, 47} Human serum concentrations of AGP are reported to range between 0.4-1.1 g L⁻¹⁴⁸ and increase up to 5 fold during acute-phase reactions such as cancer, inflammation and pregnancy.⁴⁹ Although AGP is an abundant protein its biological function remains unknown.⁵⁰

Human AGP consists of a single polypeptide chain 183 amino acids in length, having a molecular mass of 41–43 kDa when examined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis.⁵¹ The protein is approximately 45% glycosylated with the carbohydrate chains being attached at the protein core by five to six complex-type-*N*-linked glycans.⁵² Various forms of di-, tri-, and tetraantennary carbohydrate chains contribute to the human AGP glycan complexity. Significant heterogeneity of the oligosaccharides also exists due to variation in the linkage of the

terminating sugars on glycan chains to galactose (Gal) and due to the presence of fucose (Fuc) residues.⁵⁰ The high diversity of structure gives rise to many AGP glycoforms, approximately 10-20 are observed in human serum.⁵⁰

Interestingly, the AGP tryptic peptide at m/z 1753, (YVGGQEHFALLILR) was also identified by Koomen *et al*⁴⁵ in plasma from pancreatic cancer patients. As with the work presented in this Chapter, samples were tryptically digested and an initial screening procedure carried out using MALDI-TOF MS. The resulting PMF data were subjected to statistical analysis and a list of target ions generated. These significant ions were then identified using MALDI-qTOF and MALDI-TOF-TOF tandem mass spectrometry and MASCOT database searching. These findings support those in this thesis and show that the use of proteolytic digestion methodology to analyse the peptidome is an effective route to profiling differences between disease states, both in plasma and sera.

Other studies have also shown that increased levels of AGP are found in cancer.⁵³⁻⁵⁵ MALDI-TOF mass spectrometry has shown the up-regulation of AGP in breast cancer patients by analysis of proteins from nipple aspirate fluid.⁵⁶ Duche *et al*^{54, 55} have also shown that AGP is up-regulated in cancer samples. Plasma from breast, ovarian and lung cancer patients was analysed by an immunonephelometric method.

The significance of the up-regulation of AGP in melanoma stage IV serum samples needs further investigation. However, it has been found to be up-regulated in many types of cancer and thought to mediate growth rate, metastatic potential and various host responses.

57, 58

The second biomarker protein identified was complement C3, an abundant serum protein (approximately 1.2 mg mL^{-1}). This protein, produced within the kidney like AGP,⁵⁹ may be an important mediator of local inflammatory and immunological injury.⁶⁰

In support of the results reported in this Chapter, Caputo *et al*³⁴ identified presence of the peptide SSKITHRIHWESASLL (m/z 1869.5) from complement C3 protein in melanoma

samples. Although this particular peptide was not observed in the spectra obtained for the analysis of native serum samples reported in Section 5.3.1. A tryptic peptide at m/z 1093 was observed by ANNs to be significant in discriminating between melanoma stage IV and control sera and identified by LC-ESI-MS/MS as peptide NTLIIYLDK derived from complement C3 protein.

In order to enhance our understanding of cancer pathology and lead to therapeutic developments, the identification of possible biomarkers is critical. The strength of the methodology developed in this thesis and applied in this Chapter, was the ability to identify discriminatory patterns using sophisticated bioinformatics, and then target these ions using MS/MS, which resulted in protein identification. AGP and Complement C3 protein were identified as being important in discriminating between healthy serum and that from patients with stage IV melanoma.

5.5 Conclusion

In conclusion, the robust methodology developed in Chapter 4, has been used in combination with artificial intelligence algorithms for the identification of serum biomarkers. Biomarker ions have been identified that distinguish between stage IV melanoma and control patients. The data shows that stage IV melanoma can be identified with high sensitivity and specificity based upon MALDI mass spectrometry proteomic profiling. The results have demonstrated that the methodology for native serum peptides can be used to successfully discriminate between sera from melanoma stage IV patients and control sera. However, the main advantage of this novel methodology is the ability to highlight variations in the PMF between the two sample states, cancerous and control, using the ANNs and to selectively target these m/z values to gain sequence and parent protein identification by MS/MS. Three out of the six tryptic peptide ions identified as markers by ANNs in discriminating between cancer and control samples have been identified. Two of the ions were shown to belong to tryptic peptides associated with the same protein α -1-acid glycoprotein precursor, which predicted 95% i.e. (45/50) metastatic melanoma patients. Although many SELDI studies have discriminated between cancerous samples and healthy controls, identification of individual proteins making up these cancer signature profiles has not been possible. Knowledge of the identity of the individual biomarkers leads to a greater understanding of tumour biology and may help in the prognosis and development of targets for novel therapies. These data suggest that peptides in the mass range 800-3500 Da are of particular interest for characterising patients with stage IV melanoma. However, further study is needed to examine different clinical stages of melanoma.

5.6 References

- 1) American Skin Cancer Society, [online] available at <http://www.cancer.org/> [accessed 10/2/05]
- 2) Neal A., Hoskin P., Clinical oncology: A textbook for Students., Edward Arnold, Hodder Headline PLC., London,UK
- 3) Colorado Surgical Advice, 2006 [online] available at http://coloradosurgicalservice.com/article_melanoma.htm [accessed on 14/2/05]
- 4) Brenner S., Tamir E., *Clin. Dermatol.*, **20**, 203-211 (2002)
- 5) Carlson J.A., Slominski A., Linette G., Milhm C.M Jr., Ross J.S., *Expert. Rev. Mol. Diagn.*, **3**, 89-110 (2003)
- 6) Meyers M., Balch C.M., Diagnosis and treatment of metastatic melanoma. In: *Cutaneous Melanoma*. Balch CM., Houghton A., Sober A (Eds), Quality Medical Publishing Inc., St Louis, MO, USA, 325-372 (1998)
- 7) Walsh P., Gibbs P., Gonzalez R., *J. Am. Acad. Dermatol.*, **42**, 480-489 (2000)
- 8) National cancer Institute, [online] available at www.cancer.gov [accessed on 10/2/05]
- 9) Sobin L.H., Wittenkind C.H. (Ed) in TNM Classification of Malignant Tumours, Wiley-Liss Inc, NY, (1997)
- 10) Carlson J.A., Slominski A., Linette G.P., Mihm M.C. Jr., Ross J.S., *Expert Rev. Mol. Diagn.* **3**, 303-330 (2003)
- 11) Solomon E.P., Berg L.R., Martin D.W., Biology, 5th Ed Saunders College Publishing, FL, USA (1999)
- 12) Clark W.H. Jr., Elder D.E., Guerry D., Braitman L.E., Trock B.J., Schultz D., Synnestvedt M., Halpern A.C., *J. Natl Cancer Inst.* **81**, 1893-1904 (1989)
- 13) Clark W.H Jr., Elder D.E., Guerry D.T., *Hum. Pathol.* **15**, 1147-1165 (1984)
- 14) Rosso S., Zanetti R., Pippione M., Sancho-Garnier M., *Melanoma Res.*, **8**, 573-583 (1998)

- 15) Jillella A., Mani S., Nair B., *Proc Am Soc Clin Oncol.* **14**, 413 (1995)
- 16) Sirott M.N., Bajorin D.F., Wong G.Y, Tao Y., Chapman P.B., Templeton M.A., Houghton A.N., *Cancer* **72**, 3091-3098 (1993)
- 17) Eton O., Legha S.S., Moon T.E., Buzaid A.C., Papadopoulos N.E., Plager C., Burgess A.M., Bedikian A.Y., Ring S., Dong Q., Glassman A.B., Balch C.M., Benjamin R.S., *J. Clin. Oncol.* **16**, 1103-1111 (1998)
- 18) Balch C.M., Buzaid A.C., Atkins M.B., Cascinelli N., Coit D.G., Fleming I.D., Houghton A. Jr., Kirkwood J.M., Mihm M.F., Morton D.L., Reintgen D., Ross M.I., Sober A., Soong S.J., Thompson J.A., Thompson J.F., Gershenwald J.E., McMasters K.M., *Cancer*, **88**, 1484-1491 (2000)
- 19) Guo H.B., Stoffel-Wagner B., Bierwith J., Mezger J., Klingmuller D., *Eur. J. Cancer*, **31**, 924 -958 (1995)
- 20) Henze G., Dummer R., Joller-Jemelka H.I., Boni R., Burg G., *Dermatology*, **194**, 208-212 (1997)
- 21) Abraha H.D., Fuller L.C., Du Vivier A.W., Higgins E.M., Sherwood R.A., *Brit. J. Dermatol.*, **137**, 381-385 (1997)
- 22) Fagnart OC., Sindic CM., Laterre C., *Clin. Chem.*, **34**, 1387 (1988)
- 23) Martenson E.D., Hansson L.O., Nilsson B., von Schoultz E., Brahme E.M., Ringborg U., Hansson J., *J. Clin. Oncol.*, **19**, 824-831 (2001)
- 24) Seraglia R.S., Vogliardi S., Allegri G., Comai S., Lise M., Rossi C.R., Mocellin S., Scalerta R., Raggazi E., Traldi P., *Eur. J. Mass Spectrom.*, **11**, 353-360 (2005)
- 25) Petricoin E.F., Ardekani A.M., Hitt B.A., Levine P.J., Fusaro V.A., Steinberg S.M., Mills G.B., Simone C., Fishman D.A., Kohn E.C., Liotta L.A., *Lancet*, **359**, 572-577 (2002)
- 26) Hutchens T., Yip T.T., *Rapid Commun. Mass Spectrom.*, **7**, 576-580 (1993)
- 27) Merchant M., Weinberger S.R., *Electrophoresis*, **21**, 1164-1177 (2000)

- 28) Adam B.L., Qu Y., Davis J.W., Ward M.D., Clements M.A., Cazares L.H., Semmes O.J., Schellhammer P.F., Yasui Y., Feng Z., Wright G.L Jr., *Cancer Res.*, **62**, 3609-3614 (2002)
- 29) Qu Y., Adam B.L., Yasui Y., Ward M.D., Cazares L.H., Schellhammer P.F., Feng Z., Semmes O.J., Wright G.L. Jr., *Clin. Chem.*, **48**, 1835-1843 (2002)
- 30) Petricoin E.F III., Ardekani A.M., Hitt B.A., *Lancet*, 261-266 (2002)
- 31) Wu B., Abbott T., Fishman D., McMurray W., Mor G., Stone K., Ward D., Williams K., Zhao H., *Bioinformatics*, **19**, 1636-191643 (2003)
- 32) Mian S., Ugurel S., Parkinson E., Schlenzka I., Dryden I., Lancashire L., Ball G., Creaser C., Rees R., Schadendorf D., *J. Clin. Oncol.*, **23**, 5088-5093 (2005)
- 33) Ferrari L., Seraglia R., Rossi C.R., Bertazzo A., Lise ., Allegri G., Traldi P., *Rapid Commun. Mass Spectrom.*, **14**, 1149-1154 (2000)
- 34) Caputo E., Lombardi M.L., Luongo V., Moharram R., Tornatore P., Pirozzi G., Guardiola J., Martin B.M., *J. Chrom. B.*, **819**, 59-66 (2005)
- 35) Li J., Zhang Z., Rosenzweig J., Wang Y.Y., Chan D.W., *Clin. Chem*, **48**, 1296-1304 (2002)
- 36) Won Y., Song H.J., Kang T.W., Kim J.J., Han B.D., Lee S.W., *Proteomics*, **3**, 2310-2314 2003)
- 37) Xiao X.Y., Tang Y., Wei X.P., He D.C., *Biomed. Environ. Sci.*, **16**, 140-148 (2003)
- 38) Hingorani S.R., Petricoin E.F., Maitra A., Rajapakse V., King C., Jacobetz M.A., Ross S., Conrads T.P., Veenstra T.D., Hitt B.A., Kawaguchi Y., Johann D., Liotta L.A., Crawford H.C., Putt M.E., Jacks T., Wright C.V.E., Hruban R.H., Lowy A.M., Tuveson D.A., *Cancer Cell*, **4**, 437-450 (2003)
- 39) Poon T.C.W., Yip T-T., Chan A.T.C., Yip C., Yip V., Mok T.S.K., Lee C.C.Y., Leung T.W.T., Ho S.K.W., Johnson P.J., *Clin. Chem.*, **49**, 752-760 (2003)
- 40) Koopman J., Zhang Z., White N., Rosenzweig J., Fedarko N., Jagannath S., Canto M.I., Yeo C.J., Chan D.W., Goggins M., *Clin. Cancer Res.*, **10**, 860-868 (2004)

- 41) Lancashire, L, Ugurel, S, Creaser, C, Schadendorf, D, Rees, R, Ball, G. Utilizing Artificial Neural Networks to Elucidate Serum Biomarker Patterns Which Discriminate Between Clinical Stages in Melanoma. *IEEE 2005 Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, La Jolla, CA, USA. P:455-460 (2005)
- 42) Ragazzi E., Vogliardi S., Allegri G., Costa C.V.L., Lise M., Rossi C.R., Seraglia R., Traldi P., *Rapid Commun. Mass Spectrom.*, **17**, 1511-1515 (2003)
- 43) Adkins J.N., Varnum S.M., Auberry K.J., Moore R.J., Angell N.H., Smith R.D., Springer D.L., Pounds J.G., *Mol. Cell. Proteomics*, **1**, 947-955 (2002)
- 44) Tirumalai R.S., Chan K.C., Prieto D.A., Issaq H.J., Conrads T.P., Veenstra T.D. *Mol. Cell. Proteomics*, **2**, 1096-1103 (2003)
- 45) Koomen J.M., Zhao H., Li D., Baggerly K., Kobayashi R., *Rapid Commun. Mass Spectrom.*, **18**, 2537-2548 (2004)
- 46) Van Dirk W., Havenaar E.C., Brinkman-van der Linden E.C.M., *Glycoconjugate J.*, **12**, 227-233 (1995)
- 47) Hansen J.E., Larsen V.A., Bog-Hansen T.C., *Clin. Chim. Acta* **138**, 41-47 (1984)
- 48) Moore B.W., *Biochem. Biophys. Res. Commun.*, **19**, 739-744 (1965)
- 49) Gershenwald J.E., Thompson W., Mansfield P.F., Lee J.E., Colome M.I., Tseng C.H., Lee J.J., Balch C.M., Reintgen D.S., Ross M.I., *J. Clin. Oncol.* **17**, 976-983 (1999)
- 50) Fournier T., Medjoubi-N N., Porquet D., *Biochimica et Biophysica Acta* **1482**, 157-171 (2000)
- 51) Schmid K., Nimberg A., Kimura H., Yamaguchi J.P., Binette, *Biochim. Biophys. Acta*, **492**, 291-302 (1977)
- 52) Yoshima H., Matsumoto A., Mitsuochi T., Kawasaki T., Kobata A., *J. Biol. Chem.*, **256**, 8476-8484 (1981)

- 53) Kremmer T., Szöllösi E., Boldizsár M., Vincze B., Ludányi K., Imre T., Schlosser G., Vekey K., *Biomed. Chromatogr.* **18**, 323-329 (2004)
- 54) Duche J-C., Urien S., Simon N., Malaurie E., Monnet I., Barre J., *Clin. Biochem.*, **33**, 197-202 (2000)
- 55) Bleasby A.J., Knowles J.C., Cooke N.J., *Clinica Chimica Acta*, **150**, 231-235 (1985)
- 56) Alexander H., Stegner A.L., Wagner-Mann C., Du Bois G.C Alexander S., Sauter E.R., *Clin. Cancer Res.*, **10**, 7500-7510 (2004)
- 57) Mackiewicz A., Mackiewicz K., *Glycoconjugate J.*, **12**, 241-247 (1995)
- 58) Wigmore S.J., Fearon K.C., Maingay J.P., Lai P.B., Ross J.A., *Am. J. Physiol.*, **273**, E720-726 (1997)
- 59) Warner NB., The complement system, Mechanisms of activation and use as a diagnostic tool, Beckman Coulter, Inc. C-1, pg 10 (1998)
- 60) Tang S., Zhou W., Sheerin N.S., Vaughan R.W., Sacks S.H., *Journal Immunol.*, **162**, 4336-4341 (1999)

Chapter 6

Thesis Conclusions and Further Work

The work presented within this thesis described the modification of an early model of a commercially available AP-MALDI ion source (MassTech) and subsequent development of novel methodology for proteomics. The methodology developed in Chapter 2 showed that it was possible to obtain good quality spectra using the modified source controlled by the developed LabVIEW software. The combination of off-line capillary LC with AP-MALDI-QIT has been reported for the first time in Chapter 3. This novel technique was evaluated for protein identification by tryptic peptide analysis. The technique was compared with vacuum MALDI-TOF and ESI-qTOF analysis for the same BSA digest. Results showed that higher confidence of assignment for protein identification was achieved using tandem mass spectrometry for AP-MALDI MS. Interestingly, the analysis of peptides without LC separation prior to AP-MALDI-QIT MS resulted in higher sequence coverage, (comparable with MALDI-TOF analysis), than for peptides observed using LC-AP/MALDI-QIT. Although the exact reason for this is not clear it may be due to the higher acetonitrile content of the LC fractions effecting crystallisation. Further work to investigate this phenomenon would include the analysis of tryptic peptides deposited on the MALDI target plate with increasing concentration of acetonitrile to mimic LC conditions. Other future developments of the LC/AP-MALDI technique should include the analysis of tryptic peptides derived from complex protein mixtures and quantitative studies where the use of prior LC separation would be expected to lead to enhanced data quality compared to direct AP-MALDI.

These investigations showed the potential of AP-MALDI-QIT for proteomic analysis and the importance of combining data from instrumental platforms to maximise protein sequence coverage.

This thesis also reports the development of analytical protocols for the analysis of native and tryptic serum samples for peptide profiling of human serum using MALDI-TOF and AP-MALDI-MS. A number of techniques were evaluated including ultrafiltration and

protein precipitation for the removal of high abundant molecular weight proteins. A highly reproducible and robust method for the preparation of proteolytically digested serum samples was reported in Chapter 4. This is the first report of the removal of high mass abundant proteins using ZipTip clean-up prior to tryptic digestion with the resulting peptides de-salted and concentrated with further ZipTip clean-up.

Although many studies have used SELDI and MALDI for biomarker analysis^{1,2} very few groups have reported reproducibility data in terms of mass accuracy and ion intensities.³ Reproducibility studies were carried out using the optimised methods for both native and tryptic serum peptides. Results for native serum samples showed percentage relative standard deviation (% RSD) of < 0.03 % and 29 % for *m/z* and intensity variations respectively. However, tryptic peptides showed slightly poorer intensity reproducibility, with % RSD values of < 0.03 % for *m/z* and 36 % for variation in intensity.

Criticisms have been made that the differences observed in profiles between healthy and disease states may be due to pre-analytical factors, such as sample handling and storage conditions rather than biological differences.⁴ Changes in the proteome arising from room temperature and 4 °C incubation and freeze thaw cycles have been investigated and results showed that spectral differences could be observed for both native and tryptic serum peptides. Therefore future clinical samples can be assessed for degradation, ensuring that any changes observed are due to disease states rather than sample aging and storage. Also the results of this study will provide experimental data that will allow future sample collection and storage protocols for serum samples to be addressed.

The techniques developed in this thesis were applied to the analysis of serum samples taken from stage IV melanoma patients and a control population. Data were subjected to ANNs analysis to prioritise peptide/protein biomarker ions, which were differentially expressed in the melanoma samples in comparison with the healthy control samples. Identification of the peptide sequences and protein identities of the predictive biomarker

ions was carried out using MALDI-TOF PSD, AP-MALDI-QIT MS/MS and LC-ESI-MS/MS for the tryptic peptides. Results showed conclusively that two of the tryptic peptide ions which are highly significant in distinguishing between stage IV melanoma and control sera (>92.3 % accuracy) were derived from α -1-acid glycoprotein and a further discriminatory peptide was derived from complement C3 protein.

Work in this area of research is ongoing with the application of the novel high throughput methodology developed in Chapter 4 to the analysis of sera from patients diagnosed with other clinical stages of melanoma (stages I, II and III). Following the success of biomarker identification reported here for stage IV melanoma it is anticipated that it will be possible to ascertain peptide ion biomarkers and protein identifications for detection of early stages of melanoma. Further work to extend this methodology encompassing plasma samples and tissue sections (e.g. biopsies) should be conducted using the methodologies described in Chapters 4 and 5.

A number of ions identified by ANNs were significant in classifying between melanoma stage IV and control samples. For example the top discriminatory ion for native peptide was observed at m/z 2755. This ion alone was capable of predicting between the two states with 69.9 % accuracy. Therefore as previously mention in Chapter 5 section 5.3.1 identification of this ion and the parent protein should be an important future objective. Due to the limited mass range of the LCQ quadrupole ion trap (1-2000 Da) tandem mass spectrometry of this ion would need to be carried out on another mass analyser, such as a qTOF.

Parameters such as diet, age, gender and geographical location should be examined on clinical data to exclude possible variations observed for different sample sets.

The methodology developed is generic for serum analysis and could also be applied to other types of cancer and other diseases such as Alzheimer's. The methodology for tryptic serum samples can be automated using a liquid handling system and this would increase throughput by allowing 24 hour sample processing. This is of particular importance if the methodology is to be used in a clinical environment.

Additional external validation of the sample preparation methods and bioinformatic processing should be investigated. This could be achieved by inter-laboratory exchange of serum samples and protocols, giving confidence to the method in clinical settings and addressing the question of instrumental and operator variations.

Other areas of further work linking to this project could include the development of the control software developed in Chapter 1 to allow imaging of tissue sections such as cancer biopsies by AP-MALDI. This would allow a 3D map of the cancer tissue to be achieved. Further analysis of this data by bioinformatics may highlight particular regions of interest for more detailed study.

6.1 References

- 1) Petricoin E.F III., Ardekani A.M., Hitt B.A., *Lancet*, 261-266 (2002)
- 2) Wu B., Abbott T., Fishman D., McMurray W., Mor G., Stone K., Ward D., Williams K., Zhao H., *Bioinformatics*, **19**, 1636-1916 (2003)
- 3) Diamandis E.P., *J. Natl. Cancer I.*, **96**, 353-356 (2004)
- 4) De Noo, M.E.; Tollenaar R.A.E.M., Oezalp A., Kuppen P.J.K., Bladergroen M.R., Eilers P.H.C., Deelder A.M., *Anal. Chem.*, **77**, 7232-7241 (2005)

APPENDIX – Publications and Presentations

Publications

Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionisation Mass Spectrometry: A Review; Creaser CS. and Ratcliffe L., *Curr. Anal. Chem.*, **2**, 9-15 (2006)

Capillary LC/AP-MALDI-Ion Trap Mass Spectrometry: A Comparison with LC/MALDI-TOF and LC/ESI-qTOF For the Identification of Tryptic Peptides, Creaser CS., Green PS., Kilby PM., Ratcliffe L., *Rapid. Commun. Mass Spectrom.*, **20**, 829-836 (2006)

Diagnostic Biomarkers for Stage IV Melanoma Using a MALDI-TOF Mass Spectrometry-Bioinformatics Approach, Matharoo-Ball B., Ratcliffe L., Lancashire L., Ugurel S., Rees R., Sharendorf D., Ball G., Creaser CS., *Manuscript in Preparation*

Presentations

Serum protein and peptide mass fingerprinting using MALDI/TOF mass spectrometry: a question of reproducibility, Matharoo-Ball B. Ratcliffe L., Ball G, Ahmad M, Rees R and Creaser C, *Proceedings of the 28th annual meeting of the British Mass Spectrometry Society*, University of York, (4-7 September 2005) **Poster**

Integration of vacuum MALDI/TOF and AP-MALDI/QIT data for peptide and protein identification, Ratcliffe L., Creaser C and Rees R, *Analytical Research Forum*, University of Plymouth, (18-20 July 2005) **Poster**

Strategies for high-throughput proteomics using atmospheric pressure matrix-assisted laser desorption/ionisation combined with ion trap mass spectrometry, Ratcliffe LV., Nottingham Trent University (May 2005?) **Oral**

An integrated approach to the identification of serum peptides using MALDI-TOF in combination with AP/MALDI-ion trap mass spectrometry, Ratcliffe L, Creaser C and Rees R, *Proceeding of the 53rd ASMS Conference on Mass Spectrometry and Allied Topics*, San Antonio, Texas, USA (5-9 June 2005) **Poster**

Analysis of protein digests by capillary LC-AP/MALDI-ion trap, LC-MALDI-TOF and LC-ESI-qTOF mass spectrometry, Creaser C, Green P, Kilby P and Ratcliffe L, *Proceedings of the 27th annual meeting of the British Mass Spectrometry Society*, University of Derby, (5-8 September 2004) **Poster**

A comparison of capillary LC-AP-MALDI-ion trap mass spectrometry with LC-MALDI-TOF and LC-ESI-qTOF for protein digest analysis, Creaser C, Green P, Kilby P and Ratcliffe L, *Analytical Research Forum*, University of Central Lancashire, Preston, (19-21 July 2004) **Poster**

Peptide Characterisation using atmospheric pressure matrix-assisted laser desorption/ionisation-quadrupole ion trap mass spectrometry, Ratcliffe L, Mian S, Rees R and Creaser C, *Analytical Research Forum*, University of Sunderland, (21-23 July 2003) **Poster**

Development of control software for an atmospheric pressure matrix assisted laser desorption/ionisation source, Ratcliffe L and Creaser C., Nottingham Trent University, (September 2002) **Poster**