2001

Electrospray ionisation mass spectrometry of peptides and proteins: from proteomics to the characterisation of protein complexes

Jennifer A. Burgess

University of Wollongong

Recommended Citation
Electrospray Ionisation Mass Spectrometry of Peptides and Proteins: from Proteomics to the Characterisation of Protein Complexes

A thesis submitted in partial fulfilment of the requirements for the award of the degree

DOCTOR OF PHILOSOPHY

from

UNIVERSITY OF WOLLONGONG

by

JENNIFER A. BURGESS, B.Med.Chem. (Hons)

DEPARTMENT OF CHEMISTRY
2001
DECLARATION

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgement is made in the text of the thesis.

Jennifer A. Burgess
ABSTRACT

This thesis explores three different applications of electrospray ionisation mass spectrometry (ESI-MS) for protein analysis. An example of the identification of proteins following separation by 2-dimensional electrophoresis is presented with a study of proteins released from mitochondria following low-conductance permeability transition (LC-PT). A number of mitochondrial matrix proteins were identified from gels of the supernatant of mitochondria that had not undergone LC-PT. These findings raise questions regarding the release of proteins from normally functioning mitochondria. A second possibility is that the integrity of some of the mitochondrial membranes in the cell-free system may have been compromised. The identification of peroxisomal proteins in the mitochondrial preparations highlighted a major problem that is encountered with subcellular fractionation of organelles, namely contamination from other cellular components. This study also demonstrated the need for automated procedures to minimise human contamination and obtain reproducible results.

Functional proteomics is aimed at obtaining information on binding partners and protein function. An example of a functional proteomics approach developed here involved the use of affinity and ion exchange chromatography in conjunction with MS to identify proteins that associate with the T cell surface protein, CD4. The identification of a known CD4-binding protein, p56\textsuperscript{ck}, provided evidence that the method was suitable for the isolation of CD4-associating proteins. Identification of other interesting proteins (tubulin, myosin, actin, B-cell receptor-associated protein and two ribosomal proteins) has revealed potential CD4-binding partners for further immunological studies.

An interesting area in which ESI-MS has evoked significant interest has been in the analysis of noncovalent complexes. The gentle nature of ESI results in the ability to transfer complexes that are bound noncovalently into the gas phase under carefully controlled conditions. There is, however, still debate as to what extent the gas phase behaviour of complexes reflects their solution phase properties. Glutathione S-transferase (GST) is an ideal protein for investigating the applicability of ESI-MS to this type of analysis. A study of the complexes formed between GST and various ligands by ESI-MS is also presented. This study showed the formation of GST dimers within proteins of the same enzyme class. It was interesting that glutathione did not reveal the expected binding of two molecules per GST dimer. The preferential binding of two S-alkylglutathione molecules to the dimer, however, was clear from these data and the increased binding strength with increasing alkyl chain length was also apparent. Tandem mass spectrometry (MS/MS) was also found to be useful for investigating the specificity of ligand binding and for identifying the ligands bound to the GST dimer. MS/MS also showed that S-alkylglutathione ligands with longer alkyl chains required more energy to be removed from the complex that the S-alkylglutathiones of shorter alkyl chain lengths.
ACKNOWLEDGEMENTS

Throughout the course of this work many people have given me invaluable assistance. I would like to thank:

- My supervisor, Professor Margaret Sheil, for support, guidance, assistance and advice on more than just my PhD
- Professor Simon Gaskell, for the opportunity to spend time at UMIST and for guidance, assistance and supervision during my time in Manchester
- Mr Larry Hick for support, advice, technical assistance and teaching throughout my PhD
- Past and present members of the Biological Mass Spectrometry Laboratory at The University of Wollongong for their friendship and assistance
- Members of the Department of Chemistry at the University of Wollongong for creating an excellent working environment and help in surviving the refurbishments
- Mr Oliver Bernhard, Dr Andrew Sloane, Dr Tim Hochgrebe and Dr Garry Lynch for preparation of the anti-CD4 samples and advice concerning that work
- Associate Professor Mark Wilson and Ms. Alison Smail for preparation of the mitochondria and advice concerning that work
- Professor Lu-Yun Lian for the gift of the GST proteins and advice concerning that work
- Members of the Michael Barber Centre for Mass Spectrometry for advice and support during my stay in Manchester
- Micromass Australia and Micromass U.K. for advice and assistance with instrumentation and software. In particular I would like to thank Mr Steve Wilson (MMAus), Mr George Antonaides (MMAus), Dr James Langridge (MMUK), Dr Alan
Millar (MMUK), Mr Eugene Kapp (MMUK) and the late Dr Robert Bordoli (MMUK)

• The Menzies Centre for Australian Studies at the University of London for a scholarship to spend part of my PhD in the U.K.
• Micromass U.K. for further assistance to study in the U.K. and attend the International Mass Spectrometry Conference in Barcelona
• Dr Roland Annan for inspiration, encouragement and support throughout my PhD
• The late Dr Bob Bordoli, for technical assistance, particularly with the noncovalent complex work, and for encouragement and support
• Joanne Connolly, Michael Chalmers, Isabel Riba-Garcia and Joanna and Martin Lunt for friendship and support during my time in Manchester
• Amanda Cooper, Tammy Maples and Peter Edwell, for support, encouragement and down-to-earth advice
• Dad, for passing on your fighting spirit and determination to me. I can’t express how much I miss you
• Mum, for friendship, support, advice, love and so much more (and for dog-sitting for 8 months!)
• St. John, for love, patience and understanding, for proofreading this thesis, for drawing the diagrams in Chapter 4.1, for calming me, helping me and making me laugh when I most needed it
# TABLE OF CONTENTS

Declaration i)  
Abstract ii)  
Acknowledgements iii)  
List of Figures viii)  
List of Tables xi)  

## CHAPTER ONE: INTRODUCTION  

3.1 1.1 MASS SPECTROMETRY 1  
1.1.1 Electrospray ionisation 2  
1.1.2 Matrix assisted laser desorption/ionisation 6  
1.1.3 Mass analysers 7  
1.1.4 Tandem mass spectrometry 11  
1.1.5 Capillary liquid chromatography – mass spectrometry 16  

1.2 APPLICATION OF MASS SPECTROMETRY TO CONVENTIONAL PROTEOMICS 18  
1.2.1 Two-dimensional electrophoresis 18  
1.2.2 Database searching with MS information 20  

1.3 FUNCTIONAL PROTEOMICS 25  
1.3.1 Limitations of conventional proteomics 25  
1.3.2 Circumventing 2DE 26  
1.3.3 Functional proteomics 27  
1.3.4 Future directions 28  

1.4 OUTLINE OF THIS WORK 29  

## CHAPTER TWO: MATERIALS AND METHODS 32  

2.1 PREPARATION OF MITOCHONDRIAL PROTEINS FOR TWO-DIMENSIONAL ELECTROPHORESIS 32  
2.1.1 Mitochondria and permeability transition 32  
2.1.2 Chemicals and consumables for electrophoresis 33  
2.1.3 Extraction of proteins from mitochondria 34  

2.2 TWO-DIMENSIONAL ELECTROPHORESIS AND PREPARATION FOR MASS SPECTROMETRIC ANALYSIS 35  
2.2.1 Isoelectric focussing 35  
2.2.2 SDS-PAGE 36  
2.2.3 Silver staining 37  
2.2.4 In-gel tryptic digestion 38
2.3 PREPARATION OF ANTI-CD4 SAMPLES
2.3.1 Affinity capture of CD4 and co-associating proteins 38
2.3.2 Anion exchange chromatography 39

2.4 MASS SPECTROMETRY FOR PROTEOMICS APPLICATIONS
2.4.1 Desalting samples for mass spectrometric analysis 40
2.5.2 MALDI-MS analysis 40
2.5.3 NanoESI-MS and NanoESI-MS/MS 41
2.5.4 Capillary LC-MS and LC-MS/MS 42
2.5.5 Database searching 42

2.5 MASS SPECTROMETRY OF NONCOVALENT COMPLEXES
2.5.1 Preparation of samples for the analysis of complexes formed between glutathione S-transferase and ligands 43
2.5.2 Mass spectrometry for the analysis of complexes of glutathione S-transferase 45

CHAPTER THREE: AN EXAMPLE OF CONVENTIONAL PROTEOMICS
3.1 INTRODUCTION
3.1.1 The role of mitochondria in apoptosis 49
3.1.2 The Bcl-2 family 52
3.1.3 Cytochrome C 53
3.1.4 Caspases 55
3.1.5 Mitochondria and proteomics 57
3.1.6 Project aims 58

3.2 RESULTS
3.2.1 Overview of method 59
3.2.2 Extraction of proteins from mitochondria 61
3.2.3 Gels from proteins released from mitochondria during LC-PT 63
3.2.4 Examples of the identification of proteins from 2-DE gels 65
3.2.5 LC-MS of digests from gel-separated proteins 71
3.2.6 LC-MS/MS 76

3.3 DISCUSSION
3.3.1 Identification of mitochondrial proteins 79
3.3.2 Problems associated with 2-dimensional electrophoresis 80
3.3.3 Visualisation of proteins on gels 82
3.3.4 Identification of peroxisomal proteins in mitochondrial preparations 83
3.3.5 MALDI mass fingerprint and nanoES-MS/MS sequence tag searches 84
3.3.6 Capillary LC-MS 85
3.3.7 Automated procedures for conventional proteomics 86
3.3.8 Summary 87
CHAPTER FOUR: ISOLATION AND IDENTIFICATION OF CD4 AND CO-ASSOCIATING PROTEINS

4.1 INTRODUCTION
4.1.1 CD4 in T-cell activation
4.1.2 CD4 in thymocyte differentiation
4.1.3 The role of CD4 in HIV entry
4.1.4 Aims

4.2 RESULTS
4.2.1 Overview of method and method development
4.2.2 Summary of proteins identified
4.2.3 MS/MS and database searching
4.2.4 LC-MS
4.2.5 LC-MSMS and function switching
4.2.6 Preliminary results with a different anti-CD4 column
4.2.7 Importance of acidic residues

4.3 DISCUSSION
4.3.1 Method development
4.3.2 Significance of identified proteins
4.3.3 Database searching
4.3.4 Peptide fragmentation
4.3.5 Unidentified peptides
4.3.6 Capillary LC-MS
4.3.7 Data dependent MS to MS/MS switching
4.3.8 Potential for automated sample handling and data processing
4.3.9 Conclusions: Summary and future directions

CHAPTER FIVE: NONCOVALENT COMPLEXES OF GLUTATHIONE S-TRANSFERASE

5.1 INTRODUCTION
5.1.1 Noncovalent Interactions
5.1.2 Mass spectrometry of noncovalent complexes
5.1.3 Glutathione S-Transferases
5.1.4 Aims

5.2 RESULTS
5.2.1 Molecular mass determination, conventional electrospray conditions
5.2.2 Confirmation of modifications in glutathione S-transferases
5.2.3 Non-denaturing electrospray ionisation mass spectrometry of glutathione S-transferases
Chapter 5: Experimental Results

5.2.4 Addition of glutathione to GSTs
5.2.5 Addition of S-alkylglutathione inhibitors to GSTs
5.2.6 Competition experiments
5.2.7 MS/MS of glutathione S-transferase noncovalent complexes
5.2.8 In-source collision activated dissociation of noncovalent complexes
5.2.9 Effect of increased collision cell voltage
5.2.10 Titrations of GST A1-1(C112S) with various ligands
5.2.11 Formation of GST dimers within a class but not between different classes

Chapter 5: Discussion

5.3.1 ESI-MS spectra of proteins acquired under non-denaturing conditions
5.3.2 Effect of gas and voltage on ESI-MS spectra acquired under non-denaturing conditions
5.3.3 Complexes formed from the addition of GSH to GST
5.3.4 Effect of the tyrosine to phenylalanine modification
5.3.5 Effect of increasing chain length of S-alkylglutathiones
5.3.6 Non-specific binding
5.3.7 Dimer formation revealed by mass spectrometry
5.3.8 Conclusion: The applicability of ESI-MS to study non-covalent interactions of GST

Chapter 6: Conclusion

Referencens
3.8 MS/MS spectrum of doubly charged precursor ion at m/z 705.9 from an in-gel digest of spot A shown in Figure 3.5

3.9 MS/MS spectrum of triply charged precursor ion at m/z 854.1 from an in-gel digest of spot B shown in Figure 3.5

3.10 MALDI mass spectrum of an in-gel digest of spot A from Figure 3.6

3.11 Peak list of ions with m/z > 1000 Da in Figure 3.10 entered into MS-Fit for peptide mass fingerprint search of Swiss Prot database

3.12 Results from peptide fingerprint search from MS-Fit (Protein Prospector) using the m/z ratios of ions in Figure 3.10

3.13 Chromatograms from LCMS of an in-gel digest of spot C shown in Figure 3.5

3.14 Summed spectra from mass chromatograms shown in Figure 3.13

3.15 ESI mass spectra from the in-gel digest of spot C in Figure 3.5

3.16 MS/MS spectra from an LC-MS separation of an in-gel digest of spot C in Figure 3.5 (precursors selected by data-dependent MS to MS/MS switching)

CHAPTER 4

4.1 T-cell receptor recognition of antigen bound to an MHC Class II molecule

4.2 The lifecycle of HIV

4.3 Overview of method

4.4 Ion exchange chromatography profile from 60-70 minutes of the tryptic digests from the anti-CD4 (Q425) affinity column and IgG1 control column

4.5 MALDI-MS spectra of fraction 5 from an early ion-exchange chromatography separation of the column eluates after cleavage with trypsin

4.6 MALDI-MS of fractions from ion-exchange chromatography of the eluate from the anti-CD4 affinity column after cleavage with trypsin

4.7 MALDI-MS of fractions from ion-exchange chromatography of the eluate from the IgG control column after cleavage with trypsin

4.8 ESI-MS of fractions from ion-exchange chromatography of the eluate from the anti-CD4 affinity column after cleavage with trypsin

4.9 ESI-MS of fractions from ion-exchange chromatography of the eluate from the IgG control column after cleavage with trypsin

4.10 ESI-MS/MS spectrum of the precursor ion at m/z 1004.5 from preparation from the anti-CD4 column

4.11 Search parameters obtained from MS/MS of the precursor at m/z 1004.5 (Figure 4.10) entered into Peptide Search

4.12 Search results from input of the sequence tag (1044.6)NFS(1393.8) into Peptide Search

4.13 MS/MS spectrum of the precursor ion at m/z 1004.5 from preparation from the anti-CD4 column (as in Figure 4.10)

4.14 MS/MS spectrum of ion at m/z 979.0 from anti-CD4 preparation

4.15 MS/MS spectrum of the precursor ion at m/z 612.4 from the anti-CD4 preparation

4.16 MS/MS spectrum of ion at m/z 766.85 from fraction 5 of anti-CD4 preparation
4.17 MS/MS spectrum of the precursor ion at m/z 862.5 from fraction 6 of an anti-CD4 preparation

4.18 Chromatograms from LCMS of fraction 5 from an anti-CD4 preparation

4.19 Summed spectra from mass chromatograms shown in Figure 4.18

4.20 Mass chromatograms of A) m/z 615.3 and B) m/z 766.9 from LC-MS of fraction 4 from an anti-CD4 preparation

4.21 Mass chromatograms of A) m/z 615.3 and B) m/z 766.9 from LC-MS of fraction 5 from an anti-CD4 preparation

4.22 Mass chromatograms of A) m/z 615.3 and B) m/z 766.9 from LC-MS of fraction 6 from an anti-CD4 preparation

4.23 MS/MS spectrum of a precursor ion at m/z 767 acquired during LC-MS experiment of fraction 5 from anti-CD4 preparation

4.24 MS/MS spectrum of a precursor ion at m/z 851.5 acquired during LC-MS experiment of fraction 4 from anti-CD4 preparation

4.25 MS/MS spectrum of a precursor ion at m/z 719.9 from an LC-MS separation of fraction 5 from an anti-CD4 preparation

4.26 MS/MS spectrum of a precursor ion at m/z 895.9 from an LC-MS separation of fraction 5 from an anti-CD4 preparation

4.27 ESI-MS spectra of fraction 4 from ion exchange chromatography of eluate from columns after tryptic cleavage

4.28 MS/MS spectra of ion-exchange column fraction 4 from digests of elutions from the anti-CD4 column (new antibody, OKT4) and the IgG control column

4.29 The sequence of the T cell surface glycoprotein CD4 with peptides generated from a theoretical digest with trypsin labeled (T1 to T63)

4.30 The sequence of the protein tyrosine kinase p56\textsuperscript{ck} with peptides generated from a theoretical digest with trypsin labeled (T1 to T63)

CHAPTER 5

5.1 ESI-MS spectra of GST under denaturing conditions

5.2 The sequence of GST A1 with peptides generated from a theoretical digest with trypsin labelled (T1 to T35)

5.3 The sequence of GST P1 with peptides generated from a theoretical digest with trypsin labelled (T1 to T19)

5.4 MS/MS spectrum of m/z 1023.86 (3+) from a tryptic digest of A1\textsuperscript{CS}

5.5 MS/MS spectrum of m/z 453.2 (2+) from a tryptic digest of A1\textsuperscript{YF}

5.6 MS/MS spectra for confirmation of the P1 N-terminal initiator methionine

5.7 Non-denaturing ESI-MS spectra of wild-type A1, A1\textsuperscript{CS}, A1\textsuperscript{YF} and P1

5.8 ESI-MS spectra of A1 (wild-type) + GSH under non-denaturing conditions

5.9 ESI-MS spectra of GST + GSH (molar ratio 1:5) under non-denaturing conditions

5.10 ESI-MS spectra of GST A1\textsuperscript{CS} + GS-prop under non-denaturing conditions

5.11 ESI-MS spectra from the addition of GS-hex to each of the alpha class enzymes under non-denaturing conditions
5.12 Non-denaturing ESI-MS from the addition of GS-dec to the two modified A class enzymes 216
5.13 ESI-MS spectra from mixtures of A1CS, GSH and GS-hex acquired under non-denaturing conditions 218
5.14 ESI-MS spectra from mixtures of A1CS, GS-prop and GS-hex acquired under non-denaturing conditions 220
5.15 ESI-MS spectra of the mixture resulting from the addition of GS-hex and GS-dec to the modified A-class enzymes (A1CS and A1YF) under non-denaturing conditions 221
5.16 Spectra from MS/MS of [A1CSdimer + 16H+]16+ (m/z 3187) 223
5.17 Spectra from MS/MS of [A1YFdimer + 2GS-hex + 16H+]16+ (m/z 3235.5) 225
5.18 Spectra from MS/MS of [A1CSdimer + GSH + 2GS-hex + 16H+]16+ (m/z 3255) 227
5.19 Spectra from MS/MS of [A1CSdimer + GS-hex + GS-dec + 16H+]16+ (m/z 3239) 228
5.20 Spectra from MS/MS of [A1CSdimer + 3GS-hex + 15H+]15+ (m/z 3477) 230
5.21 Spectra from MS/MS of [A1CSdimer + 2GS-hex + 15H+]15+ (m/z 3451) and [A1CSdimer + 3GS-hex + 15H+]15+ (m/z 3477) 231
5.22 Spectra from in-source CAD of a mixture of A1CS + GSH (molar ratio 1:10) 234
5.23 Spectra from in-source CAD of a mixture of A1CS + GS-dec (molar ratio 1:5) 235
5.24 ESI-MS spectra from a mixture of A1CS, GSH and GS-hex (molar ratio 2:20:1) under non-denaturing conditions with increasing collision offset 236
5.25 Graph of relative intensity of complexes of A1CS dimer and GSH versus the molar excess of GSH 240
5.26 Titrations of A1CS with A) GSH, B) GS-prop and C) GS-hex 241
5.27 An enlarged portion of the titration of A1CS with GS-hex 242
5.28 MaxEnt™ spectrum of a mixture of A1CS and P1 244

LIST OF TABLES

1.1 Characteristics of miniaturised versions of HPLC 17
1.2 Protein and translated oligonucleotide sequences available via the Internet 21
1.3 Search programs that are available via the Internet for interrogation of databases with MS data 24
2.1 Voltage supply program for IEF of 13cm IPG gels with daytime focussing 36
4.1 Molecules involved in T cell activation, their location, binding partners and functions 93
4.2 Summary of identified peptides and the method used to obtain their identification 117
4.3 Summary of identified proteins with the total number of peptides sequenced from each of the proteins 119
4.4 Summary of unidentified precursors including possible sequence tags or sequence only search options and possible reasons as to why the peptides were not identified 121
4.5 The sequences of CD4 tryptic peptides of mass >1 kDa with their monoisotopic mass and the number of aspartic acid (D) and glutamic acid (E) residues in the sequence 152
4.6 Summary of the potential sequence coverage of the tryptic peptides from the identified proteins 154
5.1 Theoretical and experimental molecular masses of glutathione S-transferases 198
CHAPTER ONE: INTRODUCTION

1.1 MASS SPECTROMETRY

Mass spectrometry (MS) developed as a tool in atomic physics but has undergone dramatic developments to make it a viable tool for the study of biomolecules. The introduction of electron ionisation (EI), chemical ionisation (CI) and the coupling of gas chromatography with mass spectrometry (GC-MS) were important developments in the application of MS in chemistry. Further early developments in ionisation techniques included field desorption, liquid desorption, plasma desorption and thermospray, which had varying degrees of success for the analysis of involatile, thermally labile molecules. The introduction of the ionisation technique fast atom bombardment (FAB) by Barber and colleagues in the early 1980s (Barber et al., 1981) marked the start of an expanding field of biological mass spectrometry since it provided the first routine method for ionising peptides and carbohydrates. The development of matrix assisted laser desorption/ ionisation (MALDI) (Karas and Hillenkamp, 1988) and electrospray ionisation (ESI) (Fenn et al., 1989) dramatically increased the mass range of biomolecules that could be ionised (to approximately 150 kDa) and resulted in a rapid expansion of the application of mass spectrometry to fields such as biochemistry, immunology, molecular biology and biotechnology. Advances in instrumentation have also had a large impact on the growth of biological mass spectrometry with the re-emergence of time-of-flight (ToF) analysers, the introduction of new hybrid instruments and improvements in electronics and computational methods.
1.1.1 Electrospray Ionisation

The first reports on the use of electrospray for the production of gas-phase ions came from Dole and co-workers in the late 1960s (Dole et al., 1968). Much later, two separate groups reported the coupling of ESI to a mass spectrometer (Aleksandrov et al., 1985; Yamashita and Fenn, 1984). The transfer of protein ions to the gas phase was published several years later by Fenn and colleagues (Fenn et al., 1989). The introduction of this ionisation technique has had an enormous impact on the use of mass spectrometry for analysis of biomolecules such that it has now become an essential tool in protein chemistry.

In ESI, a high voltage is applied to a capillary through which a solution of the analyte is pumped. A mist of highly charged droplets is sprayed from the tip of the capillary and the small droplets evaporate. The evaporation process is often assisted by heating and the application of a counter-current flow of nitrogen gas. The mechanism of the formation of gaseous ions from the small droplets is thought to occur by either of two processes (discussed below). The gas phase ions then enter the mass spectrometer for separation and detection according to their mass-to-charge ratios (m/z). (Mann and Wilm, 1995)

A relatively small amount of internal energy is imparted to the analyte, so there is very little (if any) fragmentation (Mann and Wilm, 1995; Cole, 2000). ESI is therefore known as a “soft ionisation” technique and is suitable for the analysis of fragile, thermally labile biomolecules. The multiple charging that is observed in ESI mass spectra allows for the detection of high molecular mass molecules and complexes. As the molecular mass increases so does the average number of charges (Smith et al., 1990b) and, since separation and detection is based on m/z, the ions are brought into the range of detection of most mass spectrometers. Multiple charging also improves the
accuracy of the molecular mass measurement because the mass is sampled in a number of different measurements (Smith et al., 1990b). Another advantage of ESI is the ability to couple on-line solution separations with MS (see section 1.1.6).

Ion formation in electrospray ionisation

The electrospray process can be considered in a number of steps (Smith et al., 1990b; Gaskell, 1997; Kebarle and Tang, 1993). These include the formation of highly charged droplets from the spraying orifice, the decrease in size of these droplets by evaporation and Coulombic fission and the formation of desolvated gas-phase ions from the droplets. These processes are shown schematically in Figure 1.1.

The electrical field at the capillary tip causes positive or negative ions (depending on the polarity of the potential on the capillary) to accumulate at the surface of the liquid (Smith et al., 1990b; Gaskell, 1997; Kebarle and Tang, 1993; Kebarle, 2000). The repulsive forces experienced by the ions and the force of the electric field result in the formation of a cone shape (the Taylor cone) that expands out to become a thin liquid
filament (Wilm and Mann, 1994; Gaskell, 1997; Kebarle and Tang, 1993; Kebarle, 2000). When the electrostatic force becomes sufficiently strong, the surface tension is broken and droplets form. Owing to the surface tension dependence of droplet formation, it is more difficult to achieve a stable spray with water than with an aqueous organic (usually methanol or acetonitrile) solution.

Evaporation of the solvent occurs as the droplets move through the source. Droplets may also undergo fission upon reaching or exceeding the ‘Rayleigh’ limit of the droplets, i.e. the radius at which the Coulombic repulsion of the charges equals the surface tension forces holding the droplet together (Vestal, 2001). Repeated fissions and solvent evaporation continue droplet shrinkage. Two models exist for the formation of ions from these small droplets. The charge residue model (CRM) proposes that fission continues until there is only a single analyte ion in the droplet and the remaining solvent is removed by evaporation (Smith et al., 1990b; Kebarle and Tang, 1993; Gaskell, 1997). Alternatively, the ion evaporation model (IEM) proposes that an ion is ejected from the droplet when the Coloumbic repulsive forces between the charges within the droplet become too strong (Iribarne and Thomson, 1976; Thomson and Iribarne, 1979; Smith et al., 1990b; Kebarle and Tang, 1993). There is data to support both these mechanisms of gas-phase ion formation and debate remains over which model best represents the process that is occurring. However, there is mounting evidence indicating that globular proteins and noncovalent biomolecular complexes form gaseous ions through the CRM (Kebarle, 2000). Furthermore, the observation of non-specific complexes can best be accounted for by the possibility of fission resulting in two or more ions in a single droplet, which form a complex following evaporation of the last solvent molecules, providing support for the CRM (Kebarle, 2000).
Nanoelectrospray Ionisation

The ion current obtained in ESI-MS is dependent upon the concentration of the analyte but not on the flow rate (Mann and Wilm, 1995; Cole, 2000). A reduction in the flow rate therefore allows increased information to be obtained from a limited amount of sample without loss of sensitivity. In nanoelectrospray ionisation (nanoESI), the solution of analyte is introduced in a metal-coated capillary that is drawn out to a fine tip (Wilm and Mann, 1994; Valaskovic et al., 1995). Voltage applied to the capillary results in an electrostatic flow of a fine spray of solution from the capillary at very low flow rates (20-40 nL/min) (Wilm and Mann, 1994; Wilm et al., 1996b). The droplets produced by nanoESI are approximately 100 times smaller than those produced by conventional electrospray (Wilm et al., 1996b). Therefore high source temperatures and the counter-current drying gas flow rates are not required. The smaller droplets have an enhanced surface-to-volume ratio so a large proportion of analyte molecules are accessible for ionisation (Fligge et al., 1999). The capillary is placed close to the sampling cone and the percentage of analyte molecules entering the next stage is much higher than in conventional electrospray (Wilm and Mann, 1996). Sample consumption is significantly lower and there is an improvement in sensitivity (Fligge et al., 1998). The low flow rates also yield long analysis times, which improve the limit of detection and allows for successive analyses of a large number of components in a given mixture by tandem mass spectrometry (more below).

ESI-MS is intolerant to salts and detergents owing to instability of the spraying process and competition for charges that are introduced by ionic species (Mann and Wilm, 1995; Cole, 2000). An additional advantage of the nanoESI source is the increased tolerance for salts compared to conventional ESI (Wilm and Mann, 1996;
Fligge et al., 1998). NanoESI is, therefore, more suitable for the introduction of analytes to the mass spectrometer from solutions that are close to physiological conditions.

1.1.2 Matrix Assisted Laser Desorption/Ionisation

Matrix Assisted Laser Desorption/Ionisation (MALDI), unlike ESI, is a pulsed ionisation technique. A solid matrix is mixed with the analyte to absorb energy from a laser pulse, which leads to excitation of the solid and desorption of the surface layers of matrix and analyte (Roepstorff, 2000; Spengler, 2001; Karas et al., 2000). Both matrix and analyte molecules are carried into the gas phase. MALDI mass spectra are characterised by little fragmentation and the majority of ions observed are singly charged (Spengler, 2001).

The mechanism by which ions are produced in MALDI is not well understood, although there are have been many proposals since the introduction of MALDI. Ionisation most probably involves a number of different processes, which may be dependent upon the analyte, matrix, sample preparation and laser wavelength and energy (Zenobi and Knochenmuss, 1998; Karas et al., 2000). The processes involved have recently been reviewed and may include multiphoton ionisation, proton transfer from matrix to analyte, cationisation, pooling of energy from excited matrix species, excited-state proton transfer, electron transfer, desorption of pre-formed ions and thermal ionisation (Zenobi and Knochenmuss, 1998). A recent model encompasses a number of suggested mechanisms to explain the formation of ions in MALDI and explains the survival of singly charged molecular ions (Karas et al., 2000).

The ions formed by MALDI have a significant spread of velocities, which results in a decrease of mass resolution. The use of an electrostatic mirror (reflector) in the early MALDI instruments partly compensated for the velocity spread but prior to 1995
MALDI instruments had relatively poor resolution. The introduction of delayed extraction (DE) has improved the resolution of MALDI analyses (Brown and Lennon, 1995; Spengler, 2001; Vestal, 2001; Roepstorff, 2000). DE developed from a technique known as ‘time-lag focussing’ described by Wiley and McClaren in 1955 (McLuckey and Wells, 2001). After ionisation, a short delay is allowed before the ions are accelerated (Coles and Guilhaus, 1993). During the delay, ions with higher energy from the ionisation process move further than those with less energy. A fixed accelerating voltage is then applied, which results in ions of lower energy experiencing a higher voltage than the ions of higher energy that have moved further down the gradient. The small difference in accelerating potential that is applied to the ions compensates for the initial spread of velocities.

For protein and peptide analyses by MALDI-MS, sample preparation typically involves co-crystallisation of analyte and excess matrix (Zenobi and Knochenmuss, 1998). The matrix absorbs the energy of the laser pulse and may play a direct role in the ionisation of analyte molecules (Roepstorff, 2000; Karas et al., 2000). The matrix is also important for dilution of the analyte and preventing aggregation of the analyte (Karas et al., 2000).

1.1.3 Mass Analysers

Mass analysers use electric and/or magnetic fields to separate ions according to their mass-to-charge ratio (m/z). Important aspects of the mass analyser include: i) mass range, ii) mass accuracy, iii) sensitivity, iv) resolving power, v) scanning speed and vi) compatibility with the ionisation source (Jennings and Dolnikowski, 1990; McLuckey and Wells, 2001). Electric sector, magnetic sector, ToF, quadrupole, ion trap and ion cyclotron are the most common analysers. Many commercially available mass
spectrometers combine more than one analyser, resulting in improved performance and the ability to perform tandem mass spectrometry (MS/MS). In this research, two different types of mass spectrometers were used, a MALDI-ToF instrument and a hybrid instrument, the quadrupole-ToF.

Quadrupole analysers

A quadrupole analyser consists of four rods arranged in parallel where opposite rods are electrically connected (Jennings and Dolnikowski, 1990). Application of a direct current (DC) voltage and an alternating current (AC) voltage at radio frequencies (RF) to the rods (opposite polarities on adjacent rods) creates an oscillating field (Siuzdak, 1996; Yates, 1998). At a given DC and RF potential, only ions of a specific m/z are transmitted through the quadrupole. The potentials can be varied, keeping their ratio constant, to selectively transmit the full m/z range. Standard quadrupoles have an upper m/z limit of approximately 2000-4000 but modified quadrupoles have been reported that are able to scan higher m/z ranges (McLuckey and Wells, 2001).

Quadrupoles are ideal for coupling with ESI sources owing to their tolerance of relatively high pressures (Siuzdak, 1996). Triple quadrupole instruments were the first widely employed tandem mass analysers. In these instruments the first quadrupole is used for precursor ion selection, the second is used as a cell for collision-activated dissociation (with only an RF voltage) and the third quadrupole is used as a second mass analyser. In practice, the use of a quadrupole for the central collision cell is not essential and some commercial instruments employ either a hexapole or octapole, which offer superior transmission.

Time-of-flight analysers

Time-of-flight (ToF) mass spectrometry relies on the principle that ions with the same energy and different masses will have different velocities and hence different flight
times in a field-free region. The velocity of an ion is inversely proportional to the square of its m/z so ions of higher m/z will have longer flight times (Kinter and Sherman, 2000). With advances in electronics, ToF analysers now have high mass accuracy and sensitivity. Unlike scanning instruments, ToF analysers theoretically do not have an upper mass limit since larger ions merely take longer to traverse the flight tube. In practice, it is therefore the detector rather than the analyser that limits the mass range, owing to limitations with the detection of high mass ions with very low velocities (Mirgorodskaya et al., 1994). ToF analysers detect the entire mass range simultaneously, which results in fast acquisition times. Scanning analysers, such as the quadrupole, scan small regions at a time and therefore the percentage of ions that are actually detected can be quite low (poor duty cycle).

The measurement of the flight time of an ion requires a starting point (zero time). With pulsed ion sources, such MALDI, the ions are accelerated through the flight tube in packets and the pulse acts as the starting point. Sufficient time must be allowed for the largest ion to reach the detector before the next ion packet is accelerated. Continuous ion sources, such as ESI, require a storage device to provide a starting point for the ToF. Success in coupling continuous ions sources to ToF analysers has been achieved by accelerating the ions into a flight tube that is perpendicular to the ion beam (Guilhaus et al., 2000). This technique is known as orthogonal acceleration ToF (oaToF). Duty cycles of 5-50% are typical for oaToF mass spectrometers (McLuckey and Wells, 2001).

The initial range of velocities of the ions limits the resolution of a ToF analyser, which is particularly significant with ions generated by MALDI. As mentioned in section 1.1.2, there are a number of ways to correct for the initial energy spread of the ions. With the use of a reflector, ions of the same mass but with higher energy penetrate
the reflector more deeply and spend more time in the reflector, thereby compensating for the velocity spread (Coles and Guilhaus, 1993; Yates, 1998; McLuckey and Wells, 2001; Kinter and Sherman, 2000). Ions of the same mass are focussed to one point in time. The increase in path length that is introduced by the reflector is an added advantage. Increasing the flight time causes a proportionate decrease in the difference between two ions of the same mass but different energy. Similarly, higher accelerating voltages increase the energy imparted to the ions and decrease the relative contribution from the ionisation process (Coles and Guilhaus, 1993; Kinter and Sherman, 2000). The use of an oaToF also minimises the spread of energies and results in improved resolution (Guilhaus et al., 2000; Shukla and Futrell, 2000). Another technique employed to reduce the velocity dispersion of the ions generated from MALDI is delayed extraction (DE) (described in section 1.1.2).

The quadrupole-ToF

Glish and Goeringer published the first report of a hybrid quadrupole-ToF (Glish and Goeringer, 1984). The instrument consisted of a quadrupole, collision cell and a linear ToF. A lens stack prior to the ToF accelerated the ions from the collision cell into the ToF. This instrument was limited to the analysis of small, volatile molecules owing to the ionisation process and the linear arrangement of the ToF limited the attainable resolution of the instrument. It was not until the introduction the oaToF, and the development of ESI and MALDI, that interest in quadrupole-ToF instruments was again evoked.

Current commercial quadrupole-ToF instruments combine the quadrupole mass filter and hexapole or quadrupole collision cell of a standard triple quadrupole with an oaToF incorporating a reflector (Figure 1.2) (Morris et al., 1996; Shevchenko et al., 1997). The first quadrupole can act either for mass selection in MS/MS (see below) or
in RF-only mode to allow the transfer of all ions for MS analysis in the ToF analyser. The ToF analyser is therefore used to acquire both MS and MS/MS data. The collision cell is able to re-focus ions that are scattered through collisions with the collision gas (Yates, 1998). The quadrupole-ToF instruments were first coupled to ESI sources but MALDI quadrupole-ToF instruments have more recently been developed (Loboda et al., 2000; Shevchenko et al., 2000).

Figure 1.2 The Q-Tof™ (Reprinted with permission from Micromass U.K)

1.1.5 Tandem Mass Spectrometry (MS/MS)

Tandem mass spectrometry (MS/MS) has become one of the most widely used techniques for the characterisation of biologically important molecules. MS/MS uses two stages of mass analysis. In the first stage an ion of interest (i.e. the precursor ion) is
mass selected. This precursor ion is then activated via collisions with an inert gas in a reaction region or "collision cell". Ions that become sufficiently activated through multiple low-energy collisions undergo fragmentation. This process is termed collision-induced dissociation (CID) or collision-activated dissociation (CAD). The resulting fragment (or product) ions are measured in a second mass analysis stage.

Tandem mass spectrometers can be divided into two types: tandem-in-space mass spectrometers and tandem-in-time mass spectrometers (McLuckey and Wells, 2001). Tandem-in-space instruments rely on a sequential path through two or more analysers coupled together (Shukla and Futrell, 2000). In a quadrupole-ToF instrument, the precursor is selected by the quadrupole, fragmented in an RF-only hexapole collision cell and the ToF analyser separates the product ions for detection. Tandem-in-time instruments (such as ion traps) are able to perform MS/MS within the same region and are therefore separated by time rather than space.

The amount of energy imparted on an ion affects the fragmentation. In tandem-in-space instruments, the centre of mass collision energy (E_{com}) is dependent on the collision energy applied to the gas cell (E_{lab}), the mass of the collision gas (M_g) and the mass of the selected precursor ion (M_i) in the relationship

\[ E_{com} = E_{lab} \times M_g / (M_g + M_i) \]  

Low-energy collisions are generally defined as those with an E_{lab} of about 10-30 eV (Hunt et al., 1986). Low energy CAD of peptides gives rise to fragmentation along the

---

1 It is also worth noting that CAD can be performed in the atmosphere-to-vacuum interface of an ESI instrument by increasing the cone voltage albeit without prior mass selection Smith, R. D., Loo, J. A., Barinaga, C. J., Edmonds, C. G. and Udseth, H. R. (1990a) The Journal of the American Society for Mass Spectrometry, 1, 53-65; Kilby, G. W. and Sheil, M. M. (1993) Organic Mass Spectrometry, 28, 1417-1423.. This is termed "in-source CAD" and is addressed in Chapter 5.
backbone of the peptide chain. High energy CAD of peptides results in backbone plus side-chain fragmentation.

For MS/MS experiments with quadrupole-ToF instruments, the quadrupole is typically set to allow the transfer of all isotopes of the precursor into the collision cell. The high resolution afforded by the ToF analyser allows the charge on the product ions to be determined by the isotope spacing.

Another scan mode that has been found to be particularly useful for the identification of modified (e.g. phosphorylated) peptides is the precursor ion scan. With a triple quadrupole tandem mass spectrometer the first mass analyser scanned while the second is set to transmit only a single m/z value (Wilm et al., 1996a; Annan and Carr, 1997; Yates, 1998). This way, precursor ions that result in specific fragments can be identified. Quadrupole-ToF instruments are not able to perform precursor ions scans in this way but similar information can be generated by integrating the data from a large number of acquisitions, albeit with lower sensitivity than that available from the triple quadrupole precursor ion experiment.

In MALDI-ToF MS, fragmentation information may be obtained from post source decay (PSD) (Kaufmann et al., 1994). In this type of experiment, the unimolecular (metastable) decompositions and collisions with background gas as the ions pass through the drift tube produce fragments that are observed by stepwise reduction of the potential on the reflector (Kaufmann et al., 1994; Devreese and Van Beeumen, 1998; Zenobi and Knochenmuss, 1998; Chaurand et al., 1999). PSD experiments produce high-energy CAD spectra that show greater diversity of product ions and are generally more complex, than spectra from MS/MS experiments on an ESI instrument.
MS/MS of peptides

The fragmentation of protonated peptides by MS/MS has been extensively documented and low energy collisions resulting in fragmentation of the peptide backbone are typically observed with triple quadrupole and quadrupole-ToF MS. The most commonly observed ions from backbone fragmentation result from cleavage of the amide bond with charge retention on the N-terminus (B-type ions) or the C-terminus (Y-type ions) as shown in scheme 1.1 (Papayannopoulos, 1995; Roepstorff and Fohlman, 1984; Hunt et al., 1986). The formation of sequential B and/or Y ions from a peptide enables the sequence to be established from the mass difference between sequential ions in the series.

Scheme 1.1 Ions generated from fragmentation of the peptide backbone. A, B and C ions are the result of charge retention on the N-terminus. X, Y and Z ions are the result of charge retention on the C-terminus. The primes (\(^\prime\)) indicate addition of two H atoms. (Roepstorff and Fohlman, 1984)
The distribution of charges plays a critical role in the fragmentation pathway and affects the proportion of B and Y ions that are produced (Staudenmann and James, 2001; Tang et al., 1993; Downard and Biemann, 1994; Burlet et al., 1992). The localisation of charge is dependent on the presence of amino acids that have high gas phase basicity such as arginine, lysine, histidine, proline and tryptophan (Hunt et al., 1986). Tryptic peptides (except for the C-terminal peptide) possess a basic C-terminus owing to the cleavage specificity of trypsin, which cleaves on the C-terminal side of lysine and arginine residues. The presence of the C-terminal arginine or lysine typically results in charge retention on the C-terminal fragment ion, thereby favouring the formation of Y-type sequence ions (Hunt et al., 1986). For multiply charged ions, the charges may be distributed between the two types of ions (charge retention on the C-terminal arginine or lysine and the basic N-terminus) resulting in detection of both the Y ion and the complementary B ion (Smith et al., 1990b). Alternatively, the charges may be retained by one fragment, resulting in multiply charged product ions.

Other ions that are commonly observed in MS/MS spectra include the loss of neutral molecules such as water or ammonia from B or Y ions and the loss of CO from B ions, which produces the A-type ions (scheme 1.1) (Hunt et al., 1986; Tang et al., 1993). Internal fragments are also observed, particularly when a proline residue is present in the peptide (Hunt et al., 1986). These fragments result from the formation of a Y ion, which then undergoes a second fragmentation to produce a B-type ion. Immonium ions, which are of the form $\text{H}_2\text{N}^=\text{CHR}^+$, are often observed in the low m/z region and provide an indication of the amino acids present in the peptide (Hunt et al., 1986).
1.1.6 Capillary Liquid Chromatography – Mass Spectrometry

Analytical liquid chromatography (LC) was coupled to ESI-MS soon after the introduction of ESI (Bruins et al., 1987; Stults et al., 1996). The application of capillary LC has been slower, owing to problems with the construction and packing of such small columns (< 0.5 mm internal diameter) and in the generation of the low flow rates required (Moritz et al., 1994b). Advances in column design, column packing techniques, flow-splitting and pumping techniques have made capillary LC a much more accessible technique (Moritz and Simpson, 1992; Tong et al., 1997; Chervet et al., 1996). The introduction of a commercial pumping system that is capable of delivering gradient elution with homogeneous mixing and reproducible delivery has removed the need for flow-splitting (Holyoke et al., 1999). The use of standard flow-splitting techniques with conventional pumping systems, however, continues to provide comparable results in many laboratories.

A summary of the HPLC miniaturisation techniques and their analytical properties is shown in Table 1.1. As the internal diameter (I.D.) of the column decreases, the concentration of the analyte peak increases in proportion to the inverse square of the diameter (Stults et al., 1996). The optimum flow rate (and solvent consumption) decreases with the decreasing column I.D. and the resolution of components increases. Smaller columns have therefore resulted in improved sensitivity when coupled with MS (Moritz and Simpson, 1992; Chervet et al., 1996; Haynes et al., 1998). There are, however, some practical limitations with smaller columns that must also be considered. Leaks are more difficult to observe with lower flow rates and the peak broadening caused by unions is more pronounced. Sample loading is also limited and increased loads may result in peak broadening and detector saturation (Moritz and
Simpson, 1992). Smaller columns are also more prone to clogging from particulates or salt deposits.

<table>
<thead>
<tr>
<th>Name</th>
<th>Column I.D.</th>
<th>Flow Rate</th>
<th>Sample Range (pmol)</th>
<th>Relative Peak Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional HPLC</td>
<td>3.2-4.6 mm</td>
<td>0.5-2.0 mL/min</td>
<td>$10^2$-$10^5$</td>
<td>1</td>
</tr>
<tr>
<td>Microbore HPLC</td>
<td>1.5-3.2 mm</td>
<td>100-500 µL/min</td>
<td>$50$-$10^4$</td>
<td>5</td>
</tr>
<tr>
<td>Micro LC</td>
<td>0.5-1.5 mm</td>
<td>10-100 µL/min</td>
<td>$5$-$10^3$</td>
<td>20</td>
</tr>
<tr>
<td>Capillary LC</td>
<td>150-500 µm</td>
<td>1-10 µL/min</td>
<td>$0.5$-$500$</td>
<td>200</td>
</tr>
<tr>
<td>Nano LC</td>
<td>10-150 µm</td>
<td>10-1000 nL/min</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 1.1** Characteristics of miniaturised versions of HPLC

The coupling of capillary LC to MS has a number of advantages including: i) online desalting of samples; ii) concentration of components in the sample; and iii) separation of components in a complex mixture (Moritz *et al.*, 1994b). Techniques for increasing the available analysis time for a peak eluting from a column have also been explored. “Peak parking” and variable flow chromatography may increase the analysis time of a selected component by up to a factor of 10 (Davis and Lee, 1998).

Many examples of the use of capillary LC-MS and LC-MS/MS for the identification of peptides are available. Application of capillary LC-MS/MS to in-gel digested proteins from 2-dimensional electrophoresis has been particularly widespread (Moritz *et al.*, 1994a; Reid *et al.*, 1998; Heller *et al.*, 2000). More recently, with the focus on functional proteomics and to avoid the labour intensive use of 2-dimensional electrophoresis, capillary LC-MS has been used in conjunction with other separations such as 1-dimensional electrophoresis, size-exclusion chromatography, affinity
chromatography and ultrafiltration (Marvin et al., 2000; Heller et al., 2000; Opitek et al., 1998; McCormack et al., 1997; Raida et al., 1999).

1.2 APPLICATION OF MASS SPECTROMETRY TO CONVENTIONAL PROTEOMICS

The ‘proteome’ is defined as the total protein complement that is expressed by a genome or tissue (Wilkins et al., 1997). Analysis of the proteome of a given cell type, tissue, or organelle has typically involved separation and visualisation of the proteins by 2-dimensional electrophoresis, followed by excision and identification of the proteins. Mass spectrometry is routinely used for identification of in-gel digested proteins, usually in combination with database searching.

1.2.1 Two-dimensional electrophoresis

Two-dimensional electrophoresis (2DE) uses the combined techniques of isoelectric focusing (IEF) and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to enable separation and visualisation of the proteins in a particular sample. IEF, which is used for the first dimension, separates proteins by their pI, whilst SDS-PAGE in the second dimension separates the proteins according to their molecular mass (Herbert et al., 1997).

Many developments have occurred since the introduction of 2DE in the 1970s that have resulted in vast improvements in the resolution and reproducibility of 2DE gels. IEF is now performed in precast gels that contain an immobilised pH gradient (IPG), whereas previously this separation was dependent on carrier ampholytes (Hanash et al., 1987; Gorg et al., 1988; Mackin et al., 1991). IPG gels have advantages over carrier ampholytes including an increased loading capacity, improved stability in focusing, the ability to prepare narrow gradients, enhanced reproducibility and the absence of
gradient drift (Hanash et al., 1987; Gorg et al., 1988; Bjellqvist et al., 1993; Herbert et al., 1997; Herbert et al., 1998). The sample can be applied to the entire IPG gel by rehydrating the gel in sample solution (Rabilloud et al., 1994; Sanchez et al., 1997). This in-gel, sample rehydration avoids precipitation problems, eliminates cross-contamination, maximises reproducibility and is suitable for small volumes.

Advances have also occurred in sample preparation techniques. Solubilisation of hydrophobic membrane proteins, which have a tendency to aggregate and precipitate, is typically achieved with detergents such as SDS. SDS, however, interferes with IEF by masking the individual pIs of the proteins. Solubilisation for IEF must therefore be attained with non-ionic or zwitterionic detergents (such as 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS)) (Schupbach et al., 1991). Further improvements in protein solubility have been achieved by using tributylphosphine (TBP) rather than dithiothreitol as a reducing agent (Herbert et al., 1998). Combining chaotropes (urea and thiourea), detergents (CHAPS and sulfobetaine 3-10) and TBP in a stepwise extraction procedure (differential solubilisation) has also been shown to improve representation of proteins (Molloy et al., 1998). Prefractionation techniques have also been introduced to simplify samples prior to 2DE and thereby achieve representation of increased numbers of proteins (Corthals et al., 1997; Zuo and Speicher, 2000). Subcellular fractionation is often used to reduce the complexity of samples (Herbert et al., 1997; Rabilloud et al., 1998; Jung et al., 2000).

Advantages of separating proteins on gels include: i) the ability to store proteins in the gel for long periods of time; ii) qualitative and, in some cases, quantitative information is obtained through visualisation of the proteins; iii) gels are capable of separating complex samples; and iv) preliminary molecular mass and pI information is obtained for the proteins (Corthals et al., 2000a). There are, however, limitations and
disadvantages associated with 2DE (Herbert et al., 1997). Proteins with high or low pIs (> 8 or < 4) and very large proteins are difficult to resolve. The separation of hydrophobic proteins by 2DE is challenging, even with advances in solubilisation techniques. Proteins of low abundance are often not detected by 2DE despite the use of narrow range IPG gels, prefraccionation and differential solubilisation. The widespread use of similar techniques and implementation of automated procedures in large-scale laboratories has resulted in improved reproducibility within and between laboratories. Small changes in buffers, gel preparation or staining techniques, however, have a marked effect on the resulting gels and often affect gel reproducibility. Furthermore, 2DE is a time-consuming and labour-intensive technique. Nonetheless, there is currently no other technique that matches the ability of 2DE for separating and visualising thousands of proteins. Therefore the role of 2DE in the characterisation of the proteome continues to be crucial.

1.2.2 Database Searching with MS information

To identify proteins from 2DE gels, the spots are excised and enzymatically cleaved, usually with trypsin. The “in-gel” digested proteins are then analysed by MS. There are a number of ways in which protein or translated oligonucleotide sequence databases may be interrogated with MS information. The two most widely used are: i) peptide mass fingerprinting; and ii) peptide sequence tags. The use of sequence information alone may also be used in a “sequence-only” search.

Databases

Table 1.2 shows some of the most widely used databases that are available via the internet. Other databases are also available and often incorporate entries from a number of different databases. Although Swiss Prot is one of the smallest databases, it is the
most highly annotated and has low entry redundancy. After evaluation, entries from the TrEMBL database are added to the Swiss Prot database (and removed from TrEMBL) (Bairoch and Apweiler, 2000). The resulting database has a low frequency of errors, with extensive information including species origin, post-translational modifications, structural features of the sequence and references associated with the protein.

<table>
<thead>
<tr>
<th>Database name</th>
<th>Type of database</th>
<th>URL address</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWISS PROT</td>
<td>Protein and translated oligonucleotide sequence</td>
<td><a href="http://www.expasy.ch/sprot/sprot-top">www.expasy.ch/sprot/sprot-top</a></td>
<td>Extensive annotation; low error rate; low redundancy</td>
</tr>
<tr>
<td>TrEMBL</td>
<td>Translated oligonucleotide sequence</td>
<td><a href="http://www.expasy.ch/srs5">www.expasy.ch/srs5</a></td>
<td>Translation of the EMBL nucleotide sequence database; extensive annotation</td>
</tr>
<tr>
<td>GenPept</td>
<td>Translated oligonucleotide sequence</td>
<td>Download site from GenBank site</td>
<td>Translation of the GenBank nucleotide sequence database; Moderate error rate; Significant redundancy; little additional annotation</td>
</tr>
<tr>
<td>International Protein Sequence Database (PIR)</td>
<td>Protein sequence</td>
<td><a href="http://www.nbrf.georgetown.edu">www.nbrf.georgetown.edu</a></td>
<td>High level of annotation; low error rate; low redundancy</td>
</tr>
<tr>
<td>OWL</td>
<td>Protein and translated oligonucleotide sequence</td>
<td><a href="http://www.bioinf.man.ac.uk/dbbrowser/OWL">www.bioinf.man.ac.uk/dbbrowser/OWL</a></td>
<td>Composite of SWISSPROT, GenBank, PIR and NRL-3D; low redundancy</td>
</tr>
<tr>
<td>Protein DataBank</td>
<td>Translated oligonucleotide sequence</td>
<td><a href="http://www.ddbj.nig.ac.jp/PDBretriever/pdb_retriever">www.ddbj.nig.ac.jp/PDBretriever/pdb_retriever</a></td>
<td>Translation of the DNA DataBank of Japan</td>
</tr>
<tr>
<td>dbEST</td>
<td>Expressed sequence tags</td>
<td><a href="http://www.ncbi.nlm.nih.gov/dbEST/index">www.ncbi.nlm.nih.gov/dbEST/index</a></td>
<td>High error rate</td>
</tr>
</tbody>
</table>

Table 1.2 Protein and translated oligonucleotide sequences available via the Internet
Expressed Sequence Tags (ESTs) are short nucleotide sequences (approximately 250-400 base pairs in length) generated by sequencing randomly selected portions of DNA (Bairoch, 1997; Kinter and Sherman, 2000). A large amount of data is generated by this approach but the error rates are high. ESTs can, however, be of use when other databases do not result in retrieved sequences from the available information. Peptide mass fingerprinting data cannot be used to search EST databases, however, since the EST database information is fragmentary and does not represent whole proteins.

**Peptide mass fingerprinting**

MALDI-MS analysis of in-gel digested proteins provides molecular mass information on the peptides from the digest. The list of peptide masses that is generated ("peptide mass fingerprint") provides a specific tool that can be used to search databases (Henzel et al., 1993; Mortz et al., 1994; Patterson and Aebersold, 1995). The list of MALDI masses is compared to theoretically generated lists of digested proteins from the database ('in silico' digests). Proteins that have a specified number of peptides that match the MALDI masses are retrieved by the search program and displayed in a ranked list, along with information on the peptides that were matched. Peptide masses from ESI data may also be used for peptide mass fingerprinting, providing the charge on the peptides is known.

Table 1.2 lists some of the databases that are available via the internet for database searching with MALDI mass fingerprint data. Different search programs may contain different options for increasing the specificity of the search. These parameters may include pI range, protein molecular mass range, amino acid modifications and species identification.
MALDI mass fingerprinting is a rapid technique that requires minimal sample preparation following extraction of the peptides from the gel, although improved coverage of the peptides is usually obtained if the samples are desalted prior to analysis (Erdjument-Bromage et al., 1998; Courchesne and Patterson, 1997; Kussmann et al., 1997). Peptide coverage by MALDI-MS is also dependent on the response factors of individual peptides. It has been noted that peptides with C-terminal lysine residues have lower response factors compared to peptides with C-terminal arginine residues (Krause et al., 1999; Brancia et al., 2000). To improve peptide coverage, derivatisation of peptides has been successfully introduced (Brancia et al., 2000; Brancia et al., 2001).

If more than one protein is present in a particular gel-spot or sample, the protein origin of the peptides generated by enzymatic cleavage cannot be determined. Therefore MALDI-MS mass fingerprinting has limited ability for the analysis of mixtures of proteins. In some cases, the correct identification of a protein may be obtained from peptide mass fingerprinting with a mixture of peptides from different proteins. The peptides that are not accounted for by the first protein identification can be searched in a “second-pass” search to attempt to identify other components (Patterson and Aebersold, 1995). This type of search may not, however, result in a correct identification and false positive matches are also possible. With complex mixtures of proteins other search strategies must be used.

**Peptide Sequence Tags**

Peptide sequence tags are generated from MS/MS of the peptides in a digest and consist of three regions: i) the mass of the peptide fragment on the N-terminal side of the chosen sequence; ii) a short sequence of amino acids; and iii) the mass of the peptide fragment on the C-terminal side of the chosen sequence (Mann and Wilm, 1994). The sequence tag, peptide mass and fragment ion type are entered in to a search program.
Table 1.3 lists some database searching programs that are available via the internet. Retrieved sequences are compared to the entire MS/MS spectrum for confirmation of the correct match. Sequence tag searching is more specific than mass fingerprinting and does not require a large representation of peptides from a parent protein. In some cases one or two peptides may be sufficient for protein identification.

**Sequence searching without mass information**

MS/MS of tryptic peptides often generates extensive sequence information that can be interpreted relatively easily with experience in peptide fragmentation. With a small stretch of amino acids (usually five or more residues), specific searching of databases may also be achieved. “Sequence only” searching does not use the masses on either side of the stretch of sequence and does not require the peptide mass. The exclusion of mass terms in the search eliminates problems that may be encountered when modified peptides are analysed.

<table>
<thead>
<tr>
<th>Program Name</th>
<th>Search Type</th>
<th>URL address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide Search (PepSea)</td>
<td>Mass fingerprint, Sequence Tag, Sequence only</td>
<td>pepsea.protana.com</td>
</tr>
<tr>
<td>MS-Fit</td>
<td>Mass fingerprint</td>
<td>jpsl.ludwig.edu.au/ucsfhtml3.4/msfit.</td>
</tr>
<tr>
<td>MS-Tag</td>
<td>Sequence Tag</td>
<td>jpsl.ludwig.edu.au/ucsfhtml3.4/mstag</td>
</tr>
<tr>
<td>MS-Seq</td>
<td>Sequence Tag</td>
<td>jpsl.ludwig.edu.au/ucsfhtml3.4/msseq</td>
</tr>
<tr>
<td>MS-Pattern</td>
<td>Sequence Only</td>
<td>jpsl.ludwig.edu.au/ucsfhtml3.4/mspattern</td>
</tr>
<tr>
<td>TagIdent</td>
<td>Sequence Only</td>
<td>au.expasy.org/tools/tagident</td>
</tr>
<tr>
<td>PeptIdent</td>
<td>Mass fingerprint</td>
<td>au.expasy.org/tools/peptident</td>
</tr>
<tr>
<td>PepFrag</td>
<td>Sequence Tag</td>
<td>prow11.rochefeller.edu/prowl/pepfragch.</td>
</tr>
<tr>
<td>ProFound</td>
<td>Mass Fingerprint</td>
<td>129.85.19.192/profound_bin/WebProFound.exe</td>
</tr>
</tbody>
</table>

**Table 1.3** Search programs that are available via the Internet for interrogation of databases with MS data (where possible, the Australian sites are listed).
The application of MS for the identification of gel-separated proteins has resulted in a flood of research papers detailing examples of proteomics applications. Interesting examples have included the analysis of human melanoma proteins (Clauser et al., 1995), ion-channel proteins (Buhler et al., 1998), apoptosis-associated proteins (Otto et al., 1998; Brockstedt et al., 1998; Gygi et al., 1999a) and human colon cancer (Ji et al., 1994). There have also been numerous reviews discussing the merits and weaknesses of different search strategies (Patterson and Aebersold, 1995; Shevchenko et al., 1996b; Jungblut et al., 1996; Humphrey-Smith et al., 1997; Jensen et al., 1998; Lahm and Langen, 2000; Harry et al., 2000; Aebersold and Goodlett, 2001).

1.3 FUNCTIONAL PROTEOMICS

1.3.1 Limitations of Conventional Proteomics

Conventional proteomics, i.e. 2DE electrophoresis followed by MS identification of proteins, has become a valuable tool for the identification of proteins in very complex samples. The establishment of proteomes of different organisms (such as yeast) (Shevchenko et al., 1996a), organelles (e.g. the mitochondrion) (Jung et al., 2000; Rabilloud et al., 1998) and tissues (e.g. the myocardium) (Thiede et al., 1996; Otto et al., 1996) will provide information to assist in discovery of proteins that are affected by particular treatments or in certain diseases. The high level of automation has improved the throughput in large-scale laboratories and also increased reproducibility of the results (Quadroni and James, 1999; Harry et al., 2000).

Conventional proteomics, however, does not provide information on protein function and also has limitations in the analysis of low level proteins, highly basic or acidic proteins, and proteins with high or low molecular weights (Corthals et al., 2000b). Hydrophobic proteins are also difficult to analyse with conventional proteomics.
techniques. A further limitation in this approach is the difficulty of obtaining quantitative representation of the proteins (Blackstock and Weir, 1999; Aebersold and Goodlett, 2001).

The need for more information on the proteins that are identified by proteomics has resulted in the development of new proteomic technologies and increasing interest in methods that do not use 2DE. Moreover, attention has begun to focus on gathering information on the protein function from techniques used to separate the proteins prior to MS analysis.

### 1.3.2 Circumventing 2DE

The increasing use of LC-MS and LC-MS/MS data for database searching and the ability of these techniques to analyse complex mixtures has begun to reduce the need for separation of proteins prior to MS analysis. Chromatography-based methods are emerging for the partial separation of proteins prior to MS analysis (Opitek et al., 1998; Aebersold and Goodlett, 2001). The use of ion exchange chromatography and/or affinity chromatography prior to LC-MS/MS results in simplification of complex samples without the need for 2DE.

To obtain quantitative information on protein expression levels, techniques using isotope labelling have been developed (Gygi et al., 1999b; Mann, 1999). Two forms (a heavy isotope form and a light isotope form) of a derivatisation agent are used separately to label a reference sample and a second (treated or diseased) sample. The two samples are then combined, enzymatically cleaved and the labelled peptides separated by affinity chromatography. MS analysis reveals peptides derivatised with heavy labels and peptides derivatised with light labels. Equivalent peptides that differ only by the label should behave the same way and the ratio of the heavy isotope form to
the light isotope form provides a measure of the change in abundance that has occurred. In this way, quantitative information can be obtained from mass spectrometry of proteins.

1.3.3 Functional Proteomics

The focus of genomics and proteomics has been the large-scale identification of gene sequences and the proteins that are expressed by those genes. The main question that is left unanswered by these approaches is what is the function of the proteins that are expressed? The main aim of functional proteomics is to develop methods for relating the proteins that are expressed by a cell to their cellular function (Godovac-Zimmermann and Brown, 2001).

Assignment of protein function requires integration of a number of techniques (Blackstock and Weir, 1999). Important information can be gained from analysis of complexes formed between proteins. Approaches to the identification of protein binding partners have included cross-linking, co-immunoprecipitation (using antibodies), coprecipitation (using an affinity tag) and affinity chromatography (with antibodies or affinity tags) (Yu and Gaskell, 1998; Rudiger et al., 1999; Harry et al., 2000; Yates, 2000; Figeys et al., 2001). Following the use of one or more of these techniques to isolate the protein complex, the sample can be analysed directly by LC-MS/MS or subjected to 1-dimensional SDS-PAGE prior to MS analysis.

An example of an affinity method is the Tandem Affinity Purification (TAP) method (Puig et al., 2001). A known protein is tagged and introduced to the cell. The protein interacts with binding partners in the cell and the complex is recovered by affinity chromatography prior to MS analysis. Proteins that act both directly and indirectly with the tag protein are identified.
Another example, reported by McCormack et al., is the application of LC-MS/MS in identification of proteins from a complex mixture isolated using an affinity chromatography method (McCormack et al., 1997). The protein of interest was fused with glutathione S-transferase (GST), which was then noncovalently bound to immobilised glutathione. Cell lysate was passed over the column to retrieve proteins that bound to the protein fused to GST. The bound proteins were then retrieved and analysed by LC-MS/MS.

Biotinylation has also been used for the capture of protein complexes. Identification of proteins in the spliceosome complex was obtained following affinity isolation and separation of the complex by 2DE (Neubauer et al., 1998).

As an alternative to MS-based techniques, the yeast two-hybrid system has been used for high screening of protein binding pairs. This system, however, suffers from a high rate of false positives and does not identify indirect binding partners (Pandey and Mann, 2000; Aronheim, 2001; Toby and Golemis, 2001). To overcome some of the limitations of the yeast two-hybrid system, other methods have been introduced including the Sos recruitment system (SRS) and the Ras recruitment system (RRS) (Aronheim, 2001).

1.3.4 Future Directions

Techniques for the purification of protein complexes prior to mass spectrometry are currently performed on a small scale. Further development of these methods, and the application of automated procedures, should prove valuable in gaining insight into protein binding partners and hence protein function. Developments in bioinformatics will continue to support the development of functional proteomics.
Another field of development that may prove integral in functional proteomics is the use of protein chips. In its simplest form, proteins are immobilised on a chip and a sample of interest is applied to the surface. Proteins that bind to the immobilised protein are retained upon washing and can then be eluted and analysed by MS (Pandey and Mann, 2000). Developments in this technology may lead to further advances in functional proteomics.

Already 2DE is becoming a technique that is more and more confined to specialist laboratories with high levels of automation ensuring maximised throughput and reproducibility. Continued developments are predicted to result in less dependence of 2DE and an increased focus on techniques that provide information on protein function. The use of mass spectrometry in the identification of proteins from isolated complexes is of great interest. Furthermore, mass spectrometry has also proved useful in the direct analysis of noncovalent complexes (see Chapter 5).

1.4 OUTLINE OF THIS WORK

The extensive use of mass spectrometry in protein chemistry is clear from the examples mentioned above. The role of mass spectrometry in the characterisation of proteins in proteomics applications is well established and the increasing numbers of functional proteomics publications reveal the growing importance of mass spectrometry in this area. The use of mass spectrometry in the analysis of the primary structure of peptides and proteins is widespread as a result of the sequence information that is generated by MS/MS. The application of mass spectrometry for studying higher order protein structure, however, is not as widely accepted but has been, and continues to be, explored by a significant number of research groups.
The overall aim of this work is to provide insight into a number of applications to
which electrospray ionisation mass spectrometry (ESI-MS) can be applied and therefore
to address the advantages and disadvantages of the use of ESI-MS in these analyses. Three different applications of ESI-MS are presented.

A cellular process that has evoked extensive research in recent years is apoptosis, which is a highly regulated form of cell death. The importance of an increase in the permeability of the mitochondrial membrane during apoptosis has been recognised but the mechanism(s) by which this occurs are not fully understood. The subcellular fractionation and comparison of normal mitochondria and mitochondria that have experienced an increase in membrane permeability is an area well suited to proteomic analysis. The application of conventional proteomics to the analysis of proteins released from mitochondria is presented in Chapter 3.

Another area of extensive research interest is the entry of the human immunodeficiency virus (HIV) into cells. The T cell surface protein CD4 is known to be the main cellular receptor for HIV. The identification of proteins that associate with CD4 may therefore provide targets for the development of therapeutics to prevent the entry of HIV into cells. The development of a method for the identification of potential CD4-associating proteins and the results obtained are presented in Chapter 4. The identification of proteins that co-elute with CD4 from an immunoaffinity column provides an example of the use of functional proteomics and the potential that this type of approach has in the characterisation of important binding partners.

The analysis of noncovalent complexes of the protein glutathione S-transferase (GST) with various ligands is presented in Chapter 5. The GST enzyme system has been extensively characterised by conventional methods for protein structure analysis such as X-ray crystallography, nuclear magnetic resonance and fluorescence. GST is
therefore a useful model protein for investigation of the behaviour of noncovalent complexes by ESI-MS. Furthermore, a recent study of GST with a number of ligands by ESI-MS, raised some interesting questions regarding the complexes that were apparent in the ESI-MS spectra (Ishigai et al., 2000). The aims of the analysis of complexes of GST and selected ligands were therefore to investigate ligands that had not previously been used, to answer some of the questions raised by the previous study and to provide further insight into the behaviour of noncovalent complexes in the gas phase.
2.1 PREPARATION OF MITOCHONDRIAL PROTEINS FOR 2-DIMENSIONAL ELECTROPHORESIS

2.1.1 Mitochondria and permeability transition

Treated and untreated mitochondria from mouse liver were kindly provided by Associate Professor Mark Wilson and Ms Alison Smail (Department of Biological Sciences, University of Wollongong). The mitochondria were prepared by differential centrifugation and were induced to undergo permeability transition (PT) using protoporphyrin IX (PPIX). To induce low conductance PT (LC-PT), PPIX was added to the buffer to obtain a final concentration of 20 μM or 40 μM PPIX. To induce high conductance PT (HC-PT) the final concentration of PPIX was 4 mM. Following treatment, the mitochondria were washed 3 times with isotonic media whilst they were still intact and the supernatant was collected. The mitochondrial pellet and supernatant were then frozen at -80°C until required.

Attempts to remove PPIX included precipitation of the protein with ethanol or acetone (both of which caused precipitation of the PPIX along with the protein), dialysis against 20% dimethyl sulfoxide (DMSO) and extraction with diethyl ether. Of these attempts, the extraction with diethyl ether showed some success. The sample solution was acidified with acetic acid (to a final concentration of 0.875 mM CH₃COOH in a volume of 50 μL) and extracted 3 times with 100 μL diethyl ether. A small amount of bromophenol blue (BPB) was added to the solution to help visualise the aqueous and organic layers. The aqueous layer was recovered and used for 2-dimensional electrophoresis. The loss of some protein was apparent when extracted and non-
extracted control samples were compared, hence it was performed only on samples from HC-PT mitochondrial supernatants, which had a higher concentration of PPIX compared to the LC-PT mitochondria.

2.1.2 Chemicals and consumables for electrophoresis

Precast immobilised pH gradient (IPG) gels (Immobiline™ DryStrip pH 3-10 NL) were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden) in 7 cm and 13 cm formats. Pharmalyte™ 3-10 (which contains carrier ampholytes) was also purchased from Amersham-Pharmacia Biotech.

Precast polyacrylamide mini-gels (Ready Gel® Tris-HCl gels, 10-20% resolving gel, 4% stacking gel) were purchased from Bio-Rad Laboratories (Hercules, U.S.A.). Molecular weight markers (Broad Range SDS-PAGE standards) were also purchased from Bio-Rad Laboratories.

Urea, thiourea, Tris(hydroxymethyl)aminomethane (tris), sodium dodecyl sulphate (SDS), dithiothreitol (DTT), 3-([3-Cholamidopropyl]dimethylammonio)-1-propanesulfonate (CHAPS), iodoacetamide, bromophenol blue (BPB), acrylamide, N,N,N',N'-Tetramethylethylene diamine (TEMED), ammonium persulphate (APS), agarose, glycerol and tributylphosphine (TBP) of electrophoresis grade or the highest available grade were purchased from ICN Biomedicals (Costa Mesa, U.S.A).

CHAPS, SDS, sodium azide, N-decyl-N,N-dimethyl-3-ammonio-1-propanesulfonate (SB 3-10) and 40% acrylamide/bisacrylamide solution of the highest purity were purchased from Sigma-Aldrich (St Louis, U.S.A).

Thiosulphate, sodium carbonate, silver nitrate, formaldehyde, urea, ammonium bicarbonate methanol, acetic acid, ethanol and acetonitrile were of the highest available
quality from the Ajax™ range of chemicals (Asia Pacific Speciality Chemicals Seven Hills, Australia).

Sequencing-grade modified trypsin was purchased either from Boehringer Mannheim (Roche Diagnostics, Berkeley, U.S.A.) or from Promega Corporation (Madison, U.S.A.)

MilliQ™ water (Millipore Corporation, Bedford, U.S.A.) was used in the preparation of all solutions.

2.1.3 Extraction of proteins from mitochondria

Three extractions were performed on the pelleted mitochondria as described in the Australian Proteome Analysis Facility (APAF) 2D Electrophoresis Manual (May 18th-21st 1998) (APAF, 1998), which is similar to the method published by the same group (Molloy et al., 1998). The first extraction used 40 mM tris and removed water-soluble components from the mitochondria. Two hundred µL of 40 mM tris and 150 units of endonuclease were added to the pelleted mitochondria. The solution was vortexed and sonicated then left at room temperature for 20 min. The mitochondria were centrifuged at 12 000 rpm for 10 min (Eppendorf benchtop 5415C centrifuge). The supernatant was removed and concentrated to less than 10 µL by vacuum centrifugation (SpeedVac SC110, Thermo Savant, New York, U.S.A.). Buffer for IEF (8 M urea, 4% CHAPS, 2 mM TBP, 0.5% ampholytes and 40mM tris) was added to the concentrated sample.

The second extraction solution consisted of 8 M urea, 4% CHAPS, 2 mM TBP, 0.5% carrier ampholytes and 40mM tris. This solution (250 µL) was added to the mitochondrial pellet from the previous extraction and the solution was left at room temperature for 1 hour with occasional sonication. The solution was centrifuged for
approximately 20 min at 14 000 rpm. The collected supernatant was used for re-
hydration of an IPG gel.

A solution (250 μL) containing 5 M urea, 2 M thiourea, 2% CHAPS, 2% SB 3-10,
0.5% carrier ampholytes, 40 mM tris and 2 mM TBP was added to the mitochondrial
pellet following the second extraction. This solution was left at room temperature for 40
min with occasional sonication. The solution was then centrifuged for 15 min at 14000
rpm and the collected supernatant was used for re-hydration of an IPG gel.

---

**2.2 TWO-DIMENSIONAL ELECTROPHORESIS AND PREPARATION FOR MASS SPECTROMETRIC ANALYSIS**

### 2.2.1 Isoelectric focusing

Samples from the concentrated supernatant and water-soluble extraction were dissolved
in 8 M urea, 4% CHAPS, 2 mM TBP, 0.5% carrier ampholytes and 40 mM tris and then
loaded into the tray for re-hydration of IPG gels. The samples from the second and third
extractions were loaded directly into the tray for IPG gel re-hydration. Gel re-hydration
was performed overnight or for a minimum of ten hours.

Following re-hydration, IEF was performed at 20°C in the Multiphor™ IEF unit
(Amersham Pharmacia Biotech, Uppsala, Sweden) using a voltage gradient program.
An example of a daytime gradient program for 13 cm long IPG gels is given in Table
2.1. The gels were focussed for a total of 20 802.5 volt hours for the program shown in
Table 2.1. In overnight programs the gels were focussed for a total of 21 571 volt hours
with a total run time of 11.35 hours. For the 7 cm long IPG gels, the gels were focussed
for 17 755 volt hours for daytime programs. If an overnight program was used, the 7
cm long gels were focussed for 20 213 volt hours.


<table>
<thead>
<tr>
<th>Hours</th>
<th>Cumulative time</th>
<th>Voltage (V)</th>
<th>Volt hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>150</td>
<td>15</td>
</tr>
<tr>
<td>0.5</td>
<td>0.7</td>
<td>500</td>
<td>162.5</td>
</tr>
<tr>
<td>0.5</td>
<td>1.2</td>
<td>1000</td>
<td>375</td>
</tr>
<tr>
<td>2.0</td>
<td>3.2</td>
<td>3500</td>
<td>4500</td>
</tr>
<tr>
<td>4.5</td>
<td>7.7</td>
<td>3500</td>
<td>15750</td>
</tr>
</tbody>
</table>

Table 2.1 Voltage supply program for IEF of 13cm IPG gels with daytime focussing.

2.2.2 SDS-PAGE

Following IEF, those gels that were not immediately equilibrated for the second dimension were stored at -80°C. The equilibration solution for SDS-PAGE consisted of 6 M urea, 2% SDS, 20% glycerol, 5 mM TBP, 2.5% acrylamide (monomer) and 0.375 M tris. The gels were soaked in this solution for approximately 20 min. The equilibrated IPG gels were then embedded on the top of polyacrylamide gels with a solution of 1% agarose that contained a small amount of BPB. If molecular weight markers were used they were also embedded with 1% agarose in a separate lane. The precast minigels were supplied with a lane for molecular weight markers. For the gels that were cast in-house, a cut drinking straw was used as a marker lane.

The 7 cm IPG gels were embedded on the top of precast Ready Gels™, whilst the 13 cm IPG gels were embedded on the top of 1mm thick 12% polyacrylamide gels that had been cast in-house. The gels were then run under constant current conditions using a discontinuous buffer system. The anode buffer used was 0.75 M tris with 0.005% sodium azide, adjusted to pH 8.8. The cathode buffer (which was placed in the upper portion of the tank) was 192 mM glycine with 0.1% SDS, adjusted to pH 8.3. The
Chapter 2: Materials and Methods

Cathode buffer was made freshly for each run, whereas the anode buffer solution was used for up to 20 gels. For mini gels, the Ready Gel Cell™ (Bio-Rad Laboratories Hercules, U.S.A.) was run without cooling at 5 mA constant current per gel for 30 minutes and then 12 mA constant current per gel for 20 min after the bromophenol blue (BPB) front had run off the gels (usually 2 to 2.5 hours). The larger format gels (14 cm x 16 cm) were run in the Hoefer™ 600SE unit (Amersham Pharmacia Biotech, Uppsala, Sweden) at 5 mA constant current per gel for 2 hours, followed by 20 mA constant current per gel for approximately 30 min after the BPB had left the gel. The current was often increased to 30-40 mA per gel when the BPB front was approximately 1 cm from the gel bottom. A circulating water cooler maintained the temperature of the Hoefer™ unit at 4°C.

2.2.3 Silver Staining

Following SDS-PAGE, the gels were removed from the unit and placed in 50% methanol, 10% acetic acid solution to fix the proteins. The gels were left in fixer for 30 min or overnight. The gels were then stained with silver, using the protocol from the Garvan Institute of Medical Research, Sydney, Australia (Dr Garry Corthals, personal communication) with minor modifications. Following fixing, the gels were equilibrated in 5% methanol for 15 min, washed 3 times for 5 min each in MilliQ™ water, sensitised with sodium thiosulphate (0.2 g/L) for 2 min then washed 3 times for 1 min each in MilliQ™ water. Incubation in a solution of 0.2% (w/v) silver nitrate was performed for 25 min followed by three washes with MilliQ™ water for 1 min each. The silver stain was then developed with a solution containing sodium carbonate (3% w/v), formaldehyde (0.05% v/v) and sodium thiosulphate (2% of the solution used for sensitisation). The development was stopped using a 5% solution of acetic acid. The
gels were then thoroughly washed with MilliQ™ water. Fixed and stained gels were stored at 4°C in plastic bags or the staining trays prior to in-gel digestion.

2.2.4 In-gel tryptic digestion

Protein spots were cut from the 2DE gels and placed in microcentrifuge tubes. Acetonitrile was added to completely cover the gel and was left to dehydrate the gel for approximately 10 minutes. The acetonitrile was removed and the gel pieces dried by vacuum centrifugation. The gel pieces were then covered with 12.5 ng/µL trypsin in 50 mM ammonium bicarbonate and rehydrated at 4°C for 45 min. If, after rehydration with the trypsin solution, the gel pieces were no longer covered with solution, 50 mM ammonium bicarbonate was added to cover the gel. The tubes were then placed in a circulating water bath at 37°C and incubated overnight.

The supernatant from the gel pieces was collected and 20 mM ammonium bicarbonate was added to cover the gel. The gel pieces were sonicated for 5 min then left at room temperature for 15 min with occasional mixing. The pieces were centrifuged, the supernatant collected and added to the first supernatant. Three further 20 min extractions were performed with 50% acetonitrile, 5% formic acid. The combined supernatants were concentrated by vacuum centrifugation to less than 10 µL and desalted using C_{18} ZipTips™ (Millipore Corporation, Bedford, U.S.A.)(see section 2.4.1).

2.3 PREPARATION OF ANTI-CD4 SAMPLES

2.3.1 Affinity capture of CD4 and co-associating proteins

Preparation of samples by affinity capture of CD4 was performed by either Dr Andrew Sloane or Mr Oliver Bernhard at the Millennium Institute, Centre for Virus Research
(University of Sydney, Westmead, Australia). CEMT4 cells were prepared and applied to an IgG_i column and an anti-CD4 column as previously described (Sloane, 2000; Lynch et al., 1999). Briefly, the supernatant following cell lysis was passed over a sepharose column with mouse IgG_i coupled to the sepharose. This column was connected to a similar column with anti-CD4 Q425 antibody coupled to the sepharose in place of IgG_i. After column washing, the bound protein was eluted. Samples were collected from elution of both the columns and also from the IgG_i column alone.

Samples from the affinity columns in were reduced in 10 mM dithiothreitol, 6 M urea and alkylated with 40 mM iodoacetamide. After desalting on an NAP-10 desalting column, (Amersham Pharmacia Biotech, Uppsala, Sweden) samples were digested overnight with modified trypsin (Promega Corporation, Madison, U.S.A.).

2.3.2 Anion Exchange Chromatography

Anion exchange chromatography for removal of the remaining detergents from the digested peptides was also performed by either Dr Andrew Sloane or Mr Oliver Bernhard at the Millennium Institute, Centre for Virus Research (University of Sydney, Westmead, Australia). The digested samples were loaded on to a MONO-Q™ PC 1.6/5 anion exchange column (Amersham Pharmacia Biotech, Uppsala, Sweden) attached to a SMART™ micropurification HPLC system (Amersham Pharmacia Biotech, Uppsala, Sweden). The column was loaded and washed with 20 mM glycine (pH 10.0) and peptides were eluted with 350 mM NaCl in 20 mM glycine (pH 10.0). The elution of peptide was monitored by UV absorbance at 280 nm and 214 nm using a SMART™ μPeak monitor (Amersham Pharmacia Biotech, Uppsala, Sweden). The collected peptides were concentrated in a vacuum centrifuge and desalted using C_{18} ZipTips® (Millipore Corporation, Bedford, U.S.A.) (see section 2.4.1).
2.4 MASS SPECTROMETRY FOR PROTEOMICS APPLICATIONS

2.4.1 Desalting samples for mass spectrometric analysis

An equivalent volume of 10% formic acid was added to the peptides that had been concentrated by vacuum centrifugation (from in-gel digests or anion exchange chromatography). If the samples had been completely dried, 10 μL of 5% formic acid was added and the sample was sonicated. C18ZipTips™ (Millipore Corporation, Bedford, MA, U.S.A.) were wet with 50% aqueous acetonitrile and equilibrated with 5% formic acid. The peptides were bound to the C18 by flushing the solution through the ZipTip™ 10-20 times. The ZipTip™ was then washed with 5% formic acid (10-40 μL). A small volume of 50% aqueous acetonitrile with 5% formic acid was pipetted into a separate microcentrifuge tube. The bound, desalted, peptides were eluted into this small volume. The volume used to elute the peptides was varied to obtain appropriate concentrations of peptides. A maximum of 10 μL was used but typically a volume of 5-8 μL was used. These solutions were then used directly for MALDI and nanoESI analyses.

For LC-MS and LC-MS/MS analyses, the desalted sample was carefully concentrated with nitrogen. Before the solution had been removed completely, 10 μL of 1% formic acid solution was added to the tube. This was concentrated again and another 10 μL of 1% formic acid solution added.

2.4.2 MALDI-MS analysis

For MALDI-MS analyses, 0.5 μL of the desalted sample was mixed with 0.5 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich, St Louis,
U.S.A.) in 50% aqueous acetonitrile with 0.01% trifluoroacetic acid (TFA) (Sigma-Aldrich, St Louis, U.S.A) and spotted on a MALDI target.

MALDI data were acquired with a Bruker Biflex™ MALDI-TOF instrument (Bruker Daltonics, Bremen, Germany). MALDI analyses were performed using gridless delayed extraction and a reflector. The accelerating voltage was approximately 19 kV with 14 kV on the IS2 focusing lens. Laser power was optimised for each sample preparation. The instrument was calibrated using a linear two-point calibration (with ACTH 18-39 (human, clip) and angiotensin II (AusPep, Parkville, Australia)).

2.4.3 NanoESI-MS and NanoESI-MS/MS
For nanoscale electrospray ionisation, desalted peptide samples were loaded into metal-coated borosilicate capillaries (made in-house or obtained from Micromass U.K., Manchester, U.K. or Protana, Odense, Denmark). Nanoscale ESI was performed on a quadrupole orthogonal acceleration time-of-flight mass spectrometer (Q-Tof™ or Q-Tof™2) (Micromass U.K., Manchester, U.K.) with the Nanoflow-Z™ interface. The capillary voltage was optimised for each capillary (between 0.7 and 1.2 kV) and the cone voltage set between 30 and 40 V. Data was acquired in continuum mode with a 2.4 s integration time and a 0.1 s interscan delay. During MS/MS, the resolution on the quadrupole was set to allow the transmission of all the precursor isotopes. Argon was used as the collision gas and the collision cell voltage\(^2\) was optimised between 20 and 40 V. Spectra were processed using MassLynx™ software (Micromass U.K., Manchester, U.K.). The instrument was calibrated using a solution of [Glu\(^1\)]-fibrinopeptide B (Sigma-Aldrich, St Louis, U.S.A.) at a concentration of 100 fmol/\(\mu\)L.

\(^2\) The voltage applied to the collision cell is often termed the collision energy. If the ion selected for MS/MS is singly charged the voltage applied to the collision cell is equal to the laboratory collision energy; for multiply charged ions the energy is equal to the voltage multiplied by the charge. The term ‘collision cell voltage’ is used herein.
MS/MS was performed on the doubly charged molecular ion (m/z 785) with a collision cell voltage of 32 V. A two point linear calibration using the product ions at m/z 72.1 and 1285.5 was applied.

2.4.4 Capillary LC-MS and LC-MS/MS

LC-MS was performed using a Waters CapLC® System with a 0.32 x 150 mm Symmetry® 100 C_{18}, 5 µm column (Waters Corporation, Milford, U.S.A.). A 1 µL injection was performed using the microlitre pickup mode. A solvent gradient of 3.8% B to 43.8% B in 18 min was used to separate the peptides at a flow rate of 2 µL/min (Solvent A: 1% aqueous formic acid; Solvent B: 80% aqueous acetonitrile, 1% formic acid). After 18 min, the percentage B was increased to 75%. A low dead volume union was used to connect the outlet from the column directly to the Nanoflow-Z™ interface, which was plumbed with 50 µm internal diameter fused silica.

A solvent delay of approximately 20 min was used to prevent data acquisition prior to elution of the peptides. Mass spectra were acquired between m/z 500 and 1200. An integration time of 1 s was used with a 0.1 s delay.

The threshold for MS to MS/MS switching was set between 3 and 5 ion counts and precursors with 2, 3 or 4 charges were selected for MS/MS. The switch back to MS mode was set to occur after each component had been subjected to MS/MS for a certain period of time. For more information on these parameters see Chapter 4.

2.4.5 Database searching

Database searching was performed using either Peptide Search (also known as PepSea) (http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html;
http://peptsearch.protana.com) (Mann and Wilm, 1994) or MS-Seq and MS-Pattern from Protein Prospector (http://prospector.ucsf.edu) (Clauser et al., 1999).

### 2.5 MASS SPECTROMETRY OF NONCOVALENT COMPLEXES

#### 2.5.1 Preparation of samples for the analysis of complexes formed between glutathione S-transferase and ligands

Glutathione S-transferase A1(C112S), glutathione S-transferase A1(Y9F) and glutathione S-transferase P1 were the kind gift of Professor Lu-Yun Lian (Biomolecular Science, UMIST, Manchester, U.K.). Wild-type glutathione S-transferase A1 was purchased from Oxford Biomedical Research (Michigan, U.S.A.). Glutathione (GSH), S-propylglutathione, S-hexylglutathione and S-decylglutathione were purchased from Sigma-Aldrich (Dorset, U.K.) and were of the highest purity available.

GST A1(C112S) was provided as a lyophilised powder and did not require extensive purification prior to mass spectrometry. Stock solutions of the protein were prepared with concentrations of between 100 and 500 pmol/μL. A small amount of BioRad AG-501-X8 mixed bed resin (Bio-Rad Laboratories, Hercules, U.S.A) was added to the stock solution and mixed for 10-20 minutes to minimise the formation of salt adducts. Aliquots of these solutions were stored at -20°C and diluted to 5 pmol/μL with 10 mM ammonium acetate (NH₄CH₃COO) prior to analysis. For molecular mass determination of the denatured protein, a 5 pmol/μL solution in 50% aqueous acetonitrile with 0.2% formic acid was used.

GST A1(Y9F) and GST P1 were supplied in buffered solutions with estimated concentrations of 1.5 mg/mL and 11 mg/mL respectively. These were desalted using membrane centrifugation cartridges (Microcon™ concentrators, 10K NMWL, Millipore Corporation, Bedford, U.S.A). The protein was washed with at least 4 times the initial
sample volume of MilliQ™ water. The concentrated, desalted protein was then collected in MilliQ™ water and the approximate concentration calculated from the estimated load and recovered volume. This stock solution was deionised with Bio-Rad AG-501-X8 mixed bed resin (Bio-Rad Laboratories, Hercules, U.S.A.) before aliquots were stored at -20°C. Samples for mass spectrometric analysis were prepared as for GST A1(C112S).

Recombinant human GST A1 (wildtype) (100 μg determined by Bradford method) was purchased as a buffered solution and required desalting. Ultrafiltration of the samples was performed as for GST A1(Y9F) and GST P1 and the resulting stock solution in water was deionised and stored at -20°C. Samples were prepared for mass spectrometry in the same way as for the other GST samples.

Solutions of GSH, S-propylglutathione, S-hexylglutathione and S-decylglutathione were prepared in 10 mM ammonium acetate to final concentrations of between 100 and 500 pmol/μL. S-hexylglutathione and S-decylglutathione required a small amount of dimethylsulphoxide (DMSO) for solubilisation. S-hexylglutathione solutions contained between 0.5% and 1% DMSO whereas the S-decylglutathione solution contained 1.6% DMSO.

An appropriate volume of ligand solution was added to a 2 μL aliquot of GST protein stock solution and then diluted with 10 mM ammonium acetate to a final protein concentration of 5 pmol/μL. For the titration of GST A1(C112S) with glutathione, molar ratios of protein to ligand ranged from 2:1 to 1:80. For the titrations of GST A1(C112S) with S-hexylglutathione and S-decylglutathione, molar ratios ranged from 20:1 to 1:40.
2.5.2 Mass spectrometry for the analysis of complexes of glutathione S-transferase

Data were acquired on a quadrupole time-of-flight mass spectrometer (QTof™ or QTof™2, Micromass U.K., Manchester, U.K.) with the Nanoflow-Z™ interface and MassLynx™ software. Borosilicate capillaries were purchased from Micromass U.K. (Manchester, U.K.) or Protana (Odense, Denmark).

**Molecular mass determination**

For the molecular mass determination of the glutathione S-transferase proteins, samples dissolved in 50% aqueous acetonitrile with 0.2% formic acid were loaded into borosilicate capillaries. The capillary voltage was set to between 700 and 1200 V to obtain a stable spray and the cone voltage was optimised between 30 and 50 V. The flow of nitrogen desolvation gas was optimised for each sample and the source temperature was set to 80°C. Data were acquired with argon gas in the collision cell and the voltage on the collision cell was 4 V. The integration time used was 2.4 s with a delay of 0.1 s. Spectra were combined, baseline-subtracted and smoothed with the MassLynx™ software. For calculation of the molecular mass, the spectra were centred (at 80% peak height). MassLynx™ was used to calculate the average molecular mass and the mass error (using ‘Component Find’).

**Noncovalent complex analysis:**

Prior to analysis of complexes, the QTof™ (or QTof™2) was set up for high m/z detection as follows: i) the focus voltage was set to around 150 V to assist in acceleration of the ions; ii) the pusher time was set to 180 μsec\(^3\) (rather than automatic mode); iii) the RF voltage on the gas cell and hexapole ion bridge were set to their

\[^3\text{The equation } (m/z)^{1/2} = 0.595 \times T \text{ was used to calculate an appropriate pusher time. The time was set to a value slightly higher than the calculated time for the desired upper m/z limit.}\]
maximum (by setting the transfer and collision offsets to 1); iv) the voltage on the collision cell was set to 10 V; v) the source temperature was set to 50°C.

The optimal conditions for the detection of complexes of glutathione S-transferase were set using a solution of caesium iodide (CsI) at 2 mg/mL with a cone voltage of between 90 and 110 V. In the peak window of the MassLynx™ tune page two m/z windows were viewed, a low m/z range around the 133 peak of CsI and a higher mass range from approximately m/z 2900 to 3100. Once a CsI peak in the low mass range was detected, the rotary pump in-line isolation valve was partially closed to increase the pressure between the cone and extraction lens whilst monitoring the peak window. After approximately 2.5 turns of the in-line isolation valve, higher mass peaks appeared and the low mass peak decreased in intensity. The valve was then fine-tuned for transfer of the high mass peaks.

A solution of CsI (2 mg/mL) was used for calibration of the instrument after optimisation of the conditions. A multipoint, linear calibration was applied to the m/z range of approximately m/z 2500 to m/z 4000. Samples were loaded into borosilicate capillaries and the capillary voltage was set to between 700 and 1200 V to obtain a spray, often assisted with a small amount of nitrogen back-pressure. The cone voltage was optimised between 70 and 90 V, depending on the pressure and capillary voltage (see section 5.3). For experiments investigating the effects of the cone voltage and collision cell voltage, these two parameters were adjusted accordingly (see section 5.2).

Spectra were combined, baseline-subtracted and smoothed using the MassLynx™ software (Micromass U.K., Manchester, U.K.). In some cases, MaxEnt™ was used to transfer the data from an m/z scale to mass scale.
CHAPTER THREE: AN EXAMPLE OF CONVENTIONAL PROTEOMICS

IDENTIFICATION OF PROTEINS RELEASED FROM MITOCHONDRIA DURING PERMEABILITY TRANSITION

3.1 INTRODUCTION

Apoptosis is a highly regulated form of cell death that is characterised by a number of morphological changes including cell shrinkage, nuclear condensation, DNA cleavage, cell surface exposure of phosphatidylserine and blebbing of the cell membrane into apoptotic bodies (Bossy-Wetzel et al., 1998; Mignotte and Vayssiere, 1998; Thornberry and Lazebnik, 1998). Apoptosis is critical for embryologic development, T cell selection in the thymus, and normal tissue homeostasis (Narula et al., 1997; Wang and Lenardo, 1997; Tsujimoto, 1997). Diseases such as HIV, Alzheimer’s disease, Parkinson’s disease, cancers, autoimmune diseases, stroke and rheumatoid arthritis have been associated either with insufficient activation or with over-activation of apoptotic mechanisms (Wilson, 1998; Ashkenazi and Dixit, 1998; Gulbins et al., 1996).

The process of apoptosis can be divided into three phases: initiation, execution and degradation (Figure 3.1) (Kroemer et al., 1997; Wilson, 1998; Susin et al., 1999a). In the initiation phase, cells receive the stimulus that activates the apoptotic pathway. Many different changes can act as apoptotic stimuli including heat, irradiation, depletion of growth factors, cross-linking of death receptors (such as TNFR1 or Fas), or treatment with pharmacological agents (Ashkenazi and Dixit, 1998; Kroemer et al., 1997; Tsujimoto, 1997; Zamzami et al., 1996). In contrast to the heterogeneity of the
initiation phase, the execution phase shows a regular cellular response with characteristic morphological changes such as the cell shrinkage and membrane blebbing mentioned above. The degradation phase involves the removal of any remainder of the cell, including phagocytosis of the apoptotic bodies (Mignotte and Vayssiere, 1998).

Figure 3.1 Stages of apoptosis
The mitochondria (i.e. the energy-producing organelles of cells) are known to be key regulators of the execution phase of apoptotic cell death and there are a number of mechanisms that may lead to cell death that involve mitochondria. These mechanisms include disruption of the energy production pathway (electron transport, oxidative phosphorylation and ATP production) and Ca\(^{2+}\) homeostasis, release of apoptogenic factors from the mitochondria, alteration of the redox potential and an increase in the production of reactive oxygen species (ROS) (Cai and Jones, 1999; Green and Reed, 1998; Scarlett and Murphy, 1997; Loeffler and Kroemer, 2000; Bernardi et al., 2001).

### 3.1.1 The role of mitochondria in apoptosis

Mitochondria possess a proton electrochemical gradient across the inner mitochondrial membrane as shown in Figure 3.2 (Petit et al., 1990). The uneven distribution of protons gives rise to a membrane potential (\(\Delta \psi_{\text{mit}}\)) and a chemical (pH) gradient both of which are crucial for normal mitochondrial function (Petit et al., 1990; Kroemer et al., 1997). Disruption of the \(\Delta \psi_{\text{mit}}\) often precedes apoptosis and occurs before the exposure of phosphatidylserine or degradation of DNA (Zamzami et al., 1996; Kroemer et al., 1997; Zamzami et al., 1997; Mignotte and Vayssiere, 1998). Experiments have indicated that the loss of \(\Delta \psi_{\text{mit}}\) results from a process known as permeability transition (PT) (Zamzami et al., 1997; Kroemer et al., 1997).
Permeability Transition

Permeability transition (PT) is the sudden increase in the permeability of the mitochondrial membrane to solutes ≤ 1500 Da (Kroemer et al., 1997). It is thought to result from the formation of pores in the inner membrane at contact sites between the inner and outer mitochondrial membranes (Green and Reed, 1998; Zamzami et al., 1997; Scarlett and Murphy, 1997). The opening of these pores allows equilibration of ions between the mitochondrial matrix and the intermembrane space and therefore dissipates the proton electrochemical gradient (Green and Reed, 1998). The pores also allow the influx of water (owing to the osmolality of the matrix) resulting in enlargement of the mitochondria (Green and Reed, 1998). The folded inner membrane has the capacity to expand but the outer membrane does not possess the large surface area of the inner membrane (Figure 3.2). The outer membrane eventually ruptures, releasing the contents of the intermembrane space (Scarlett and Murphy, 1997; Green and Reed, 1998; Vander Heiden et al., 1997). Despite the loss in integrity of the outer
membrane, the inner membrane may continue oxidative phosphorylation and the production of ATP, which may be necessary for ordered destruction of the cell (Vander Heiden et al., 1997). The order of these changes in the mitochondria may be dependent on the cell type and the stimulus that resulted in initiation of apoptosis (Susin et al., 1999a).

The PT pore can operate in either a low-conductance state (LC-PT), in which the pore opens and closes rapidly, or an irreversible high-conductance state (HC-PT) in which the pore remains open (Bossy-Wetzel et al., 1998; Green and Reed, 1998; Pinton et al., 2001). HC-PT results in a long-term loss of $\Delta \psi_{\text{mit}}$ whereas LC-PT results in a transient loss of $\Delta \psi_{\text{mit}}$. In the LC state, only small molecules and ions are able to pass through the pore, whilst in the HC state larger molecules pass through and subsequently the outer mitochondrial membrane may be disrupted (Pinton et al., 2001).

**The PT pore**

The PT pore is believed to be composed of the adenine nucleotide translocator (ANT, an inner membrane protein), the voltage-dependent anion channel (VDAC, which is an outer membrane protein also known as porin or the peripheral benzodiazepine receptor (PBA)), creatine kinase (an intermembrane protein), cyclophilin D (a mitochondrial matrix protein) and hexokinase (a cytosolic protein) (Zamzami et al., 1997; Kroemer et al., 1997; Shimizu et al., 1999; Gottlieb, 2000). Furthermore the ANT has been shown to interact with Bcl-2 and Bax (Shimizu et al., 1999) (see section 3.1.2) and to possess the ability to form non-specific pores in response to several inducers of apoptosis (Loeffler and Kroemer, 2000). These non-specific pores may lead to permeabilisation of the inner membrane, causing swelling and a resulting disruption of the outer membrane (Loeffler and Kroemer, 2000).
In some examples, certain inhibitors of components of the PT pore did not prevent permeabilisation of the mitochondrial membrane (Loeffler and Kroemer, 2000). It is, therefore, possible that a number of mechanisms operate to cause PT.

**AIF**

Apoptosis-inducing factor (AIF), a flavoprotein of approximately 57 kDa, is released from the mitochondrial intermembrane space during PT (Susin *et al.*, 1996; Susin *et al.*, 1999b). AIF is able to induce chromatin condensation and fragmentation in isolated nuclei (Susin *et al.*, 1999b). AIF also affects the function of mitochondrial membranes, so that AIF released from mitochondria may act on other mitochondria (Susin *et al.*, 1999b). Bcl-2, which is associated with the outer mitochondrial membrane, inhibits PT and the release of AIF from mitochondria, although the mechanism by which this occurs remains unknown (Mignotte and Vayssiere, 1998).

Other proteins that are released from the mitochondria of cells undergoing apoptosis include cytochrome C, procaspase-2, procaspase-9 and a DNAse (Susin *et al.*, 1999a). The release of these proteins is also inhibited by Bcl-2 (Susin *et al.*, 1999a).

### 3.1.2 The Bcl-2 Family

The Bcl-2 family consists both of inhibitors and of activators of apoptosis. Bcl-2 and Bcl-X<sub>L</sub> are inhibitors of apoptosis, whilst Bad, Bax, Bid and Bim promote apoptosis (Du *et al.*, 2000). The mechanisms by which the Bcl-2 family regulate apoptosis are not fully understood. It has been shown that competition between Bcl-2 and Bax exists but each of these proteins is able to independently regulate apoptosis (Mignotte and Vayssiere, 1998).

Bcl-2 and Bcl-X<sub>L</sub> are located primarily in the outer mitochondrial membrane, nuclear envelope and endoplasmic reticulum (Mignotte and Vayssiere, 1998; Kroemer
et al., 1997). Bcl-2 has been shown to prevent the release of cytochrome C from mitochondria but does not prevent apoptosis induced by the addition of exogenous cytochrome C (Kluck et al., 1997). In a similar fashion, Bcl-2 prevents the release of AIF from mitochondria but does not protect the cell from AIF present in the cytosol (Susin et al., 1996; Susin et al., 1999b). Bcl-2 has also been shown to protect mitochondria from a loss of Δψ_{mit} and induction of PT (Susin et al., 1996; Kluck et al., 1997). Bcl-X\textsubscript{L} prevents membrane hyperpolarisation and mitochondrial swelling apparent prior to cytochrome C release and the loss of Δψ_{mit} (Vander Heiden et al., 1997).

The Bcl-2 family proteins are structurally similar to some pore-forming domains of bacterial toxins (Vander Heiden et al., 1997; Kluck et al., 1997; Bossy-Wetzel et al., 1998; Green and Reed, 1998). Furthermore, Bcl-2, Bcl-X\textsubscript{L} and Bax have been shown to form channels in artificial bilayers (Vander Heiden et al., 1997; Green and Reed, 1998; Bossy-Wetzel et al., 1998) and Bax is thought to be part of the PT pore (Antonsson and Martinou, 2000). Hence, the Bcl-2 family may regulate the flow of small ions (and maybe larger molecules) across mitochondrial membranes (Vander Heiden et al., 1997; Bossy-Wetzel et al., 1998; Green and Reed, 1998). Experiments have also indicated that the Bcl-2 family may regulate the conformation of VDAC in the PT pore (Shimizu et al., 1999).

### 3.1.3 Cytochrome C

Cytochrome C resides in the intermembrane space of mitochondria and is an electron carrier in oxidative phosphorylation (Liu et al., 1996; Bossy-Wetzel et al., 1998). Cytochrome C is released from mitochondria in cells undergoing apoptosis and is able to activate downstream caspases and DNA fragmentation (Liu et al., 1996; Bossy-
Wetzel et al., 1998; Kluck et al., 1997). Cytochrome C binds to Apaf-1 (apoptosis protease activating factor-1), dATP (deoxyadenosine triphosphate) and procaspase-9, resulting in the autoactivation of caspase-9, which in turn activates downstream caspases (caspases 3, 6 and 7) (Du et al., 2000).

The release of cytochrome C has been shown to be an early event in the execution phase of apoptosis and can occur prior to the loss of $\Delta \psi_{\text{mit}}$ and independently of caspase activation (Vander Heiden et al., 1997; Bossy-Wetzel et al., 1998). Inducers of PT have been shown to cause release of cytochrome C from isolated mitochondria (Bossy-Wetzel et al., 1998) but a decrease in the $\Delta \psi_{\text{mit}}$ of isolated mitochondria does not necessarily result in the release of cytochrome C (Kluck et al., 1997).

The mechanism(s) responsible for the release of cytochrome C from the mitochondrial intermembrane space are unknown (Bossy-Wetzel et al., 1998). It is possible that the release of cytochrome C occurs through the PT pore and does not require HC-PT (and is therefore apparent prior to a sustained loss of $\Delta \psi_{\text{mit}}$) (Scarlett and Murphy, 1997; Bossy-Wetzel et al., 1998). Another possibility is that upon disruption of the outer membrane in a small subpopulation of mitochondria, a general disruption of $\Delta \psi_{\text{mit}}$ might not be apparent (Mignotte and Vayssiere, 1998; Scarlett and Murphy, 1997). The Bcl-2 family members may also act as pore-forming proteins and regulate the release of cytochrome C (Mignotte and Vayssiere, 1998; Bossy-Wetzel et al., 1998; Gottlieb, 2000). It has been suggested that a channel formed by Bax and VDAC may allow the passage of cytochrome C out of the mitochondria (Gottlieb, 2000; Antonsson and Martinou, 2000). It is also of interest that the import of newly translated, apocytochrome C into the mitochondria does not require an import receptor, proteolytic processing, high $\Delta \psi_{\text{mit}}$ or ATP (Liu et al., 1996; Bossy-Wetzel et al., 1998; Mignotte
and Vayssiere, 1998). It is, therefore, possible that the import of cytochrome C is related to its export.

The release of cytochrome C may have different consequences on different cells (Green and Reed, 1998). Those cells containing high levels of cytochrome C may be able to continue ATP production and the induction of apoptosis therefore occurs via caspase activation. In cells with lower levels of cytochrome C and/or high levels of caspase inhibitors, apoptosis may occur because of the depletion of ATP and production of ROS.

3.1.4 Caspases

Caspases are cysteine proteases that cleave substrates on the C-terminal side of aspartic acid residues (Du et al., 2000). These proteases require recognition of specific tetrapeptide sequences on the N-terminal side of the aspartic acid (Thornberry and Lazebnik, 1998). The activation of caspases is one of the key steps in the execution of apoptosis (Bratton et al., 2000). They are synthesised as inactive enzymes and, in cells undergoing apoptosis, are converted to the active form by proteolytic cleavage (Thornberry and Lazebnik, 1998; Bossy-Wetzel et al., 1998). Activated caspases then activate other caspases ("caspase cascade") and cleave many intracellular substrates that would otherwise protect the cell from apoptosis (Thornberry and Lazebnik, 1998; Du et al., 2000).

Activation of individual caspases can occur by a number of mechanisms and cells that have been induced to undergo apoptosis through different pathways may activate distinct caspase cascades (Thornberry and Lazebnik, 1998; Bratton et al., 2000). Activation of the initial caspase appears to involve cofactors. For example, procaspase-9 activates itself (autoactivation) after binding to a complex of cytochrome C, Apaf-1 and
dATP (see also section 3.1.3) (Green and Reed, 1998). Caspase-9 and caspase-3 are released from mitochondria as proenzymes and become activated during, or immediately after, their release (Susin et al., 1999a).

Another group of proteins important in the regulation of apoptosis are the "inhibitors of apoptosis proteins" (IAPs) (Thornberry and Lazebnik, 1998). IAPs are able to bind to procaspase-9 and prevent its activation (Du et al., 2000). IAPs also bind and inhibit other activated caspases (Thornberry and Lazebnik, 1998; Du et al., 2000). The inhibition of IAPs is opposed by another protein called Smac (second mitochondria-derived activator of caspase) (Du et al., 2000). Smac is a mitochondrial protein that is released to the cytosol during apoptosis and promotes the activation of procaspase-9 by bindings IAPs (Du et al., 2000; Gottlieb, 2000).

Activated caspase-3 activates DFF/CAD (DNA fragmentation factor, which appears to be a complex of caspase-activated deoxyribonuclease and its inhibitor (Wilson, 1998)), which triggers DNA fragmentation (Susin et al., 1999b). Capase-3 also cleaves polyADP-ribose polymerase (PARP, a DNA-repairing enzyme), protein kinase C δ, DNA-dependent protein kinase and other proteins (Bossy-Wetzel et al., 1998). DNA repair proteins, DNA replication enzymes and factors important for mRNA splicing are cleaved by caspases (Thornberry and Lazebnik, 1998). Bcl-2-protective proteins are also inactivated by caspase-catalysed cleavage, which appears to result in a pro-apoptotic fragment (Thornberry and Lazebnik, 1998). The pro-apoptotic protein Bid is also activated by caspases (Pinton et al., 2001). Cellular structure is affected by caspase cleavage of structural proteins such as lamins (causing destruction of nuclear lamina) and actin, as well as focal adhesion kinase (FAK), gelsolin and p21-activated kinase-2, which are proteins that regulate the cytoskeletal components (Nicholson and Thornberry, 1997; Thornberry and Lazebnik, 1998).
Caspases are also capable of inducing PT pore opening (Green and Reed, 1998). In cells that show release of cytochrome C prior to the loss of $\Delta \psi_{\text{mit}}$, activation of the caspase cascade by cytochrome C may result in PT (Green and Reed, 1998). This would cause further caspase activation through an increase in cytosolic cytochrome C and AIF.

Caspases have an enormous capacity for destruction of cellular activities. Cell-to-cell contact, cytoskeleton and nuclear structure organization, DNA replication and repair and splicing events are all disrupted by caspase activation (Thornberry and Lazebnik, 1998). Many of these events are poorly understood and there are likely to be caspase effects that remain unknown.

An important role exists for many other components in the apoptotic pathway including $\text{Ca}^{2+}$, serine/threonine protein kinases (such as PKC, MAPK and PKB), ceramide and calpains (Pinton et al., 2001). Although significant advances in the past decade have greatly improved our understanding of the components and pathways involved in apoptosis, the complicated and diverse mechanisms involved in the execution phase of apoptosis are not clearly understood. For application of inducers or inhibitors of apoptosis to disease therapies, an improved understanding of the underlying mechanisms is necessary.

### 3.1.5 Mitochondria and proteomics

The complexity of mitochondria, their importance in energy production and the central role they play in the execution phase of apoptosis indicate that establishment of the proteome of normal functioning mitochondria would be useful. Mitochondria have recently evoked interest from a proteomics perspective and the development of a mitochondrial proteome has commenced (Rabilloud et al., 1998). Advances in the approaches used to establish the mitochondrial proteome have included a sucrose
density gradient separation of the mitochondrial proteins prior to 2-dimensional electrophoresis (Hanson et al., 2001) and affinity fractionation using mini-spin columns to retain calcium-binding proteins, glycoproteins or hydrophobic proteins (Lopez et al., 2000b). An analysis of the proteins released from mitochondria following HC-PT has also been reported, in which 79 known proteins were identified (Patterson et al., 2000). Proteomics techniques have significant potential for identifying proteins released from the mitochondria during apoptotic processes and changes that occur within the compartments of the mitochondria. Comparisons of mitochondrial fractions from normally functioning mitochondria and mitochondria undergoing PT may provide further insight into proteins that are implicated in apoptosis. Combining the affinity fractionation reported from Lopez et al. (Lopez et al., 2000b) with control and PT samples may reveal interesting changes in the proteins retained. Currently there have been no reports of changes that accompany LC-PT being studied by proteomic techniques.

3.1.6 Project aims

Prior to the commencement of this project, 2-dimensional electrophoresis had not been used in our laboratory. The initial part of this work therefore involved setting up equipment and establishing protocols for 2-dimensional electrophoresis. Application of the technique to mitochondrial samples was then planned.

The development of a cell-free system for the analysis of PT in mitochondria was undertaken by Ms Alison Smail and Dr Mark Wilson (Department of Biological Sciences, University of Wollongong). The aim of this project was to develop a method for the proteomic analysis of the mitochondria and, in particular, to assess the changes in the protein complement released from mitochondria that had undergone LC-PT.
3.2 RESULTS

3.2.1 Overview of method

Mitochondria were prepared in collaboration with Dr Mark Wilson and Ms Alison Smail (Biological Sciences, University of Wollongong). The mitochondria and supernatant from the mitochondria were stored at -80°C until required.

Figure 3.3 shows an outline of the method for two-dimensional electrophoresis of the mitochondrial samples. Mitochondria were induced to undergo permeability transition (PT) with protoporphyrin IX (PPIX) (Alison Smail, Biological Sciences, University of Wollongong). The intact mitochondria (a sample that was treated with PPIX and a control sample) were washed with isotonic media to collect the proteins that were released during PT and the mitochondria were pelleted. Three extractions were performed on the mitochondria. The first extracted water-soluble components using a solution of 40 mM Tris buffer. The second and third extractions removed hydrophobic proteins from the mitochondria. The second extraction used urea and the detergent CHAPS whilst the third extraction buffer contained urea and CHAPS as well as thiourea and SB3-10 detergent.

The second and third extractions from mitochondria were applied directly to the IPG gels by rehydrating the gels in the extraction solutions. The mitochondrial wash and water-soluble extraction required further preparation before they were applied to the IPG gels. Some early preparations contained KCl, which was removed prior to IEF using ultrafiltration devices. The samples were then lyophilised and redissolved in 2-dimensional electrophoresis buffer (as used for the second extraction solution). The IPG gels were then rehydrated in these solutions.

After performing 2DE, the gels were fixed and the proteins visualised by silver staining. It was found that staining with Coomassie Brilliant Blue was not sufficiently
Chapter 3: Conventional Proteomics

Figure 3.3 Overview of method – Preparation of samples for MS
sensitive to detect the low concentrations of proteins loaded on the gels. The gels from
the control mitochondrial samples and the samples from mitochondria that had been
induced to undergo PT were then compared to detect protein differences. Protein spots
of interest were excised from the gels and the proteins in the spot were enzymatically
cleaved with trypsin.

Extracted peptides from in-gel digests were desalted using C18ZipTips™ and then
analysed by MALDI-MS and/or nanoESI-MS. The information obtained from MS was
used to interrogate databases of protein and translated oligonucleotide sequences using
the programs from Protein Prospector (http://prospector.ucsf.edu) (Clauser et al., 1999)
and/or Peptide Search (also known as PepSea) (http://pepsea.protana.com) (Mann and
Wilm, 1994).

3.2.2 Extraction of proteins from mitochondria

Figure 3.4 shows the gels obtained from a sequential extraction of proteins from
mitochondria that had not been treated with PPIX. Figure 3.4A shows the gel from the
extraction of water-soluble proteins. The dark staining in the high mass region suggests
that the gel was overloaded and therefore the proteins in this region could not be
resolved. The lower mass region shows many resolved proteins. The second gel (Figure
3.4B) is from the extraction performed on the pelleted membranes after the first
extraction. The protein concentration appeared to be lower on this gel but the
background in the high mass region was still quite dark. Some of the spots in the first
gel were also apparent in the gel from the second extraction but there was also evidence
of proteins that were not apparent in the first extraction. The gel pictured in Figure 3.4C
was from the final extraction of the same mitochondria, which had been pelleted after
the second extraction. The gel from this extraction did not appear to be overloaded but
Figure 3.4 Sequential extraction of proteins from mitochondria A) First extraction with 40mM Tris, water soluble proteins, B) Initial extraction from membranes of mitochondria, C) Final extraction from membranes of mitochondria
there were fewer spots in the low mass region of the gel. Again there were proteins that were evident in the two previous gels. The majority of detected protein spots were in the basic pI region of the gel. The three gels show the decreasing concentration of proteins through the extractions. The lack of proteins in the low mass and acidic regions of the gel from the third extraction suggests that although this gel did not reveal the dark background indicative of overloading, some proteins were not detected as a result of their low concentration.

3.2.3 Gels from proteins released from mitochondria during LC-PT

Figure 3.5 shows two gels obtained from 2DE of the supernatant from control mitochondria. These gels were from the same preparation and were obtained under identical conditions. Small differences in the intensity of the silver stain were observed in the gels, even though they had been treated using the same experimental conditions. The large number of spots on the gels in Figure 3.5 indicates that many proteins were soluble in the mitochondrial wash, either from mitochondrial surface proteins or proteins that were released from the mitochondria even before the induction of PT.

The gels in Figure 3.6 are from the supernatant of mitochondria from a different mouse liver preparation. The gel from the supernatant of mitochondria that had not undergone PT (control gel, Figure 3.6A) appears to be significantly different from the two control gels in Figure 3.5, even though the mitochondria were prepared in the same way. Fewer protein spots were apparent on the gel and an intense series of spots was apparent in the basic-pI/high-mass region. The gels from the supernatant of mitochondria that had undergone low conductance (LC) PT (Figure 3.6 B-D) also showed this intense series of proteins. There were, however, a number of different spots in the LC-PT gels that were not apparent in the control gel.
Figure 3.5 Silver stained gels from 2-dimensional electrophoresis of the supernatant from a control mitochondrial preparation. Gels obtained under identical conditions.

Figure 3.6 Silver stained gels from 2-dimensional electrophoresis of the supernatant from mitochondrial preparations. A) Control mitochondria, B) C) and D) Mitochondria treated with 20μM PPIX.
The gels from another preparation of control mitochondria and LC-PT mitochondria are shown in Figure 3.7. The control gel from this preparation (Figure 3.7A) appears to be similar to the gels in Figure 3.5 but showed some differences to the control gel in Figure 3.6A. The intense series of spots seen in Figure 3.6 was not apparent in the gels in Figure 3.7. Of the gels in Figure 3.7, the intensity of the stain in the control gel (Figure 3.7A) appears to be greater than in the LC-PT gel (Figure 3.7B). The ability to distinguish novel spots in the LC-PT gel is therefore made more difficult by uncertainty in the intensity of the silver stain and the vast difference between the gels in Figure 3.6 and 3.7.

3.2.4 Examples of the identification of proteins from 2-DE gels

Figure 3.8 shows the MS/MS spectrum obtained from a doubly charged ion of m/z 705.9 that was apparent in the nanoESI mass spectrum of an in-gel digest of spot A from Figure 3.5. The digest was performed on four equivalent combined spots in order to maximise the sample amount. The MS/MS spectrum was used to obtain a sequence tag for searching the available databases. The ions at m/z 817.4, 916.5, 1013.6, 1084.7 and 1183.6 resulted in the sequence tag (817.4)VPAV(1183.6). This was entered into the program Peptide Search with the peptide molecular mass, 1409.78 Da. The search resulted in the identification of the rat mitochondrial matrix protein, electron transfer flavoprotein alpha-subunit (ETFα) (Swiss Prot accession number P13803, the equivalent mouse protein was not present in the database). The remainder of the product ions in the MS/MS spectrum matched the retrieved sequence (LNVAPVSDIIEIK). Along with B ion and Y ion series, internal fragments from the preferential cleavage on the N-terminal side of proline were apparent in the spectrum (this is discussed further in
Figure 3.7 Silver stained gels from 2-dimensional electrophoresis of the supernatant from mitochondrial preparations. Approximate molecular masses are shown on the right-hand side of the figure. A) Control mitochondria, B) Mitochondria treated with 20μM PPIX.
Figure 3.8 MS/MS spectrum of doubly charged precursor ion at m/z 705.9 from an in-gel digest of spot A shown in Figure 3.5. Ions from the identified sequence, LNVAPVSDIEIK, are shown.

Figure 3.9 MS/MS spectrum of triply charged precursor ion at m/z 854.1 from an in-gel digest of spot B shown in Figure 3.5. Ions from the identified sequence, LVQDVANNTNEEAGDGTATVLAR, are shown.
Chapter 3: Conventional Proteomics

A number of peptides from ETFα were sequenced by MS/MS from this particular digest (data not shown).

Figure 3.9 shows an MS/MS spectrum from an in-gel digest of equivalent spots combined from two gels (spot B in Figure 3.5). MS/MS of the triply charged precursor at m/z 854.1 resulted in protein identification through sequence tag searching with the tag (933.6)GDGA(1233.6). The retrieved sequence from the mitochondrial precursor of heat shock protein 60 (accession number P19226) closely matched the remaining MS/MS data and the Y and B ions are labelled in Figure 3.12. A number of other peptides from the same protein were also identified by MS/MS, which further confirmed the identity of spot B in Figure 3.5.

An example of a MALDI mass spectrum of an in-gel digest is shown in Figure 3.10. This spectrum was obtained from the tryptic digest of the same spot combined from three separate gels of mitochondria that had undergone LC-PT. The spot chosen (spot A, Figure 3.6D) corresponded to a spot from the intense series of spots mentioned above that was not apparent in the gels in Figures 3.5 and 3.7. The list of ions with m/z over 1000 Da, which were used for a mass fingerprint search, are shown in Figure 3.11. Using Protein Prospector to search human and mouse entries in the Swiss Prot database with these masses resulted in 93 matches, the top 10 of which are shown in Figure 3.12. The entry with the highest rank, with 20 out of 41 peptides matched, was mouse catalase (accession number 24270). The identity of this protein was confirmed by nanoESI-MS/MS and database searching with sequence tags from the MS/MS data (data not shown). Also noted in the mass fingerprint search results was a human keratin with 7 matched peptides, which was apparent in a number of samples.
Figure 3.10 MALDI mass spectrum of an in-gel digest of spot A from Figure 3.6

Figure 3.11 Peak list of ions with m/z > 1000 Da in Figure 3.10 entered into MS-Fit for peptide mass fingerprint search of Swiss Prot database.
MS-Fit Search Results

Sample ID (comment): spot A
Database searched: SwissProt.01September2001
Molecular weight search (1000 - 100000 Da) selects 94640 entries.
Full pI range: 100228 entries.
Species search (HUMAN MOUSE) selects 11762 entries.
Combined molecular weight, pI and species searches select 10474 entries.
MS-Fit search selects 93 entries (results displayed for top 10 matches).

Parameters Used in Search

- Considered modifications: | Peptide N-terminal Gin to pyroGlu | Oxidation of M | Protein N-terminus Acetylated |
- Min. # Peptides to Match | Peptide Mass Tolerance (+/-) | Peptide Masses are | Digest Used | Max. # Missed Cleavages | Cysteines Modified by | Peptide N terminus | Peptide C terminus |
- 5 | 0.300 Da | monoisotopic | Trypsin | 1 | acrylamide | Hydrogen (H) | Free Acid (O H) |

Result Summary

<table>
<thead>
<tr>
<th>Rank</th>
<th>MOWSE Score</th>
<th># (%): Masses Matched</th>
<th>Protein MW (Da)/pI</th>
<th>Species</th>
<th>Accession #</th>
<th>Protein Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.88e+011</td>
<td>20/41 (48%)</td>
<td>59765.8/7.72</td>
<td>MOUSE</td>
<td>P24270</td>
<td>CATALASE</td>
</tr>
<tr>
<td>2</td>
<td>1.48e+006</td>
<td>10/41 (24%)</td>
<td>59756.7/6.90</td>
<td>HUMAN</td>
<td>P04040</td>
<td>CATALASE</td>
</tr>
<tr>
<td>3</td>
<td>3.24e+003</td>
<td>8/41 (19%)</td>
<td>69872.9/7.79</td>
<td>HUMAN</td>
<td>P43403</td>
<td>TYROSINE-PROTEIN KINASE ZAP-70 (70 KDA ZETA-ASSOCIATED PROTEIN) (SYK-RELATED TYROSINE KINASE)</td>
</tr>
<tr>
<td>4</td>
<td>2.33e+003</td>
<td>7/41 (17%)</td>
<td>95202.8/8.94</td>
<td>MOUSE</td>
<td>Q60934</td>
<td>GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 1 PRECURSOR (GLUTAMATE RECEPTOR 5) (GLUR-5)</td>
</tr>
<tr>
<td>5</td>
<td>1.6e+003</td>
<td>7/41 (17%)</td>
<td>65865.8/8.07</td>
<td>HUMAN</td>
<td>P35908</td>
<td>KERATIN, TYPE II CYTOSKELETAL 2 EPIDERMAL (CYTOKERATIN 2E) (K2E) (CK 2E)</td>
</tr>
<tr>
<td>6</td>
<td>1.22e+003</td>
<td>6/41 (14%)</td>
<td>81278.4/6.10</td>
<td>HUMAN</td>
<td>Q92503</td>
<td>SEC14-LIKE PROTEIN 1</td>
</tr>
<tr>
<td>7</td>
<td>821</td>
<td>5/41 (12%)</td>
<td>47773.9/8.37</td>
<td>MOUSE</td>
<td>Q62226</td>
<td>SONIC HEDGEHOG PROTEIN PRECURSOR (SHH) (HHG-1)</td>
</tr>
<tr>
<td>8</td>
<td>601</td>
<td>5/41 (12%)</td>
<td>97657.5/5.53</td>
<td>HUMAN</td>
<td>Q13435</td>
<td>SPlicing FACTOR 3B PROTEIN, SUBUNIT 2 (SF3B150) (SPlicingOSOME ASSOCIATED PROTEIN 145) (SAP 145)</td>
</tr>
<tr>
<td>9</td>
<td>476</td>
<td>5/41 (12%)</td>
<td>37206.7/6.67</td>
<td>HUMAN</td>
<td>Q95154</td>
<td>AFLATOXIN B1 ALDEHYDE REDUCTASE 2 (AFB1-AR 2)</td>
</tr>
<tr>
<td>10</td>
<td>455</td>
<td>5/41 (12%)</td>
<td>73095.1/6.96</td>
<td>HUMAN</td>
<td>Q15523</td>
<td>DEAD-BOX PROTEIN 3, Y-CHROMOSOMAL</td>
</tr>
</tbody>
</table>

Figure 3.12 Results from peptide fingerprint search from MS-Fit (Protein Prospector) using the m/z ratios of ions in Figure 3.10.
Another example of an identification obtained from the gels in Figure 3.6 was the protein acetyl-CoA C-acyl transferase. This protein (spot B, Figure 3.6D, combined from three gels) was identified through MS/MS and sequence tag searching (data not shown). Acetyl-CoA C-acyl transferase is also known as 3-ketoacyl-CoA thiolase and the rat sequences retrieved from the Swiss Prot database were P21775 and P07871 (no equivalent mouse entries were found in the Swiss Prot database). Catalase and acetyl-coA C-acyl transferase are both peroxisomal proteins.

3.2.5 LC-MS of digests from gel-separated proteins

Capillary LC-MS was performed on an in-digest of spot C shown in Figure 3.5. The digest was performed on identical spots combined from four gels. The total ion current (TIC) chromatogram (Figure 3.13F) provided minimal information in regards to the elution of peptides from the column. The base peak intensity (BPI) chromatogram was more informative for the elution of components from the column (Figure 3.13E). The BPI chromatogram and individual mass chromatograms showed that the peptides eluted between approximately 33 minutes and 41 minutes. Figure 3.13A-D shows four individual mass chromatograms and the corresponding summed spectra from the peaks in these chromatograms are shown in Figure 3.14. The first multiply charged component that was detected by MS from the capillary column was a doubly charged ion at m/z 566.3 at a retention time of 32.8 minutes. The portion of the spectrum illustrating the isotopic resolution of this ion is in Figure 3.14A. The abundant ions that were also apparent in the spectrum were from singly charged ions (m/z 571.3, 672.4, 743.4 and 782.4). The mass chromatogram from the doubly charged ion at m/z 639.8 (Figure 3.13B) showed a higher signal-to-noise ratio than the other three mass chromatograms in Figure 3.13. The summed spectra in Figure 3.14 also reflected the high signal-to-
Figure 3.13 Chromatograms from LCMS of an in-gel digest of spot C shown in Figure 3.5. A) Mass chromatogram of m/z 566.3, B) Mass chromatogram of m/z 639.8 C) Mass chromatogram of m/z 581.2, D) Mass chromatogram of m/z 899.9, E) Base peak intensity chromatogram, F) Total ion current chromatogram.
Figure 3.14 Summed spectra from mass chromatograms shown in Figure 3.13. Average retention times A) 32.8 min B) 34.5 min C) 37.5 min D) 41.1 min.
noise of the ion at m/z 639.8. Ions of lower abundance were also apparent in this spectrum but the ion at m/z 639.8 dominated the spectrum. In contrast, the summed spectra from the other chromatograms did not show the most abundant ion to be that with which the chromatograms were generated. The doubly charged ion at m/z 581.2 eluted at a retention time of 37.5 min along with a number of other multiply charged ions, including the doubly charged ion at m/z 583.3, which can also be seen in the enlarged section of Figure 3.14C. The most abundant ion in this spectrum was from an ion at m/z 652.4. The summed spectra from the peak in the chromatogram of m/z 899.9 showed the ion at m/z 563.5 to be the most abundant. The ion at m/z 899.9 did, however, have a high signal-to-noise ratio and was clearly doubly charged, as can be seen in the enlarged section of the spectrum in Figure 3.14D.

Separation of the digest by LC-MS revealed many ions that were not apparent in the nanoESI mass spectrum of the same sample (Figure 3.15A). The most abundant ion in the nanoESI mass spectrum is from a singly charged ion at m/z 672.4. The high signal to noise ratio of the doubly charged ion at m/z 639.8, which was apparent in Figure 3.14B, is clear in Figure 3.15. The ions at m/z 581.2 and 583.3, which were apparent in the LC-MS experiment at a retention time of 37.5 min (Figure 3.14C) were, however, not visible above the background noise in the nanoESI mass spectrum (Figure 3.15A). The spectrum resulting from summation of the individual spectra from 32.5 minutes to 41.5 minutes of the LC-MS experiment (Figure 3.15B) was quite similar to the nanoESI mass spectrum of the same sample (Figure 3.15A). Both spectra showed the same singly charged ion at m/z 672.4, which was apparent at a retention time of 32.8 min in the LC-MS experiment (Figure 3.14A). The multiply charged components that were apparent in the nanoESI mass spectrum were also apparent in the summed LC-MS spectra but the nanoESI mass spectrum showed higher signal-to-noise ratios for the
Figure 3.15 ESI mass spectra from the in-gel digest of spot C in Figure 3.5 A) nanoES-MS B) Summed spectra from LC-MS experiment in Figure 3.13
majority of these components. It was clear from Figure 3.15B that the spectrum obtained by summing all the spectra in the LC-MS experiment resulted in lower signal-to-noise ratios than the individual spectra in the LC-MS experiment.

3.2.6 LC-MS/MS

LC-MS/MS experiments can be performed automatically during an LC-MS experiment when criteria for selection of a suitable precursor are met. Correct settings for switching from MS mode to MS/MS mode and then from MS/MS mode back to MS mode are crucial for obtaining useful data in this type of experiment (known as “function switching”). This is further discussed with other examples in Chapter 4.

Figure 3.16 shows three MS/MS spectra that were obtained from “function switching” during an LC-MS experiment. Sequence tags from these spectra resulted in the identification of three peptides from the mitochondrial matrix protein ornithine carbamoyltransferase (also known as ornithine transcarbamylase or OTcase, Swiss Prot accession number P11725). The Y ion series from the sequence LS(OxM)TNDPLEAAR (where OxM represents oxidised methionine) is shown in Figure 3.16A. The doubly charged ion at m/z 639.8 from the peptide with the sequence VLSS(OxM)TDVLARAR was selected for MS/MS. The resulting MS/MS spectrum (Figure 3.16B) showed a number of Y ions and another short series of ions, which were 64 Da below the corresponding Y ions (m/z 828.4, 915.5 and 1002.5). The loss of 64 Da is often observed from peptides containing oxidised methionine and corresponds to the loss of HSOCH$_3$ (Kinter and Sherman, 2000). The B$_2$ and A$_2$ ions were also abundant in Figure 3.16B. The MS/MS spectrum from the doubly charged ion at m/z 715.3 from the peptide LQAFQGYQVT(OxM)K is shown in Figure 3.16C. The Y ions in this spectrum resulted in peptide identification and are labelled along with the B$_2$ and B$_3$ ions. This
Figure 3.16 MS/MS spectra from an LC-MS separation of an in-gel digest of spot C in Figure 3.5. Precursors selected by data-dependent MS to MS/MS switching: A) m/z 667.3, 34.1 minutes, LS(OxM)TNDFEAAR; B) m/z 639.8, 34.9 minutes, VLSS(OxM)TDAVLAR; C) m/z 715.3, 35.8 minutes, LQAFQGYQVT(OxM)K.
spectrum also showed the presence of ions that were 64 Da below the Y₄", Y₉" and Y₁₀" ions.

The MS/MS spectra from function switching during an LC-MS experiment (Figure 3.16) differ from the spectra acquired from nanoES MS/MS (Figure 3.8 and 3.9) mainly because of the lower ion counts in the LC-MS/MS spectra. The time constraints when acquiring data during an LC-MS experiment often mean that only a few spectra are acquired on a single precursor ion. Therefore series of product ions are often not complete, as can be seen in Figure 3.16C where the Y₅" and Y₆" ions are not evident in the spectrum. Sufficient information for identification may not require extensive coverage of the possible product ions, however, and a small number of ions may result in assignment of the peptide sequence with high confidence. This is explained further in Chapter 4.
3.3 DISCUSSION

3.3.1 Identification of mitochondrial proteins

The identification of mitochondrial proteins from 2DE gels was achieved with both MALDI mass fingerprinting data and sequence tag data derived from ESI MS/MS. A number of proteins from the mitochondrial matrix were identified from the supernatant of mitochondria that had not undergone permeability transition (PT). These proteins included an electron acceptor for dehydrogenases (the electron transfer flavoprotein), an enzyme important for arginine biosynthesis (OTCase) and a heat shock protein (HSP60), which is implicated in mitochondrial import and the folding of imported proteins. There are a number of reasons that may account for the presence of these mitochondrial matrix proteins. It is possible that a loss of mitochondrial membrane integrity may have occurred in a fraction of the mitochondria prior to the addition of PPIX, resulting in release of matrix proteins to the supernatant. Alternatively, a mechanism for the controlled release of proteins from normal functioning mitochondria may be in operation. The appearance of the gels from the mitochondrial supernatant suggests that a general release of the contents of the matrix had not occurred, as limited spots were apparent. Further studies are required to establish the reason for the presence of matrix proteins in the supernatant.

The spots identified as mitochondrial matrix proteins mentioned above were also apparent in the gels from the supernatant of mitochondria that had undergone LC-PT. Hence, the presence of these matrix proteins was apparently not affected by LC-PT. It was difficult to compare relative intensities between gels so it could not be concluded as to whether the concentrations of these proteins had changed. The LC-PT gels therefore did not provide further insight into possible reasons for the presence of the mitochondrial matrix proteins. Furthermore, visually finding protein spots that appeared...
to be different in the control and LC-PT gels proved to be very difficult. The use of a fluorescent stain such as SYPRO Ruby (see also section 3.3.2) may improve the staining reproducibility and help to determine whether the concentration of proteins differed between the supernatant from control mitochondria and the supernatant from mitochondria that had undergone LC-PT. The process of quantifying spots on a gel is, however, not very accurate, even with computer-assisted methods. More accurate methods for protein quantification have been developed using isotope labelling of peptides prior to mass spectrometry (Mann, 1999; Gygi et al., 1999b). These techniques have improved the ability to demonstrate concentration differences between specific proteins from different samples and the peptides can be labelled in the enzymatic cleavage process. Such methods are discussed further in Chapter 4.

3.3.2 Problems associated with 2-dimensional electrophoresis

A range of difficulties was encountered with initial attempts to separate proteins from the mitochondrial samples by 2DE. Two-dimensional electrophoresis had not been used previously in our laboratory, thus this study required the purchase and set-up of instrumentation. Standard samples were employed to obtain the expertise required for 2DE and the technique was then applied to mitochondrial samples. Optimisation of suitable samples loads, conditions for separation, equilibration of the IPG strips prior to the second dimension SDS-PAGE and staining techniques was necessary. Notable problems that were encountered included the presence of KCl in early preparations, the presence of BSA in various samples and the use of PPIX to induce PT. KCl, which interfered with the IEF stage of 2DE, was removed by ultrafiltration and was later excluded from the sample preparation. BSA, which resulted in a large broad stained region on the gels, was removed after cell lysis through buffer exchanges (Alison Smail,
Department of Biological Sciences, University of Wollongong, N.S.W.). Removal of PPIX proved to be more difficult.

The PPIX used to induce PT in mitochondria caused problems in IEF of the samples. Varying concentrations of PPIX were used to induce LC-PT. For high conductance PT (HC-PT), 4 mM PPIX was used in the buffer to induce PT. The residual PPIX in the supernatant fractions and the water-soluble fractions from the HC-PT samples caused a dramatic loss in the resolution of protein spots on the gels (data not shown). For LC-PT, a concentration of 20 μM PPIX was used. This lower concentration was less detrimental to the gels but also resulted in lower resolution of proteins. The resolution was increasingly compromised as the sample load increased. To improve the resolution of the gels from samples containing PPIX, various attempts were made to remove the PPIX prior to 2DE (data not shown). The most useful procedure involved acidification of the sample and extraction with diethyl ether. The major problem with this method was the significant decrease in the number of gel spots, suggesting that the procedure caused precipitation and loss of proteins. Therefore gels from mitochondria that had undergone HC-PT could not be analysed further. Gels from mitochondria that had undergone LC-PT were useful even though the PPIX affected the resolution. Increasing the sample load resulted in a decrease in the resolution of the spots in the gels from the supernatant and first extraction of mitochondria that had undergone LC-PT. This was presumably a consequence of an increase in the concentration of PPIX. Therefore, to increase the concentration of peptides in the digests, multiple gels were obtained and equivalent spots were combined from a number of gels.

PPIX was not completely soluble and the presence of insoluble PPIX presumably resulted in the diminished resolution. IEF is affected by even low concentrations of insoluble material or particulates (Janson and Ryden, 1989). Centrifugation did not
successfully remove the PPIX and, since the extractions were found to cause protein loss, it was difficult to resolve this issue to achieve more definitive results.

3.3.3 Visualisation of proteins on gels

Coomassie Brilliant Blue staining of the proteins in the gels was not sufficiently sensitive and as a result silver staining was used. The intensity of the stain and the darkness of the background were affected by small differences in the time of each step in the staining process. Hence, obtaining similar staining of equivalent samples run under identical conditions was very difficult.

Silver staining is reported to be up to 100 times more sensitive than Coomassie staining (Lauber et al., 2001). The detection limit of silver staining lies between 1 and 10 ng, compared to 100 ng for Coomassie staining (Shevchenko et al., 1996c). The reproducibility of silver staining is, however, affected by temperature and the complex staining procedures may result in up to 20% variation in intensity (Patton, 2000). Silver stains have also been shown to result in lower sequence coverage than Coomassie stains (Patton, 2000). The sequence coverage of proteins from MALDI-MS analyses of in-gel digests may be improved by destaining of the gel pieces with potassium ferricyanide and sodium thiosulphate prior to digestion (Gharahdaghi et al., 1999). Another problem with silver staining is that the stain is not quantitative because the intensity is dependent upon individual protein sequence and post-translational modifications (Lauber et al., 2001; Herbert et al., 1997).

Comparisons between gels were complicated by differences in the intensity of staining, and recovery of peptides from in-gel digests of the spots was also problematic. To obtain peptides for detection by mass spectrometry, it was found that gel spots from a number of equivalent gels had to be combined. The combining of gel spots was
complicated by gel-to-gel staining variations since the same spot had to be carefully chosen from gels that had different intensities of silver stain. With protein spots of low intensity, some gels did not show the spots and therefore could not be used to pool equivalent proteins.

Fluorescent staining has become the method of choice for detection of proteins on two-dimensional gels (Lauber et al., 2001; Malone et al., 2001; Patton, 2000). Fluorescent stains, such as SYPRO Ruby, are MS compatible and have low limits of detection (Malone et al., 2001). A further advantage is the large dynamic range and an improvement in recovery of peptides from in-gel digests (Lopez et al., 2000a). For small laboratories, however, the expense of these commercial stains may restrict their use, particularly when re-use of the stain is not recommended.

3.3.4 Identification of peroxisomal proteins in mitochondrial preparations

Another problem that was encountered with the identification of proteins from mitochondria was the discovery of peroxisomal proteins. The identification of these proteins suggested that the preparation of mitochondria was contaminated with peroxisomes. The contamination of other cellular components in the subfractionation of organelles is a common problem associated with the analysis of organelle proteomes (Rabilloud et al., 1998). Organelles of similar densities may cause contamination of mitochondrial fractions, regardless of the methods used (Lopez et al., 2000b). The separation of mitochondria from other components and organelles may also be limited because of interactions between organelles (such as the mitochondria and the endoplasmic reticulum) (Hanson et al., 2001; Lopez et al., 2000b). Catalase, which was identified in this work, is a highly concentrated protein that is found within peroxisomes and catalyses the reduction of hydrogen peroxide to water and oxygen (Mathews and
van Holde, 1990). Lopez et al. also identified catalase in an affinity preparation of mitochondria (Lopez et al., 2000b). The other peroxisomal protein that was identified in this work was acetyl-CoA C-acyl transferase. The presence of these proteins suggests that the contents of the contaminating peroxisomes were released into the supernatant of the mitochondrial preparation.

3.3.5 MALDI mass fingerprint and nanoESI-MS/MS sequence tag searches

Identification of proteins from 2-dimensional electrophoresis using mass spectrometry and database searching has focussed on two main approaches: i) MALDI mass fingerprinting and ii) sequence tag searching with ESI MS/MS data. Factors that have resulted in preferential use of one technique over the other have included access to instrumentation, sample preparation, sample consumption, time required for data acquisition and analysis, quality of the results obtained, the ability to analyse mixtures of different proteins and the use of automated procedures.

MALDI-TOF mass spectrometers have generally been more accessible to proteomic laboratories owing to their lower cost and ease of operation. Other reasons for the preferred use of MALDI mass fingerprinting include lower sample consumption, less time required for analysis, and the potential for automation. Prior to the introduction of nanoESI techniques, the small sample volumes required for MALDI compared to ESI analyses resulted in the wide application of MALDI analysis for identification of proteins from gels. With current technologies, both MALDI and nanoESI require only small volumes and subpicomole levels (usually hundreds of femtomoles) are all that is required for identification of gel-separated proteins (Blackstock and Weir, 1999). The time required for sample analysis is typically greater for MS/MS analyses, particularly given automated procedures exist for in-gel digests,
sample clean-up and MALDI target spotting. Automated sample handling may also be implemented for MS/MS analyses, particularly with LC-MS applications (further discussed in Chapter 4), but this is limited when off-line nanoESI from borosilicate capillaries is used.

The data obtained from MS/MS analyses is often more useful than the information from MALDI analyses. Mass fingerprinting data is limited when more than one protein is digested in the same sample as there is no way of distinguishing which peptides belong to a particular protein. Protein spots from 2DE may contain more than one protein, particularly in regions of the gel that are highly populated and tend to have overlapping proteins. MS/MS data requires only a single peptide for identification and is therefore highly suited to the analysis of complex mixtures (as demonstrated in Chapter 4). Mass fingerprinting data is also not suitable for searching the database of expressed sequence tags (ESTs). The ESTs only cover a portion of a gene sequence and may not represent sufficient peptides for identification through peptide mass fingerprinting (Aebersold and Goodlett, 2001). Sequence tag data therefore allows access to a vast amount of information that peptide mass fingerprinting does not cover. An added advantage with MS/MS data is the potential to assign large stretches of sequence in many cases. The sequence can be used to search databases without any mass information in a “sequence only” search. This type of search may be useful for the identification of proteins from modified peptides and unexpected enzymatic cleavages. Examples of the application of “sequence only” searches are presented in Chapter 4.

3.3.6 Capillary LC-MS

Capillary LC-MS has enhanced the sensitivity of protein identification by MS based methods and improved the identification of peptides in complex mixtures.
Technological advances in miniaturisation of HPLC columns and solvent delivery systems have resulted in the capability for interfacing HPLC systems to low-flow ESI sources, with or without flow splitting. The CapLC™ system used in this work is capable of delivering flow rates as low as 1µL/min for gradient separations and 200 nL/min for isocratic separations.

Capillary LC-MS has a number of advantages for proteomic applications. The separation of components in a mixture results in improved sequence coverage, and in the case of complex mixtures, identification of proteins that might otherwise be missed. Background contamination, which is often a significant problem in the identification of gel-separated proteins, may interfere with detection of peptide ions. The improvement in sensitivity in LC-MS results in a lower limit of detection and may facilitate protein identification that might not be achieved with off-line nanoESI-MS/MS.

LC-MS and LC-MS/MS have a greater potential for automation than do off-line nanoESI-MS and nanoESI-MS/MS. The ability to automatically load samples and acquire data results in higher throughput of samples. This is crucial in conventional proteomics applications when many in-gel digests are generated and there is a demand for analysis of a large number of samples. With MS detection, resolution of individual peptides is not necessary as multiple components may be detected and subjected to MS/MS simultaneously. Therefore fast gradient analyses may be used, increasing sample throughput (Tomer, 2001).

3.3.7 Automated procedures for conventional proteomics

The need for methods to increase the throughput of samples for proteomic analysis is evident from the significant amount of time required for sample preparation, 2DE, gel staining, analysis of the gels for novel spots and subsequent excision and processing of
samples for MS analysis. Other important advantages of automated procedures are that gel-to-gel variations are minimised and the chance of introducing contaminants through human handling (such as keratin) is reduced. Large-scale proteomics laboratories are able to run numerous gels in batches to reduce gel-to-gel variations. The use of densitometers and computer software in the identification of novel spots assists with gel comparisons. The software also enables protein assignments and labelling for excision by gel-cutting robots. Robots also perform in-gel digests and the resulting extracted peptides are automatically loaded on to MALDI targets, or on to microtitre plates for LC-MS.

Small laboratories that do not have access to equipment for automation require significantly more time to produce gels and may suffer from increased gel-to-gel variation and contamination. In our laboratory, up to twelve IPG gels could be run simultaneously and four second-dimension gels could be run simultaneously (in two tanks connected with the same power supply and cooling system). For mini-gels, only two second-dimension gels could be run simultaneously. Gel-to-gel variations were also introduced in the casting of gels for the second dimension. A maximum of four gels could be cast at one time with the available equipment. Precast mini-gels minimised variations but were costly and have insufficient capacity for identification of in-gel digested peptides by MS. As mentioned above, the complicated steps in silver staining produced significant gel-to-gel variations, which further complicated the comparison of gels from different samples.

3.3.8 Summary

Conventional proteomics has relied upon 2DE to separate proteins in a complex sample prior to analysis by MS. The ability of 2DE to provide an image of the protein
complement of an organism, tissue or organelle has enabled comparisons of normal and diseased or treated samples. The ability to distinguish differences between a control sample and a diseased sample, and to subsequently identify the proteins of interest has resulted in the identification of many interesting proteins.

The establishment of proteomes of organisms and subcellular fractions from tissues is continuing to increase the information available for rapid identification of proteins. The identification of peroxisomal proteins in the gels from mitochondria in this study, however, shows one of the potential problems with the analysis of subcellular fractions. Careful separation of the organelles is required to minimise the contamination by other organelles and cellular components.

The application of 2DE to the analysis of mitochondrial samples was accomplished in this work after some initial difficulties with particular components in the samples such as BSA, KCl and PPIX. The complexity of the mitochondrial samples was clear from the number of proteins that were resolved following differential solubilisation of the mitochondria. The supernatant from the mitochondrial samples also showed a large number of resolved proteins, some of which were known mitochondrial matrix proteins. The variation between gels and difficulty in recognising proteins of interest, however, complicated identification of changes that occurred in the mitochondrial supernatant following LC-PT. The ability to detect differences in the concentration of a particular protein between gels was also found to be a limiting factor.

Two-dimensional electrophoresis is a time-consuming and labour-intensive technique. In addition, the information obtained usually does not provide any detail on the function of the protein or possible protein-protein interactions that may be important. The results from 2DE and MS analysis of the proteins in small scale laboratories is expected to be surpassed by analyses in large scale laboratories owing to
greater access to automated procedures. Development of alternative techniques for the separation of proteins is of interest to many research groups. Techniques that not only provide separation but also give some indication of protein function, i.e. functional proteomics projects, are increasingly being developed. Long-term goals may include the removal of 2DE from many laboratories but current methodologies have not found a replacement that allows the visualisation provided by 2DE.
CHAPTER FOUR: ISOLATION AND IDENTIFICATION OF CD4 AND CO-ASSOCIATING PROTEINS

4.1 INTRODUCTION

CD4 is a type I transmembrane glycoprotein from the immunoglobulin superfamily (IgSF) of proteins. CD4 is expressed on the surface of T cells and also, to a lesser extent on Langerhan’s cells, dendritic cells, macrophages and monocytes (Bour et al., 1995). CD4 has four immunoglobulin-like domains, D1, D2, D3, and D4, a transmembrane region and a cytoplasmic tail of 38 amino acids (Bour et al., 1995). The most widely known function of CD4 is amplification of T lymphocyte activation. CD4 has been implicated in autoimmune disorders and transplantation reactions and is known to be the major cellular receptor for the human immunodeficiency virus (HIV) (Huang et al., 1997).

4.1.1 CD4 in T-cell activation

The immune system can be divided into the humoral (or antibody) response and cell-mediated immune responses (Alberts et al., 1994; Kuby, 1997). The two main cell types involved are the B lymphocytes and the T lymphocytes (B and T cells). B cells are antibody-producing cells, which are recognised as the major effector cells of the humoral response. T cells are crucial in the cell-mediated response in which specialised cells recognise foreign antigens displayed on the surface of other cells. T cells can be divided into two groups, cytotoxic T cells and T helper cells (Kuby, 1997). Both these groups are important in cell-mediated immunity and T helper cells are also crucial for assisting B cells
in humoral immune responses. Another important aspect of the immune system is its ability to respond very quickly and with more intensity towards an antigen that has been recognised previously. The initial encounter with an antigen elicits a primary immune response (Alberts et al., 1994; Kuby, 1997). The primary immune response results in clonal expansion of cells that are able to recognise the antigen. Some of these cells are memory cells, which can live for long periods of time (possibly years) and are able to produce a secondary response if the same antigen is encountered at a later stage (Alberts et al., 1994; Kuby, 1997).

T cells recognise an antigen only once it has been processed within, and displayed upon, an antigen-presenting cell (APC) (Jensen, 1997; Vignali et al., 1993; Krummel et al., 2000). The antigen is bound to a major histocompatibility complex (MHC) molecule on the surface of an APC as shown in Figure 4.1. There are two types of these heterodimeric MHC proteins, MHC class I and MHC class II. The T cell receptor (TCR), in association with another protein complex called CD3, binds to the complex of peptide antigen and MHC molecule. The TCR is also a heterodimer and is always found associated with CD3. T cells do not contain a cytoplasmic domain so CD3 serves as the signal transduction component of the T cell receptor complex (Krummel et al., 2000). The affinity of the TCR for a peptide-MHC complex is quite low (K_D >100µM) (Huang et al., 1997; Krummel et al., 2000), however, there are numerous other molecules involved in binding of the TCR/CD3 complex to peptide-MHC and/or subsequent T cell activation and these are outlined in Table 4.1.
Figure 4.1 T-cell receptor recognition of antigen bound to an MHC Class II molecule. The T-cell receptor along with CD3 is expressed on the surface of a T helper (T_H) cell. CD4 is also expressed on the T_H cell and its cytoplasmic domain is associated with the protein tyrosine kinase, p56^{Lck}. The MHC class II molecule is expressed on the surface of an antigen-presenting cell (APC). See text for further details.
<table>
<thead>
<tr>
<th>Name</th>
<th>Cell type</th>
<th>Binds to</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>T cell receptor (TCR)</td>
<td>T cell</td>
<td>Peptide-MHC</td>
<td>Recognition of antigen on APC, T cell activation</td>
</tr>
<tr>
<td>CD3</td>
<td>T cell</td>
<td>TCR, Peptide-MHC</td>
<td>Signal transduction of TCR engagement</td>
</tr>
<tr>
<td>MHC class I</td>
<td>APC</td>
<td>TCR/CD3, CD8</td>
<td>Antigen presentation</td>
</tr>
<tr>
<td>MHC class II</td>
<td>APC</td>
<td>TCR/CD3, CD4</td>
<td>Antigen presentation</td>
</tr>
<tr>
<td>CD4</td>
<td>T cell (usually helper T cell)</td>
<td>TCR/CD3/peptide-MHC class II</td>
<td>Enhance T cell activation, recruitment of p56\text{lk}</td>
</tr>
<tr>
<td>CD8</td>
<td>T cell (usually cytotoxic T cell)</td>
<td>TCR/CD3/peptide-MHC class I</td>
<td>Enhance T cell activation, recruitment of p56\text{lk}</td>
</tr>
<tr>
<td>LFA-1</td>
<td>T cell</td>
<td>ICAM-1 (on APC)</td>
<td>Enhance T cell activation</td>
</tr>
<tr>
<td>CD2 (LFA-2)</td>
<td>T cell</td>
<td>LFA-3 (on APC)</td>
<td>Enhance T cell activation</td>
</tr>
<tr>
<td>CD28</td>
<td>T cell</td>
<td>B7.1, B7.2 (on APC)</td>
<td>Enhance or suppress T cell activation</td>
</tr>
<tr>
<td>CD45</td>
<td>T cell</td>
<td>p56lck, CD4, CD45-AP</td>
<td>Regulate p56lck activity</td>
</tr>
</tbody>
</table>

Table 4.1 Molecules involved in T cell activation, their location, binding partners and functions.

CD4 and CD8 are co-receptors involved in T cell recognition (Jensen, 1997; Krummel et al., 2000). T cells that express CD4 on their surface (CD4 positive cells, CD4⁺)
recognise complexes of peptide antigen with MHC class II molecules on APCs, whilst T cells expressing CD8 (CD8+ cells) bind to peptide-MHC class I. MHC class I molecules are found on most cells of the body whereas MHC class II molecules are specifically found on APCs such as macrophages, dendritic cells and B lymphocytes (Jensen, 1997). The MHC class I molecules display peptides that have been synthesised and degraded within a cell and may be derived either from foreign proteins (such as those a virus would cause to be produced) or self-proteins. T cells able to recognise self-antigen are destroyed in the process of T cell development that occurs in the thymus (see below), therefore MHC class I molecules in a complex with self-peptide antigens are not recognised by mature T cells. If a CD8+ cytotoxic T cell recognises peptide antigen in a complex with an MHC class I molecule, the antigen will usually have derived from viral infection and the cell displaying the antigen is destroyed (Jensen, 1997). The peptides displayed by MHC class II molecules are vastly different to those displayed by class I (Jensen, 1997). MHC class II molecules are able to bind much longer polypeptides than class I and the peptides may be derived from internalised proteins that are from exogenous sources (Lederman and Suciu-Foca, 1999). Most CD4+ T cells are T helper cells and play an important role in the regulation of immune responses such as antibody production and cytotoxic T cell function. Although CD4+ cells are usually of the helper phenotype and the majority of CD8+ are cytotoxic or suppressor T cells, there is some crossover (Bour et al., 1995; Lederman and Suciu-Foca, 1999). The distinction between CD4 and CD8 expression is therefore closely associated with MHC class specificity rather than the T cell type.

The binding of CD4 to the recognition machinery amplifies T-cell activation by a factor of 10-100 (Krummel et al., 2000). CD4 binds weakly to the peptide-MHC class II
complex through domains D1 and D2. The TCR/CD3 complex associates with CD4 via the extracellular portion of CD4, most probably through domain D3 (Vignali and Vignali, 1999) and is also connected via a protein tyrosine kinase (p56\textsuperscript{ck}) associated with the cytoplasmic tail of CD4 (Sweet \textit{et al.}, 1991; Vignali \textit{et al.}, 1993). The association of TCR/CD3, peptide-MHC class II and CD4 at the cell surface result results in optimal activation of the T cell (Figure 4.1) (Li \textit{et al.}, 1998b; Bour \textit{et al.}, 1995; Vignali and Vignali, 1999). There is mounting evidence that suggests the active form of CD4 is dimeric or oligomeric (Huang \textit{et al.}, 1997; Li \textit{et al.}, 1998b; Gratton \textit{et al.}, 2000) and studies have shown that T cell activation requires co-aggregation between CD4/p56\textsuperscript{ck} and the TCR (Gratton \textit{et al.}, 2000).

CD4 and CD8 are both associated with p56\textsuperscript{ck}, an src-related tyrosine kinase (Veillette \textit{et al.}, 1988). The association occurs via two cysteine residues of CD4 and two cysteine residues in the N-terminal portion of p56\textsuperscript{ck} via Zn\textsuperscript{2+} (Turner \textit{et al.}, 1990; Gratton \textit{et al.}, 2000). The activation of p56\textsuperscript{ck} is required for initiation of the tyrosine kinase cascade that occurs intracellularly following TCR engagement (Zamoyska, 1998). An increase in the phosphotyrosyl content of a number of proteins occurs just prior to T cell activation (Gratton \textit{et al.}, 2000). Another tyrosine kinase, ZAP-70, is also required for T cell activation and is thought to be activated by p56\textsuperscript{ck} (Gratton \textit{et al.}, 2000; Bosselut \textit{et al.}, 1999). The TCR/CD3/peptide-MHC association with CD4 brings p56\textsuperscript{ck} in close proximity to the TCR/CD3 complex and improves phosphorylation of sites on CD3 and ZAP-70 (Sweet \textit{et al.}, 1991; Bosselut \textit{et al.}, 1999). ZAP-70 inturn phosphorylates LAT (linker for activation of T cells), which is also associated with CD4 (Bosselut \textit{et al.}, 1999). P59\textsuperscript{fyn} is
another src-family tyrosine kinase that may be activated upon TCR/CD3 engagement with peptide-MHC (Bosselut et al., 1999).

Following the association of CD4 with the TCR/CD3 complex, CD4 is phosphorylated, it dissociates from p56\textsuperscript{ck} and is removed from the surface of the T cell (Vignali et al., 1993). The dissociation of p56\textsuperscript{ck} from CD4 has been proposed to result from phosphorylation of CD4 by protein kinase C (PKC) (Sweet et al., 1991; Vignali et al., 1993).

There are many other molecules that are known to be important in the adhesion and activation of T cells. LFA-1, a member of the integrin family on the surface of a T cell, recognises ICAM-1 on an APC once the TCR has engaged with peptide-MHC (Krummel et al., 2000). Two molecules on the T cell surface, CD28 and CTLA-4, bind to the molecules B7.1 (CD80) and B7.2 (CD86) on other cells (Krummel et al., 2000). Binding of B7.1 enhances the T cell activation whilst binding to B7.2 suppresses T cell activation. CD28 is important in amplification of activated T cell responses and is often required for naïve T cell responses (Krummel et al., 2000), i.e. in the primary immune response of T cells.

The ability of CD45 to regulate p56\textsuperscript{ck} is crucial in T cell activation (Veillette et al., 1999; Bonnard et al., 1997). CD45 has been shown to associate with CD4 in resting T cells and has also been shown to modulate T cell activation through this association (Bonnard et al., 1997; Gratton et al., 2000). A transmembrane protein known as CD45 associating phosphoprotein (CD45-AP) and CD45 form part of the large multiprotein complex for optimal T cell activation that includes the TCR, CD4 and p56\textsuperscript{ck} (Veillette et al., 1999). An isoform of CD45 that is expressed on memory T cells remains permanently associated with
CD4 and may play an important role in the activation of memory T cells (Bour et al., 1995).

Interleukin 16 (IL-16) is another ligand for CD4. The binding site of IL-16 is distinct from the MHC site (Center et al., 2000). The binding of IL-16 to CD4 causes crosslinking of CD4. If this binding occurs in the absence of antigen, CD4 no longer interacts with the TCR and the cell enters a state of anergy. In this way, CD4 can act to either activate the T cell by interacting with TCR/MHC when antigen is presented by the correct MHC molecule, or, to inhibit T cell activation via IL-16 but still produce inflammatory signals (Center et al., 2000). The state of unresponsiveness of the IL-16 T cells is reversible after sufficient time has elapsed and the T cell is then able to be activated upon binding an antigen-MHC class II complex. IL-16 may therefore play a role in the accumulation of T cells for future antigen recognition or in preventing prolonged activation of T cells (Center et al., 2000).

4.1.2 CD4 in thymocyte differentiation

T cells develop from thymocytes in the thymus. It is important that the T cells are able to recognise foreign antigens presented by self-MHC molecules but that they are not reactive to self-antigen (Finkel et al., 1991). In a process termed ‘positive selection’, the T cells that express TCRs which recognise self-MHC molecules survive, but those that express TCRs not specific for self-MHC molecules are destroyed because they would not be useful to the cell (Alberts et al., 1994). A second process, termed ‘negative selection’, ensures any T cells that form high-affinity complexes with self-antigen bound to MHC molecules are also destroyed in the thymus (Finkel et al., 1991; Alberts et al., 1994). The thymocytes begin as
double negative (DN) cells that express neither CD4 nor CD8. These cells then begin to express both CD4 and CD8 and are termed double positive (DP) thymocytes, as they are both CD4$^+$ and CD8$^+$. The DP cells then undergo a process so that they express either CD4 or CD8 but not both. DP cells expressing TCRs that recognise MHC class I molecules mature into CD8$^+$ cells, whilst DP cells with TCRs that recognise MHC class II molecules become CD4$^+$. The co-receptors CD4 and CD8 are crucial in both positive and negative selection but the mechanism by which a fully functional single positive T cell is produced is not completely understood (Finkel et al., 1991). Two models have been proposed to explain the change in expression of CD4 and CD8 from a double positive to a single positive T cell (Sweet et al., 1991; Killeen et al., 1993; Hernandez-Hoyos et al., 2000). The instructional model proposes that the association of CD4 and a TCR/MHC class II complex causes the expression of CD8 to be stopped. Likewise, an instructional signal produced upon the association of CD8 with a TCR/MHC class I complex informs the cell to cease expression of CD4. The second model, known as the stochastic model, proposes that the co-receptor selection occurs independently of the TCR specificity (Hernandez-Hoyos et al., 2000). In this model the expression of one of the co-receptors is stopped independently of TCR association with peptide-MHC. Since recognition of peptide-MHC is dependent on expression of the appropriate co-receptor, CD8$^+$ T cells expressing TCRs that recognise MHC class II molecules do not mature, whilst CD8$^+$ cells expressing TCRs that recognise MHC class I molecules are able to be activated. Correspondingly, CD4$^+$ cells expressing TCRs that recognise MHC class I molecules do not mature, whilst the CD4$^+$ cells with the correct TCR specificity are able to mature into fully activated T cells.
Recently a role of p56\textsuperscript{ck} in the determination of co-receptor lineage has also been proposed (Hernández-Hoyos et al., 2000) that supports the instructional model of co-receptor commitment. Hernández-Hoyos et al. found that the activity of p56\textsuperscript{ck} during positive selection of self-MHC molecules has an effect on the selection of either CD4 or CD8. The thymocytes are able to determine quantitative differences between the p56\textsuperscript{ck} signals and thereby control the expression of only one of the co-receptors. P56\textsuperscript{ck} is associated with CD4 or CD8 and its activity (phosphorylation of CD3) is dependent on the association of a co-receptor with a TCR. The co-receptor itself therefore has the ability to influence the commitment to CD4\textsuperscript{+} or CD8\textsuperscript{+}.

4.1.3 The role of CD4 in HIV entry

\emph{Human immunodeficiency virus}

It is now widely established that the retrovirus HIV is the causative agent of acquired immunodeficiency syndrome (AIDS). It is estimated that there are 36.1 million people worldwide who are living with HIV and in the two decades since the virus was identified, 21.8 million people have died owing to implications directly resulting from AIDS (UNAIDS, 2000). The development of more effective preventative measures and therapeutic treatments are crucial in the fight against this epidemic.

The discovery of HIV in 1986 revealed two different variants, which were called HIV-1 and HIV-2 (Kuby, 1997). HIV-1 is much more widespread than HIV-2 and appears to be much more virulent (Kuby, 1997).

The genome in retroviruses is encoded by RNA and is reverse-transcribed to DNA in the host cell (Turner and Summers, 1999). The genome of HIV encodes 2 envelope
proteins, 3 structural proteins, 3 enzymes and 6 accessory proteins (Turner and Summers, 1999). A simplified schematic of the HIV life-cycle is shown in Figure 4.2. The initial event of infection involves fusion with the cell membrane. Once the virus is within the cytosol of the host protein, reverse transcription occurs and the viral DNA produced is transported to the cell nucleus. The viral DNA is then integrated into the genome of the host. Messenger RNA (mRNA) is synthesised from the DNA, as in a healthy, uninfected cell, and the mRNA is transported out of the nucleus. Translation of the mRNA into the viral proteins occurs and the proteins, along with unspliced RNA, form new viral particles that bud off from the cell membrane. (Turner and Summers, 1999; Kuby, 1997; Bour et al., 1995; Weiss, 2001; Bour et al., 1994; Alberts et al., 1994)

Initial infection with HIV causes a dramatic increase in the viral load and a lack of T helper cell response to the virus (Altfeld and Rosenberg, 2000). Activated T cells are more susceptible to HIV and T cell activation greatly enhances the replication of HIV-1 (Bour et al., 1995; Picker and Maino, 2000). Following the initial production of virus, the virus load decreases to a steady low level but there is still a lack of T helper cell response (Altfeld and Rosenberg, 2000). Cytotoxic T cell responsiveness is apparent through the rise and fall in viral load and these cells are critical in the immune response to HIV (Picker and Maino, 2000). The loss of CD4\(^+\) T cells occurs well after the loss of T helper cell function (Altfeld and Rosenberg, 2000). A marked increase in the viral load and a decrease in the cytotoxic T cell function lead to progression of the disease (Altfeld and Rosenberg, 2000).
Figure 4.2 The lifecycle of HIV\textsuperscript{1}. Binding of the HIV surface glycoprotein, gp120, to CD4 and recruitment of a chemokine receptor results in fusion of the HIV membrane with the cell membrane. The RNA of HIV is reverse transcribed into viral DNA, which enters the nucleus of the cell and is integrated into the genome of the host. Viral RNA is produced through the cell's own machinery and proteins are translated from the viral RNA. A new virion is assembled and buds off from the cell membrane. A mature HIV particle is thus produced.

\textsuperscript{1} Artist St John Skilton
**HIV fusion with the cell membrane**

The surface of HIV particles is coated with glycoproteins. A surface (SU) glycoprotein, gp120 is anchored to the membrane by another transmembrane (TM) glycoprotein, gp41 (Sweet et al., 1991; Turner and Summers, 1999). Each of these glycoproteins are generated from the cleavage of the gp160 envelope glycoprotein precursor (Sweet et al., 1991; Sattentau and Moore, 1993; Bour et al., 1995). The binding of an HIV particle to a cell is mediated through interactions between D1 of CD4 and gp120 (Sweet et al., 1991). In addition, although they are separate, there is some overlap between the gp120 binding site and the MHC class II binding site of CD4 (Bour et al., 1995). The binding of gp120 to CD4 induces a conformational change in both gp120 and CD4 (Yachou and Sekaly, 1999) but alone is not sufficient for HIV entry (Moore et al., 1997). After initial binding to CD4, a chemokine receptor is recruited and binds to the gp120/CD4 complex (Fauci, 1996). The association with the chemokine receptor is critical for viral entry. It is thought that binding of gp120 to the chemokine receptor induces further conformational changes in gp120 that culminate in fusion (Kwong et al., 1998; Turner and Summers, 1999; Horuk, 1999). These conformational changes in gp120 cause it to dissociate from gp41, which possibly induces conformational changes in gp41, exposing sites on gp41 that are necessary for fusion (Bour et al., 1995; Yachou and Sekaly, 1999; Turner and Summers, 1999). A hydrophobic fusion peptide of gp41 is inserted into the cell membrane, which is followed by fusion of the host cell membrane with the viral membrane (Wyatt and Sodroski, 1998; Kwong et al., 1998).

HIV-1 varies in its ability to infect particular cells. Immediately after infection, the virus usually infects and replicates within monocytes and macrophages that express CD4. The virus is said to be M-tropic, even though there is some ability to replicate in T cells
(Moore et al., 1997). M-tropic strains are responsible for initial HIV infection (Fauci, 1996). A virus that preferentially infects T cells is termed T-tropic and usually dominates with the progression of the infection (Fauci, 1996; Choe et al., 1998; Horuk, 1999). T-tropic strains are involved in laboratory cell lines and in vivo forms often have the ability to infect T cells and macrophages (dual tropic) (Choe et al., 1998). The dual tropic forms appear to dominate later in the course of infection (Choe et al., 1998; Wyatt and Sodroski, 1998) and are thought to be intermediate in the change from M-tropic to T-tropic strains (Berson and Doms, 1998). The T-tropic and M-tropic strains of HIV-1 use different chemokine receptors as co-receptors for fusion with the host cell. T-tropic virus utilises CXCR4 whilst CCR5 is the co-receptor for M-tropic strains (Deng et al., 1996; Dragic et al., 1996; Moore et al., 1997; Fauci, 1996; Horuk, 1999). CCR5 is the most commonly used co-receptor through most of the infection (Choe et al., 1998; Wyatt and Sodroski, 1998).

Gp120 is also able to interact with CD4 independently of gp41. Gp120 antibodies in complex with gp120, free gp120, and multimeric gp120 are able to bind to CD4 at the surface of uninfected cells. This binding may induce T cell cytolysis, anergy or apoptosis (Bour et al., 1995).

**Down-regulation of CD4 from the cell membrane**

The reduction in circulating CD4+ cells has been used as a marker for the progression of HIV to AIDS (Richman, 2001; McCune, 2001). The killing of infected CD4+ cells is not the only mechanism that results in the reduction in CD4+ cells. HIV invokes the down-regulation of CD4 from the surface of infected cells. There are a number of ways in which this down-regulation of CD4 may be achieved (Bour et al., 1995). A reduction in the levels
of CD4 mRNA in HIV infected cells has been reported and may be the result of the inhibition of CD4 transcription or of a decrease in the stability of the CD4 mRNA (Bour et al., 1994). CD4 has also been shown to bind to gp160 (the precursor to gp120 and gp41) in the endoplasmic reticulum (Bour et al., 1994). Such intracellular complexes impair the expression of CD4 on the cell surface. A protein encoded by the HIV genome called Vpu also binds to CD4 in the endoplasmic reticulum and promotes CD4 degradation (Bour et al., 1994). The trapping of CD4 by gp160 enhances CD4 degradation by Vpu (Bour et al., 1995). Another protein produced by the HIV genome is the Nef protein. Nef has the capability of binding to the cytoplasmic tail of CD4 and to p56^ck (Turner and Summers, 1999; Bour et al., 1995) and is known to down-regulate cell-surface CD4, independently of the transcription and translation of CD4 (Bour et al., 1995). It acts at the surface to cause internalisation and degradation of CD4 in lysosomes (Turner and Summers, 1999).

**Anti-HIV therapies**

The drugs that are currently available for the treatment of HIV target two enzymes in the virus life-cycle: reverse transcriptase and the protease enzyme. The available drugs are usually used in combinations to improve the efficacy of attack on HIV (Turner and Summers, 1999).

The first drugs developed for HIV were nucleosides and targeted the enzyme reverse transcriptase (Figure 4.2) (Richman, 2001). Non-nucleoside inhibitors of reverse transcriptase have also been developed. Reverse transcriptase inhibitors are still crucial for anti-HIV therapy but cause undesirable side effects such as liver toxicity. The virus has shown the ability to develop resistance to reverse transcriptase inhibitors, particularly when administered alone. There are currently 6 nucleoside reverse transcriptase inhibitors and 3

The other available antiviral drugs target the protease enzyme that is required for processing of the viral proteins produced by the host cell. There are 5 currently available protease inhibitors (World Health Organization, 2000; HIV/AIDS treatment information service, 2001). Side effects from protease inhibitors have including insulin intolerance, increased cholesterol levels and redistribution of body fat (Richman, 2001). Resistance to protease inhibitors also develops (Turner and Summers, 1999). Combination therapies have been shown to more effective in HIV control and protease inhibitors are usually used in combination with reverse transcriptase inhibitors.

The third enzyme in the HIV life cycle, integrase, is involved in the integration of viral DNA into the host genome and has also been targeted for HIV therapy. There are, however, no integrase inhibitors currently available for anti-HIV treatment. Preventing the infection by vaccination has been shown to be very difficult (Landau, 1999). Other possible targets that are currently being developed include inhibitors of gp120, gp41, the chemokine receptors CCR5 and CXCR5, Tat (important for replication of HIV), Nef and Rev (crucial for transport of viral DNA). Many inhibitors of these targets have been found to be inefficient or produce toxic side effects. Further understanding of the life cycle of HIV will assist in the development of therapies and identify new targets for anti-HIV drugs.

4.1.4 Aims

The use of a functional proteomics approach may provide insight into proteins that have not previously been found to associate with CD4. The identification of such proteins could
expand our understanding of the many roles that CD4 plays and may reveal potential targets for the prevention of HIV entry and treatment of immune disorders. The aim of this project was to develop a method for the isolation and identification of proteins that associate with CD4.
4.2 RESULTS

4.2.1 Overview of method and method development

Figure 4.3 shows a summary of the method developed for the isolation and characterisation of proteins that co-elute with CD4 from an affinity column. The first four steps, CEMT4 cell preparation and lysis, IgG₁ and anti-CD4 separation and the anion exchange chromatography were performed at the Millennium Institute, Centre for Virus Research (University of Sydney, Westmead, Australia) by Mr Oliver Bernhard, Dr Andrew Sloane or Dr Tim Hochgrebe.

Ion Exchange Chromatography

A portion of a typical ion exchange chromatography profile in which the peptides of interest eluted is shown in Figure 4.4. Figure 4.4B shows the profile obtained from the digest of IgG₁ column eluate. The profile in 4.4A is from the digest of eluate that passed through both the IgG₁ column and the anti-CD4 column (hereafter referred to as ‘anti-CD4 column’). The bound peptides were eluted using a step gradient of 350 mM NaCl (35% buffer B), which was introduced at 62 minutes, and fractions were collected from this point. Twelve fractions were collected over the elution period (62 to 68 minutes) and these are labelled on the elution profiles. The highest absorbance was recorded in fractions 4 and 5 for digests from both the IgG₁ and the anti-CD4 columns. The percentage buffer B was increased to 60% after 68 minutes to completely flush the column of bound species.
Figure 4.3 Overview of method

Cell preparation and lysis

IgG₁ control column for non-specific binding capture

Anti-CD4 affinity column

Anion exchange chromatography

Desalt fractions (C18 ZipTips™)

MALDI-MS
- Screen for sample viability
- Any significant difference between control and anti-CD4 fractions?

NanoESI-MS
- Multiply charged ions apparent over background?
- Any significant difference between control and anti-CD4 fractions?

NanoESI-MS/MS (Section 4.2.3 for examples)

LC-MS (Section 4.2.4)
- Multiply charged ions apparent?
- Determine threshold for LC-MS/MS

LC-MS/MS with data dependent MS to MS/MS switching (Section 4.2.5)

Database searching
- 'Sequence tag'
- 'Sequence only'
(Section 4.2.3 and 4.2.5 for examples, advantages and limitations)
**Figure 4.4** Ion exchange chromatography profile from 60-70 minutes of the tryptic digests from A) anti-CD4 (Q425) affinity column, and B) IgG1 control column. The UV absorbance at 280nm (blue line) is shown on the left hand axis (arbitrary units). The right hand axis shows the percentage of Buffer B (1M NaCl, 20mM glycine) used for the elution. The buffer used to elute the peptides was 350mM NaCl in 20mM glycine (35% buffer B), shown on the graph by the red line over the elution profile. Salts and detergent eluted over the first 30 minutes and were detected as a large broad absorbance at 280nm (data not shown).
Fractions were tested for CD4 antigen by spotting 2 µL of each fraction on to a nitrocellulose membrane and analysing with a CD4 polyclonal antibody (poly T4-5) (data not shown). Fractions 4, 5 and 6 from the anti-CD4 column digest contained CD4. The same test on the fractions from the IgG₁ column did not indicate the presence of CD4.

**MALDI-MS**

MALDI-MS was used to check for the presence of peptides in the collected fractions. It was necessary to desalt the samples prior to MALDI analysis. Fractions 4-6 from the anti-CD4 preparation and the IgG₁ control preparation were desalted using C₁₈ ZipTips™ (Millipore Corporation, Bedford, MA, U.S.A.). In the initial experiments, when MALDI-MS spectra of the samples from an ion exchange chromatography experiment did not show peptide ions, or there was no apparent difference between the control and anti-CD4 samples, the method was further developed to obtain clearer spectra. Developments included: an increase in the number of cells grown, modifications in the elution of peptides from the IgG₁ control column and the anti-CD4 column (Oliver Bernhard, personal communication) and changes in the solutions used to elute the peptides from the ion-exchange chromatography column (Sloane, 2000).

Figure 4.5 shows MALDI-MS spectra of fraction 5 from an early preparation of the anti-CD4 and control IgG₁ samples. The spectrum from fraction 5 of the anti-CD4 digest (Figure 4.5A) contained many components that were also present in the spectrum of the corresponding fraction from the IgG₁ control digest (Figure 4.5B). The large peak at m/z 1533.4 in the spectrum from the anti-CD4 preparation was not present in the spectrum from the IgG₁ preparation.
Figure 4.5 MALDI-MS spectra of fraction 5 from an early ion-exchange chromatography separation of the column eluates after cleavage with trypsin A) digest of eluate from anti CD4 column, B) digest of eluate from IgG control column
Figures 4.6 and 4.7 show MALDI-MS spectra from fractions 4-6 of the anti-CD4 and control IgG\(_1\) preparation after increasing the number of cells used, exclusion of iodoacetamide from the cell lysis solution and elution from the affinity column with a solution containing glycine rather than EDTA (as in Figure 4.5). Each of the fractions from the anti-CD4 preparation resulted in significantly different MALDI spectra (Figure 4.6A-C), which indicated that further analysis of each of these fractions was warranted. Comparing the spectra from Figure 4.6 to those in Figure 4.7 revealed that there were also significant differences between the anti-CD4 preparation and the IgG\(_1\) preparation. The MALDI-MS spectra also showed components that were present in both the anti-CD4 preparation and the IgG\(_1\) preparation, such as the components of approximate m/z values of 1516 and m/z 3312 in fraction 5 from both preparations (Figure 4.6B and Figure 4.7B).

**NanoESI-MS**

Following desalting and MALDI analysis, the samples were analysed by nanoESI-MS. Representative ESI-MS spectra of fractions 4, 5 and 6 from ion exchange chromatography of the digest from the anti-CD4 column are shown in Figure 4.8. As seen in the MALDI-MS spectra (Figure 4.6), there are noticeable differences between the fractions. Singly and multiply charges ions can be readily distinguished in the ESI mass spectra because the high resolution of the quadrupole-ToF instrument enabled the separation of individual isotopes of each species. The spectra show significant background ions and many singly charged ions were apparent. Fraction 6 from this particular preparation (Figure 4.8C) contained abundant singly charged ions, such as m/z 683.6, 711.6, 739.7, 1041.9, 1069.9 and 1098.0 that dominated the spectrum. Even with the extensive background and singly charged ions,
Figure 4.6 MALDI-MS of fractions from ion-exchange chromatography of the eluate from the anti-CD4 affinity column after cleavage with trypsin. A) Fraction 4, B) Fraction 5, C) Fraction 6
Figure 4.7 MALDI-MS of fractions from ion-exchange chromatography of the eluate from the IgG control column after cleavage with trypsin.
A) Fraction 4, B) Fraction 5, C) Fraction 6
multiply charged components were also apparent in the spectra. For example, a doubly charged ion was apparent at m/z 862.5 in the spectrum from fraction 6 (Figure 4.8C).

ESI-MS spectra of the corresponding fractions from the IgG₁ control column samples are shown in Figure 4.9. The background is also apparent in these spectra and the majority of ions below m/z 600 in these spectra are from singly charged ions. Comparison of the ESI-MS spectra from the anti-CD4 preparation and the IgG₁ preparation showed that some multiply charged ions were present in fractions from both columns, such as m/z 1105 in fraction 5. Close investigation, however, revealed many multiply charged ions in the spectra from the anti-CD4 fractions that were not present in the spectra from the IgG₁ fractions, such as the doubly charged ions at m/z 612.5 and 615.5 (Figure 4.8B). These two abundant components with high signal-to-noise ratios were easily seen above the background but many others were distinguishable from the singly charged background only by the spacing between their isotopes (1 m/z between isotopes of a singly charged ion, 0.5 m/z for a doubly charged ion and 0.33 m/z for a triply charged ion).

Following nanoESI-MS, the fractions containing peptides were then used for nanoESI-MS/MS analyses and LC-MS experiments. LC-MS/MS data was obtained from samples once an initial LC-MS experiment provided sufficient information to set up the LC-MS/MS experiment. MS/MS spectra then provided sequence information, which was entered into various search programs for interrogation of protein and translated oligonucleotide databases. Examples of nanoESI-MS/MS spectra and a search protocol are given in section 4.2.3. Sections 4.2.4 and 4.2.5 show examples from LC-MS and LC-MS/MS, respectively.
Figure 4.8 ESI-MS of fractions from ion-exchange chromatography of the eluate from the anti-CD4 affinity column after cleavage with trypsin.
A) Fraction 4, B) Fraction 5, C) Fraction 6, magnified by 6 from m/z 750
Figure 4.9 ESI-MS of fractions from ion-exchange chromatography of the eluate from the IgG control column after cleavage with trypsin.
A) Fraction 4, B) Fraction 5, C) Fraction 6
4.2.2 Summary of proteins identified

Table 4.2 shows a list of the precursors selected for MS/MS from the anti-CD4 samples that resulted in identification of the peptide sequence. The table does not include peptides that were also found in the IgG1 control samples or peptides identified from trypsin autolysis or known contaminants such as human keratins. Identifications that were uncertain are also excluded from Table 4.2.

Examination of the sequences of the peptides in Table 4.2 shows that all of the peptides contained at least one acidic amino acid (glutamic acid or aspartic acid) and all but two of the peptides contained two or more acidic residues. The two peptides that contained only one acidic amino acid were the peptide from the B cell receptor-associated protein and the peptide of mass 1353.1 Da from p56\textsuperscript{ck}.

The myosin regulatory light chain peptide NAFACamFDEEATGTIQEDYLR was sequenced from four different precursors. Two of these precursor ions, which are listed in Table 4.2, were of the expected peptide mass of 2349.0 Da. The other two ions were consistent with the loss of ammonia from the triply charged peptide and the addition of a sodium ion to the protonated peptide. Another precursor that was from the same tryptic peptide was a doubly charged ion at m/z 547.3. This corresponded to a fragment of the tryptic peptide with the sequence GTIQEDYLR, which indicated that the peptide bond between threonine and glycine had cleaved. This peptide is not included in Table 4.2.

Of the peptides sequenced from the anti-CD4 samples, a total of 10 proteins were identified. These proteins and their Swiss Prot accession numbers (A.N.) are listed in Table 4.3. The T cell surface glycoprotein, CD4, and the protein tyrosine kinase, p56\textsuperscript{ck}, were identified. Three different myosin proteins were identified, two from various light chains.
<table>
<thead>
<tr>
<th>Peptide Mass</th>
<th>m/z</th>
<th>Charge</th>
<th>Sequence</th>
<th>Residues</th>
<th>Protein</th>
<th>Method of Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1789.8 (1789.9)</td>
<td>895.9</td>
<td>2</td>
<td>SYELPDGQVITIGNER</td>
<td>239-254</td>
<td>actin, cytoplasmic 1 or 2 (beta-actin or gamma-actin)</td>
<td>nanoES</td>
</tr>
<tr>
<td>1085.0 (1084.6)</td>
<td>543.5</td>
<td>2</td>
<td>EIIDLVLDR</td>
<td>113-121</td>
<td>alpha tubulin</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1701 (1700.9)</td>
<td>851.5</td>
<td>2</td>
<td>AVFVDLEPTVIDEVR</td>
<td>65-79</td>
<td>alpha tubulin</td>
<td>LC MS/MS</td>
</tr>
<tr>
<td>2005.6 (2006.9)</td>
<td>1003.8</td>
<td>2</td>
<td>TIGGGDDSNTFFSETGAGK</td>
<td>41-60</td>
<td>alpha tubulin</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1723 (1722.9)</td>
<td>862.5</td>
<td>2</td>
<td>IPWFQYPPIYDIR</td>
<td>72-84</td>
<td>B cell receptor associated protein</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1531.7 (1531.7)</td>
<td>766.9</td>
<td>2</td>
<td>LTGSgelWWWQAER</td>
<td>232-244</td>
<td>CD4</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1956 (1956.0)</td>
<td>979</td>
<td>2</td>
<td>EGEQVEFSFPLAFTVEK</td>
<td>215-231</td>
<td>CD4</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>2492.3 (2492.2)</td>
<td>831.8</td>
<td>3</td>
<td>DFSALESQLODTQELLQENR</td>
<td>588-608</td>
<td>myosin heavy chain nonmuscle type A</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>2493 (2492.2)</td>
<td>1246.5</td>
<td>2</td>
<td>DFSALESQLODTQELLQENR</td>
<td>588-608</td>
<td>myosin heavy chain nonmuscle type A</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>2499.8 (2500.2)</td>
<td>1250.9</td>
<td>2</td>
<td>LQQEQLQAEHELCAEELR</td>
<td>883-903</td>
<td>myosin heavy chain nonmuscle type A</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1961.2 (1960.9)</td>
<td>981.6</td>
<td>2</td>
<td>TQLEELEDELQATEDAK</td>
<td>826-842</td>
<td>myosin heavy chain nonmuscle type A or B</td>
<td>LC MS/MS</td>
</tr>
<tr>
<td>1233.6 (1232.6)</td>
<td>617.8</td>
<td>2</td>
<td>EGNGTVMGAER</td>
<td>98-109</td>
<td>myosin light chain alkali</td>
<td>LC MS/MS</td>
</tr>
<tr>
<td>1531.7 (1531.6)</td>
<td>766.9</td>
<td>2</td>
<td>Ac-CamDFTEDQTAEFK</td>
<td>1-12</td>
<td>myosin light chain alkali</td>
<td>nanoES MS/MS and LC-MS/MS</td>
</tr>
<tr>
<td>1543.7 (1543.7)</td>
<td>772.9</td>
<td>2</td>
<td>DQGYEDYVEGLR</td>
<td>81-93</td>
<td>myosin light chain alkali</td>
<td>nanoES MS/MS and LC-MS/MS</td>
</tr>
<tr>
<td>1887.9 (1887)</td>
<td>630.3</td>
<td>3</td>
<td>VLDFEHFLMLQTVAK</td>
<td>63-78</td>
<td>myosin light chain alkali, (other components present)</td>
<td>nanoES MS/MS</td>
</tr>
</tbody>
</table>

Table 4.2 Summary of identified peptides and the method used to obtain their identification
<table>
<thead>
<tr>
<th>Peptide Mass (m/z)</th>
<th>m/z</th>
<th>Charge</th>
<th>Sequence</th>
<th>Residues</th>
<th>Protein</th>
<th>Method of Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1228.6 (1227.6)</td>
<td>615.3</td>
<td>2</td>
<td>LNGTDPEDVIR</td>
<td>92-102</td>
<td>myosin regulatory light chain</td>
<td>nanoES MS/MS and LC-MS/MS</td>
</tr>
<tr>
<td>2349 (2349.0)</td>
<td>1175.5</td>
<td>2</td>
<td>NAFACamFDEEATGTDQEDYLR</td>
<td>103-122</td>
<td>myosin regulatory light chain</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>2349.1 (2349.0)</td>
<td>784.03</td>
<td>3</td>
<td>NAFACamFDEEATGTDQEDYLR</td>
<td>103-122</td>
<td>myosin regulatory light chain</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1222.8 (1222.6)</td>
<td>612.4</td>
<td>2</td>
<td>LIEDNEYTAR</td>
<td>387-396</td>
<td>p56lck</td>
<td>nanoES MS/MS and LC-MS/MS</td>
</tr>
<tr>
<td>1274.6 (1274.6)</td>
<td>638.3</td>
<td>2</td>
<td>ILEQSGWVK</td>
<td>89-98</td>
<td>p56lck</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1353.1 (1351.7)</td>
<td>677.6</td>
<td>2</td>
<td>NLDNGFYISPRI</td>
<td>184-195</td>
<td>p56lck</td>
<td>nanoES MS/MS and LC-MS/MS</td>
</tr>
<tr>
<td>1455.6 (1455.7)</td>
<td>728.8</td>
<td>2</td>
<td>ESESTAGFSLSVR</td>
<td>154-167</td>
<td>p56lck</td>
<td>LC MS/MS</td>
</tr>
<tr>
<td>1463.6 (1463.7)</td>
<td>732.8</td>
<td>2</td>
<td>ANSLEPEWFFK</td>
<td>118-129</td>
<td>p56lck</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>1693.5 (1692.8)</td>
<td>565.5</td>
<td>3</td>
<td>ERPERPTFDYLR</td>
<td>478-490</td>
<td>p56lck</td>
<td>nanoES MS/MS</td>
</tr>
<tr>
<td>2056.4 (2055.9)</td>
<td>1029</td>
<td>2</td>
<td>SVLDFEFTATEGQYQQPOP</td>
<td>491-508</td>
<td>p56lck</td>
<td>nanoES MS/MS and LC-MS/MS</td>
</tr>
<tr>
<td>1359.7 (1359.7)</td>
<td>680.9</td>
<td>2</td>
<td>EGDVTLLESER</td>
<td>52-63</td>
<td>40S ribosomal protein</td>
<td>LC MS/MS</td>
</tr>
<tr>
<td>1416.7 (1416.7)</td>
<td>709.4</td>
<td>2</td>
<td>ILDSVGEADDLR</td>
<td>26-38</td>
<td>60S acidic ribosomal protein P2</td>
<td>LC MS/MS</td>
</tr>
</tbody>
</table>

Table 4.2 (continued) Summary of identified peptides and the method used to obtain their identification.
Two ribosomal proteins were identified, a 40S subunit and a 60S subunit, but only one peptide from each was sequenced. Alpha tubulin was identified from three separate peptides and each of these peptides matched a human alpha tubulin in Swiss Prot that was brain specific (P04687). Three mouse tubulin proteins also contained each of the peptides and they were from alpha chains 1, 2 and 6 (P02551, P05213 and P05216). Translated oligonucleotide sequences from various human tissues also contained each of the three peptides and all of these translated sequences were from alpha 1 tubulin (such as AAC31959 and AAH06468). A single peptide from an actin protein and a single peptide from the B cell receptor-associated protein were also sequenced.

<table>
<thead>
<tr>
<th>Protein identification</th>
<th>peptides</th>
<th>Swiss Prot A.N.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actin (α or γ)</td>
<td>1</td>
<td>P02570 or P02571</td>
</tr>
<tr>
<td>BCR-associating protein</td>
<td>1</td>
<td>Q99623</td>
</tr>
<tr>
<td>CD4</td>
<td>2</td>
<td>P01730</td>
</tr>
<tr>
<td>Myosin heavy chain type A</td>
<td>3</td>
<td>P35579</td>
</tr>
<tr>
<td>Myosin light chain alkali</td>
<td>4</td>
<td>P16475</td>
</tr>
<tr>
<td>Myosin regulatory light chain</td>
<td>2</td>
<td>P19105</td>
</tr>
<tr>
<td>p56^{lck}</td>
<td>7</td>
<td>P06239</td>
</tr>
<tr>
<td>40S Ribosomal protein</td>
<td>1</td>
<td>P25112</td>
</tr>
<tr>
<td>60S Acidic Ribosomal protein P2</td>
<td>1</td>
<td>P05387</td>
</tr>
<tr>
<td>Tubulin alpha-1 chain</td>
<td>3</td>
<td>P04687</td>
</tr>
</tbody>
</table>

Table 4.3 Summary of identified proteins with the total number of peptides sequenced from each of the proteins. The Swiss Prot accession number (A.N.) is listed in the final column.
A number of MS/MS spectra (either from nanoESI-MS/MS or LC-MS/MS) from precursor ions that were apparent in the anti-CD4 samples but not in the IgG\textsubscript{1} control samples did not enable identification of a peptide that could be assigned to a known protein. A summary of these unidentified precursor ions is given in Table 4.4. Some of these spectra (e.g. the peptide of molecular mass 1764.8 Da) showed clear fragmentation but the information entered for database searching did not result in any retrieved sequences. In other cases (e.g. the peptide of mass 1828.9 Da) sequences were retrieved from the searches but they did not match the full MS/MS spectrum. Finally, a number of spectra, particularly those from LC-MS/MS, either did not contain sufficient product ions to establish information with which to search the databases, or alternatively, the MS/MS spectra was sufficient to generate a tag but when a search retrieved a sequence (e.g. peptide of molecular mass 1365.6 Da), confirmation of the sequence was difficult, if not impossible. One curious result was for MS/MS of the doubly charged ion at m/z 877.1, which resulted in a very informative spectrum but database searching did not reveal a sequence that matched the data (Table 4.4). \textit{De novo} sequencing of the peptide was attempted but there were a number of combinations of residues leading to ambiguity of the full sequence. These are denoted by the square brackets in Table 4.4. In the first set of brackets either two glycine residues or one asparagine residue agreed with the MS/MS spectrum. The second set of brackets shows that either combination of aspartic acid and valine (in any order) or leucine/isoleucine and threonine (in any order) was possible. "Sequence-only" searching with any of these combinations did not result in identification of the protein.
<table>
<thead>
<tr>
<th>Peptide Mass</th>
<th>Precursor m/z</th>
<th>Charge</th>
<th>Sequence Tags</th>
<th>Possible Reasons not Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1002.7</td>
<td>502.3</td>
<td>2</td>
<td>(597.5)LD(825.6) (175.1)LLL(514.4)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1094.8</td>
<td>548.4</td>
<td>2</td>
<td>(635.4)EAF(982.6) (764.5)AF(982.6)</td>
<td>precursor mass possibly incorrect (incorrect charge on ion), modified peptide, more than one species (both m/z 147 and 175 present), dominant ions that could not fit into tags or prospective sequences</td>
</tr>
<tr>
<td>1120.6</td>
<td>561.32</td>
<td>2</td>
<td>(661.3)L(773.5)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1156.5</td>
<td>579.3</td>
<td>2</td>
<td>(573.4)DADT(975.5)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1183.6</td>
<td>592.8</td>
<td>2</td>
<td>(614.4)DLE(971.5)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1187.6</td>
<td>595.3</td>
<td>2</td>
<td>(961.6)D(1076.5) C-terminal K</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1257.6</td>
<td>630.8</td>
<td>2</td>
<td>(423.2)ELYNF(1089.6)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1265.6</td>
<td>633.8</td>
<td>2</td>
<td>(980.5)A(1051.6)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1365.6</td>
<td>683.8</td>
<td>2</td>
<td>(1008.5)GS(1152.6)</td>
<td>LC-MS/MS therefore minimal data, insufficient data to confirm or rule-out possible sequences but likely match ITSGPFEPDLYK</td>
</tr>
<tr>
<td>1390.6</td>
<td>696.3</td>
<td>2</td>
<td>(201.1)E(330.2) (849.5)TL(1063.5) (849.5)TN(1064.5)</td>
<td>LC-MS/MS therefore minimal data, insufficient data to confirm or rule-out possible sequences such as ISELTSELTDER</td>
</tr>
<tr>
<td>1498.8</td>
<td>750.4</td>
<td>2</td>
<td>(175.1)H(312.2) (765.5)S(852.6)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
</tbody>
</table>

Table 4.4 Summary of unidentified precursors including possible sequence tags or sequence only search options and possible reasons as to why the peptides were not identified
<table>
<thead>
<tr>
<th>Peptide Mass</th>
<th>Precursor m/z</th>
<th>Charge</th>
<th>Sequence Tags</th>
<th>Possible Reasons not Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1507.9</td>
<td>754.94</td>
<td>2</td>
<td>(1121.8)D(1236.8)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1512.8</td>
<td>757.4</td>
<td>2</td>
<td>(812.5)DESDV(1357.8)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1574.8</td>
<td>788.4</td>
<td>2</td>
<td>(1221.6)Q(1349.8)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1608.9</td>
<td>805.5</td>
<td>2</td>
<td>(1097.6)Q(1225.7)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1623.9</td>
<td>812.4</td>
<td>2</td>
<td>(969.6)VD(1183.7)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>1752.2</td>
<td>877.1</td>
<td>2</td>
<td>(1072.8)NL(1299.0)</td>
<td>possibly modified peptide, A_{2}/B_{1} 183.2/211.2 (I/L)/P, de novo sequencing LP[GGorN]ELNP[DVorLT]SEVEGLK may be not in database</td>
</tr>
<tr>
<td>1764.8</td>
<td>883.4</td>
<td>2</td>
<td>(933.5)DGE(1234.6)</td>
<td>modified peptide therefore incorrect masses, more than one peptide at precursor mass</td>
</tr>
<tr>
<td>1764.8</td>
<td>883.4</td>
<td>2</td>
<td>(175.1)LEQM(676.3)</td>
<td>modified peptide therefore mass incorrect, numerous possibilities for tags, difficult to confirm sequences</td>
</tr>
<tr>
<td>1808</td>
<td>905.03</td>
<td>2</td>
<td>(991.6)ENGE(1420.7)</td>
<td>part of the sequence matches light chain myosin but the remainder does not match</td>
</tr>
<tr>
<td>1828.9</td>
<td>915.45</td>
<td>2</td>
<td>(736.4)DEYTQGD(1544.8)</td>
<td>spectrum dominated by two doubly charged fragments, insufficient information, possibly modified peptide</td>
</tr>
<tr>
<td>2019</td>
<td>674</td>
<td>3</td>
<td>(1566.8)DP(1778.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1566.8)PD(1778.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1566.8)I/LV(1778.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1566.8)II/L(1778.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1566.8)YG(395.2)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4 (continued)
<table>
<thead>
<tr>
<th>Peptide Mass</th>
<th>Precursor m/z</th>
<th>Charge</th>
<th>Sequence Tags</th>
<th>Possible Reasons not Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>2030.9</td>
<td>1016.5</td>
<td>2</td>
<td>(175.1) L(288.2) (288.2)EA(488.3)</td>
<td>LC-MS/MS therefore minimal data, modified peptide therefore incorrect mass, insufficient data to confirm or rule-out possible sequences</td>
</tr>
<tr>
<td>2046.04</td>
<td>683.02</td>
<td>3</td>
<td>sequence only ELDHD (719.54)DH(1598.0) (1499)DL(1727) (1500)NL(1727) A2/B2 m/z 292/320</td>
<td>doubly charged fragments dominate and complicate spectrum above precursor, modified peptide,</td>
</tr>
<tr>
<td>2062</td>
<td>688.3</td>
<td>3</td>
<td>(1566.7)DP(1778.8) (1566.7)PD(1778.8) (1566.7)ILV(1778.8) (1566.7)VI/L(1778.8) (175.1)TN(390.2) (147.1)DF or FD(409.3)</td>
<td>related to precursor at m/z 674(3+), same tags for high mass range, insufficient information for identification</td>
</tr>
<tr>
<td>2224</td>
<td>742.3</td>
<td>3</td>
<td>too many to list</td>
<td>modified peptide therefore incorrect masses, more than one peptide at precursor mass</td>
</tr>
<tr>
<td>2441.1</td>
<td>1221.5</td>
<td>2</td>
<td>(1345.7)CamLY(1781.8) (1345.7)CamLL(1731.7) (824.3)L(927.4) (873.5)DE(1117.5)</td>
<td>modified peptide therefore incorrect masses, more than one peptide at precursor mass</td>
</tr>
<tr>
<td>2464.5</td>
<td>822.5</td>
<td>3</td>
<td>(1291.8)EOF(1695.9) sequence only LEGNEQF</td>
<td>appears to be an unusual autolysis product of trypsin therefore sequence tags did not reveal the identity</td>
</tr>
<tr>
<td>?</td>
<td>997.7</td>
<td>?</td>
<td>(1311.9)L(1424.5)</td>
<td>difficult to induce fragmentation, unsure of charge on precursor ion</td>
</tr>
<tr>
<td>?</td>
<td>1000.7</td>
<td>?</td>
<td>(1044.9)NF(1306)</td>
<td>few fragments, unsure of charge on precursor ion, insufficient information</td>
</tr>
<tr>
<td>1443.0</td>
<td>722.5</td>
<td>2, 3</td>
<td>(883.4)LSTF(1331.6) (996.4)ST(1184.5) (930.4)E(1059.5)</td>
<td>possibly contains peptide from prohibitin P35232 IFTSIGEDYDER (1443.7 Da), complicated due to more than one peptide, incomplete sequence information for full sequence assignment</td>
</tr>
<tr>
<td>2656.5 (3+)</td>
<td>886.5</td>
<td>≥ 3</td>
<td>(2130.1)WD(2431.21)</td>
<td>unsure of charge on precursor ion, multiply charged ions above precursor complicate spectrum</td>
</tr>
</tbody>
</table>

Table 4.4 (continued)
Other difficulties in assignment of the MS/MS spectra arose when the charge on the ion could not be established. With ions that had low signal-to-noise ratios the isotopic resolution was often obscured by the background noise. The overlap of two species also complicated determination of the charge and, if both species were subjected to MS/MS, the spectrum contained product ions from both precursors.

Examples of identifications from Table 4.2 are presented in the following sections and illustrate the applicability of different search techniques for protein identification. An example of an unidentified peptide is also included and illustrates the limitations of database searching for the identification of some proteins.

### 4.2.3 MS/MS and database searching

Multiply charged ions that were apparent in the spectra of fractions from the anti-CD4 preparation but not in the spectra of fractions from the IgG₁ preparation were selected for further analysis by MS/MS. Figure 4.10 shows the MS/MS spectrum obtained from fragmentation of a doubly charged ion at m/z 1004.5. The region above the precursor, which is the base peak in the spectrum, shows a series of singly charged product ions. An example of a possible sequence tag is indicated by the arrows and labelled amino acids in Figure 4.10. The four ions selected form a tag that could be used to search protein databases in a number of ways. One possibility is to enter the four product ion masses along with the mass of the precursor into a search program such as Protein Prospector (http://prospector.ucsf.edu, (Clauser et al., 1999)). Alternatively, the mass of the ions at either end of the sequence with the amino acid sequence and the precursor mass can be entered into the search program Peptide Search (also known as PepSea) (Mann and Wilm,
1994) (http://pepsea.protana.com/). The example in Figure 4.10 generated the sequence tag \((1044.6)\text{NFS}(1393.8)\) and the mass of the precursor was 2000.98 Da. Figure 4.11 shows the parameters in Peptide Search that are entered for searching the Swiss Prot database (http://au.expasy.org/sprot) (as explained in Chapter 3). Sequence tags were searched as both Y ion series and B ion series and missed cleavages by trypsin were also allowed. A 2 Da error tolerance in the peptide mass was typically used but was decreased in searches that retrieved large numbers of proteins.

A portion of the results from the search parameters displayed in Figure 4.11 is shown in Figure 4.12. Only 13 of the 34 matches retrieved are displayed but the remaining matches also corresponded to the peptide sequence shown in the left hand column of Figure 4.12. When this sequence was compared to the MS/MS spectrum from which the sequence tag was obtained, a clear Y ion series was apparent, as well as other sequence ions from the retrieved sequence. Figure 4.13 shows the MS/MS spectrum with the Y ion series of the retrieved sequence labelled. The sequence matched the MS/MS data very well and alpha tubulin was therefore identified in the sample. Of the alpha tubulin proteins retrieved by Peptide Search from the Swiss Prot database that were not translated oligonucleotide sequences, only one corresponded to a human alpha tubulin (accession number P04687). This protein was described as human tubulin alpha-1 chain but was brain specific. The T cells used were from a human cell line and were not from brain tissue. The translated nucleotide sequences identified tubulin alpha 1 (accession numbers AAH06468, AAC31959, AAH00431, AAH06481 and AAH08330) and alpha 6 (accession number Q9BQE3) from various different tissues for example muscle, bone marrow, and skin.
Figure 4.10 ESI-MS/MS spectrum of the precursor ion at m/z 1004.5 from preparation from the anti-CD4 column. The peptide sequence tag (1044.6)NFS(1393.8) was obtained from the labelled peaks.

<table>
<thead>
<tr>
<th>Search parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence tag</td>
</tr>
<tr>
<td>Protein mass range</td>
</tr>
<tr>
<td>Cleavage agent</td>
</tr>
<tr>
<td>Peptide mass accuracy</td>
</tr>
<tr>
<td>Methionine is</td>
</tr>
<tr>
<td>Cysteine is</td>
</tr>
<tr>
<td>Peptide mass</td>
</tr>
<tr>
<td>Match regions</td>
</tr>
<tr>
<td>Search by</td>
</tr>
<tr>
<td>Allowed number of errors</td>
</tr>
<tr>
<td>N terminal specific</td>
</tr>
<tr>
<td>C terminal specific</td>
</tr>
</tbody>
</table>

Figure 4.11 Search parameters obtained from MS/MS of the precursor at m/z 1004.5 (Figure 4.10) entered into Peptide Search (http://pepsea.protana.com/)
## Search result

34 matches were found.

<table>
<thead>
<tr>
<th>Peptide Sequence matched/Peptide found</th>
<th>Mass [kDa]</th>
<th>Database accession</th>
<th>Protein Name</th>
<th>Digest</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.891</td>
<td>sprembl:Q94978</td>
<td>Q94978 ALPHA-2 TUBULIN.//:trembl</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.809</td>
<td>sprembl:Q94985</td>
<td>Q94985 ALPHA-1 TUBULIN.//:trembl</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.692</td>
<td>sprembl:Q26595</td>
<td>Q26595 ALPHA-TUBULIN.//:trembl</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.893</td>
<td>spremblnew:AAG15319</td>
<td>AAG15319 ALPHA TUBULIN.//:</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.685</td>
<td>spremblnew:AAG15365</td>
<td>AAG15365 ALPHA TUBULIN.//:</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.837</td>
<td>trembl:K00558</td>
<td>HSTUBAK_1 human alpha-tubulin mRN</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.564</td>
<td>trembl:X07045</td>
<td>XLTUBAG_1 product: &quot;alpha-tubulin</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.780</td>
<td>pironly:C24903</td>
<td>C24903 tubulin alpha-3 chain - C</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>43.388</td>
<td>pironly:S33517</td>
<td>S33517 tubulin alpha chain - mar</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.843</td>
<td>swissnew:P04687</td>
<td>TBA1_HUMAN TUBULIN ALPHA-1 CHAI</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.590</td>
<td>swissnew:P52273</td>
<td>TBA_BOMMO TUBULIN ALPHA CHAIN.</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.837</td>
<td>swissprot:P05209</td>
<td>TBA1_CRIGR TUBULIN ALPHA-1 CHAIN.</td>
<td>•</td>
</tr>
<tr>
<td>TIGGGDDSFNTFFSETGAGK</td>
<td>50.821</td>
<td>swissprot:P02551</td>
<td>TBA1_MOUSE TUBULIN ALPHA-1 CHAIN.</td>
<td>•</td>
</tr>
</tbody>
</table>

*Figure 4.12* Search results from input of the sequence tag (1044.6)NFS(1393.8) into Peptide Search (http://pepsea.protana.com/)
Figure 4.13 MS/MS spectrum of the precursor ion at m/z 1004.5 from preparation from the anti-CD4 column (as in Figure 4.10). The region above the precursor is enlarged by 5. The y ion series from the sequence TIGGGDSFNTFFSETGAGK is labelled.

Figure 4.14 MS/MS spectrum of ion at m/z 979.0 from anti-CD4 preparation. The main singly charged Y and B ions of the sequence EGEQVEFSFPLAFTVEK (CD4 T29) are shown.
Two further examples of MS/MS spectra that enabled unambiguous assignment of the peptide sequence from peptide sequence tags are shown in Figures 4.14 and 4.15. Figure 4.14 shows the MS/MS spectrum obtained from a doubly charged precursor at m/z 979.0. The clear series above the precursor enabled a sequence tag to be generated (i.e. (1051.7)SFE(1414.9)) and the retrieved sequence EGEQVEFSFPLAFTVEK closely matched the MS/MS data. This peptide corresponded to a tryptic peptide from the T-cell surface glycoprotein CD4. Figure 4.15 shows an MS/MS spectrum of a doubly charged precursor of m/z 612.4. The matching sequence LIEDNEYTAR was found in a number of protein tyrosine kinases, including p56\textsuperscript{Lck}.

Figure 4.16 shows a more complicated MS/MS spectrum, which was from the precursor at m/z 766.85. MS/MS of the same precursor from an earlier sample had shown a much simpler spectrum (data not shown) and the sequence was identified as LTGSGELWWQAER from the T-cell surface glycoprotein CD4. The Y ion series of this sequence is labelled in Figure 4.16 (red labels). With this prior information and the clearly evident product ions that did not belong to the CD4 peptide, a second partial peptide sequence was established. The sequence between the ions at m/z 595.32 to 1331.6, QDETFD, was searched without any mass information, i.e. a “sequence-only” search. If the series was assumed to be Y ions (i.e. using the sequence DFTEDQ) 11 matches resulted from the Swiss Prot database being searched with the Peptide Search “sequence-only” option. All of the matches corresponded to myosin light chain alkali and the N-terminal tryptic peptide CamDFTEDQTAEFK (where Cam represents a carboxyamidomethylated cysteine) matched the MS/MS data. The peptide mass (1531.7 Da) did not, however match the theoretical mass of the tryptic peptide sequence, which would be expected to have
Chapter 4: Identification of CD4 and co-associating proteins

**Figure 4.15** MS/MS spectrum of the precursor ion at m/z 612.4 from the anti-CD4 preparation. The fragment ions labelled are from the sequence LIEDNEYTAR.

**Figure 4.16** MS/MS spectrum of ion at m/z 766.85 from fraction 5 of anti-CD4 preparation. Two sequences apparent: Red CD4: LTGSGELWWQAER, Blue Myosin: CamDFTEDQTAEFK
a molecular mass of 1489.6 Da (with the modified cysteine). The difference (42 Da) can be attributed to acetylation of the N-terminal residue. The Swiss Prot entry for this protein confirmed that the protein was N-terminally acetylated (accession number P16475). To check that the same product ions from the MS/MS spectrum in Figure 4.16 could not be attributed to B ions, the sequence stretch was entered into the search program in the reverse order (QDETFD). No matches were returned for this search. Using a sequence tag search did not identify this myosin peptide, as the peptide mass calculated for the sequence is 1489.6 Da. With a mass of 1531.7 Da, no matches for myosin were retrieved with the sequence tag (838.5)ET(1068.8). With the peptide mass set to 1489.6 Da the myosin sequence was retrieved with the same tag. This example demonstrates an advantage of using "sequence-only" searching.

Figure 4.17 shows an MS/MS spectrum from the precursor ion at m/z 862.5. A sequence tag from the ions above the precursor in this spectrum was used to identify the sequence of this peptide and the resulting match, IPWFQYPIIYDIR, corresponded to a tryptic peptide from a B cell receptor-associated protein (Swiss Prot accession number Q99623). The Y ion series from this peptide was apparent in the MS/MS spectrum and the ions are labelled in Figure 4.17A. Many of the abundant product ions below the precursor did not, however, correspond to the Y or B ion series, but can be attributed to the formation of internal fragment ions from cleavage at the proline residue and a second peptide bond in the backbone chain (see Section 4.3.4). An enlarged portion of the spectrum is shown in Figure 4.17B and the internal fragments are labelled with the amino acid sequence in the fragments (PW, PWF, PWFQ, PIYD, PIYDI and PWFQY). The identification of these product ions further confirmed the peptide sequence but its presence in a T cell preparation
Figure 4.17 MS/MS spectrum of the precursor ion at m/z 862.5 from fraction 6 of an anti-CD4 preparation. The Y and B ions labelled are from the sequence IPWFQYP11YDIR. A) Full m/z range (50-1600) B) Enlarged region from m/z 77-750
was unexpected. The B cell receptor-associated protein sequence was submitted for BLAST searching to retrieve any homologous proteins that may have not specifically associated with the B cell receptor. The proteins retrieved that contained the tryptic sequence from Figure 4.17 were all from various forms of the B cell receptor-associated protein, (also known as prohibitin) (data not shown). A “sequence-only” search (using MS Pattern, from Protein Prospector) of the non-redundant database of human ESTs (updated 01.08.01) retrieved 295 matches when the sequence was searched with unknown specificity of leucine/isoleucine residues. The same search with the Genpept database (updated 01.08.01) showed four translated proteins that were B cell receptor-associated proteins and one that was a repressor of estrogen activity (accession number AF150962). There were no other peptides sequenced from the B cell receptor-associated protein.

4.2.4 LC-MS

Separation of the peptides within each of the three fractions was performed using capillary LC-MS. Figure 4.18 shows the LC-MS profile of fraction 5 from an anti-CD4 preparation. The total ion current (TIC) chromatogram (Figure 4.18A) did not provide detailed information on the elution of peptides from the column. The base peak intensity (BPI) chromatogram (Figure 4.18B) indicated that elution of components from the column became apparent after approximately 27 minutes in this separation. Investigation of the spectra from this separation showed that elution of the multiply charged components was complete by approximately 40 minutes. From the data obtained in the LC-MS experiment, mass chromatograms for particular masses or mass ranges can be generated. Figure 4.18C shows the mass chromatogram for m/z 612.4, which was previously identified as a tryptic peptide from the protein tyrosine kinase, p56\(^\text{lck}\). A well-defined peak at a retention time of
Figure 4.18 Chromatograms from LCMS of fraction 5 from an anti-CD4 preparation. A) Total ion current chromatogram, B) Base peak intensity chromatogram, C) Mass chromatogram of m/z 612.3, D) Mass chromatogram of m/z 615.3 E) Mass chromatogram of m/z 766.9.
29.4 minutes was apparent in the chromatogram. The mass chromatogram of m/z 615.3 (Figure 4.18D) also shows a well-defined peak, with a retention time of 32.1 minutes. The chromatogram in Figure 4.18E is that of m/z 766.9. The chromatogram of m/z 766.9 revealed two separate peaks at retention times of 34.6 minutes and 36.1 minutes. The signal-to-noise ratio of the peaks in the chromatogram of m/z 766.9 was lower than the chromatogram of m/z 615.3 and the total ion count from the chromatogram of m/z 766.9 was significantly lower than that of m/z 615.3 (102 counts versus 685 counts).

The summed mass spectra from the mass chromatograms in Figure 4.18C-E are shown in Figure 4.19. The spectra summed from retention times indicated by the chromatograms of each of m/z 612.4 and m/z 615.3 (Figure 4.19A and B) showed the base peak in each spectrum corresponded to the m/z from which the chromatogram was generated. Both these ions were doubly charged, as was clear from their isotopic resolution (insets Figure 4.19A and B). The two separate peaks in the mass chromatogram of m/z 766.9 were summed and the resulting spectra are shown in Figure 4.21C and D. At a retention time of 34.6 minutes (Figure 4.21C), the spectrum showed the base peak was a doubly charged ion at m/z 679.3 which did not appear to be present in the LC-MS data for fraction 5 from the IgG1 preparation (data not shown). The ion at m/z 766.9 was also clearly evident in the spectrum and the isotopic resolution revealed the ion to be doubly charged (inset Figure 4.21C). The spectrum at a retention time of 36.1 minutes showed the base peak to be a doubly charged ion at m/z 971.1, which again was not apparent in the LC-MS acquisition of the equivalent IgG1 control sample. This peptide was later identified to be a fragment from the trypsin used in the digest (data not shown). The ion at m/z 766.9 had a low signal-to-noise ratio but was still discernible and was clearly doubly charged.
Figure 4.19 Summed spectra from mass chromatograms shown in Figure 4.18. Average retention times A) 29.4 min B) 32.1 min C) 34.6 min D) 36.1 min
Figures 4.20, 4.21 and 4.22 show the mass chromatograms of m/z 615.3 (A) and m/z 766.9 (B) from LC-MS of fractions 4, 5 and 6 of an anti-CD4 preparation. The chromatograms of m/z 615.3 from each of the three fractions showed a well-defined peak. The summed spectra from this peak in each of the different fractions showed that the ion at m/z 615.3 was doubly charged (data not shown for fractions 4 and 6, Figure 4.19B for fraction 5). In contrast, the chromatograms of m/z 766.9 (Figures 4.20B, 4.21B and 4.22B) showed less well-defined peaks. As seen in Figure 4.18E, two peaks were apparent in the mass chromatogram of m/z 766.9 from LC-MS of fraction 5 and the summed spectra for each of these peaks showed the ions at m/z 766.9 to be doubly charged (Figure 4.19 C-D). The chromatogram of m/z 766.9 in fraction 4 showed a single peak at a retention time of approximately 37.7 minutes (Figure 4.20B). The summed spectra from the elution of m/z 766.9 resulted in a spectrum in which the base peak was a singly charged ion at m/z 819.4 (data not shown). This spectrum showed that the ion at m/z 766.9 was singly charged, and therefore unlikely to be a tryptic peptide. Similarly, the chromatogram of m/z 766.9 from fraction 6 revealed only one peak at a retention time of approximately 34.2 minutes. The summed spectra from the chromatogram peak showed the ion at m/z 766.9 was not very abundant but was clearly doubly charged (data not shown). Many other components were present in the LC-MS data of fractions 4, 5 and 6 from the anti-CD4 preparation that were not apparent in the equivalent LC-MS data from the IgG\textsubscript{i} preparation. LC-MS/MS was therefore performed to obtain more information, particularly on peptides that were not observed in the nanoESI experiments.
Figure 4.20 Mass chromatograms of A) m/z 615.3 and B) m/z 766.9 from LC-MS of fraction 4 from an anti-CD4 preparation

Figure 4.21 Mass chromatograms of A) m/z 615.3 and B) m/z 766.9 from LC-MS of fraction 5 from an anti-CD4 preparation

Figure 4.22 Mass chromatograms of A) m/z 615.3 and B) m/z 766.9 from LC-MS of fraction 6 from an anti-CD4 preparation
4.2.5 LC-MSMS and function switching

To obtain MS/MS information from peptides as they eluted from the RP-HPLC column, the mass spectrometer was set to automatically detect potential precursors in MS mode and switch to MS/MS when particular criteria were met. After acquisition of MS/MS data for a specified time, the mass spectrometer switched back to MS mode to check for newly eluting components. The ability to perform this data-dependent switching during an LC-MS experiment allows a large amount of information to be generated from one experiment. The data obtained from the MS/MS spectra are highly dependent on the switch criteria and on the parameters used for acquisition of the MS/MS data. Prior knowledge of the samples from previous LC-MS runs without function switching is helpful for establishing ideal parameters for a function switching experiment. The ion counts from single spectra in the initial LC-MS experiments were used as an indicator of ideal thresholds for detection of suitable precursors. Thresholds between 3 and 5 ion counts were used for the LC-MS/MS acquisitions in this work.

Figure 4.23 shows the MS/MS spectrum of a precursor ion at m/z 766.9, which was acquired by data-dependent switching during an LC-MS separation of fraction 5 from an anti-CD4 preparation. The spectrum in Figure 4.23 is summed from 8 separate spectra, each of which was acquired for 1 s. The data-dependent switching software also enables MS/MS acquisitions on multiple components that elute from the column at the same time. In between the acquisition of each of the 8 spectra of the precursor at m/z 766.9, a second precursor at m/z 679.4 was selected and subjected to MS/MS. The instrument alternated between MS/MS of m/z 766.9 and m/z 679.4. A previous LC-MS experiment had shown that these two components eluted together (Figure 4.19C). The MS/MS spectrum from the
precursor at m/z 679.4 (data not shown) enabled this component to be identified as a tryptic peptide of human keratin. The MS/MS spectrum of the doubly charged ion at 766.9 (Figure 4.23) is very different to the spectrum acquired using nanoESI MS/MS, owing to its much lower ion count. Comparing this spectrum to that in Figure 4.16, however, revealed that the ion series above the precursor matched some of the ions in Figure 4.16. The ions at m/z 838.5, 967.4, 1068.6 and 1331.7 corresponded to Y ions of the myosin light chain alkali N-terminal tryptic peptide CamDFTEDQTAEFK. The ions from the CD4 tryptic peptide were not present in the spectrum shown in Figure 4.23 as the CD4 peptide had a different retention time and therefore the MS/MS spectrum of the myosin light chain alkali peptide was much less complicated.

Figure 4.24 shows the MS/MS spectrum from a precursor ion at m/z 851.5 acquired during an LC-MS experiment of fraction 4 from an anti-CD4 preparation. The spectrum was generated from 8 separate scans and the retention time was approximately 40.4 minutes. The sequence tag (928.6)ELDV(1384.9) produced 49 matches from Swiss Prot (when the ions were assumed to be Y ions) and all of the matches were for the same sequence from alpha tubulin (AVFVDLEPTVIDEVR). The sequence closely matched the MS/MS data. This peptide had not been identified from the sample without prior separation of peptides by LC-MS.
**Figure 4.23** MS/MS spectrum of a precursor ion at m/z 767 acquired during LC-MS experiment of fraction 5 from anti-CD4 preparation.

**Figure 4.24** MS/MS spectrum of a precursor ion at m/z 851.5 acquired during LC-MS experiment of fraction 4 from anti-CD4 preparation.
The MS/MS spectrum of a precursor ion at m/z 719.9 is shown in Figure 4.25 and is the result of the sum of 9 separate spectra, each with an integration time of 1 s. The precursor was selected by data-dependent function switching during an LC-MS experiment of fraction 5 from an anti-CD4 preparation. The doubly charged ion eluted at a retention time of approximately 31.2 minutes and the measured peptide mass was 1437.8 Da. This peptide had not been identified previously by MS/MS of the unseparated sample. The sequence tags generated from the MS/MS data in Figure 4.25 did not retrieve any sequences from the protein databases that matched the MS/MS data. The product ions in the spectrum enabled the assignment of stretches of sequence but “sequence-only” searches did not result in identification of the peptide. De novo sequencing of the peptide resulted in two possible sequences but could not distinguish between leucine and isoleucine. These sequences were FDGE[IL]VD[IL]NFDR and FNGE[IL]VD[IL]NFDR. The difference between the two likely sequences was the possibility of either an asparagine or an aspartic acid at position 2. The peptide mass matched the former sequence very well, whilst the latter sequence differed from the measured peptide mass by approximately 1 Da (i.e. the difference between D and N). Further searching with these proposed sequences, allowing for the possibility of incorrect residues, did not result in the identification of a tryptic peptide that matched the MS/MS data.

Figure 4.26 shows an MS/MS spectrum acquired during LC-MS of fraction 5 from an anti-CD4 preparation. The precursor selected was m/z 896.46, but closer examination of the MS spectrum revealed that the ion selected for MS/MS was the doubly charged peptide ion containing one $^{13}$C atom. The peptide ion containing only $^{12}$C atoms (i.e. the monoisotopic ion) was the ion at m/z 895.9. The monoisotopic ion gave a measured peptide mass of 1789.8 Da. The spectrum did not show a clear series of ions above the precursor but there was sufficient information for sequence tag searching. An important
The feature of the spectrum was the intense $A_2$ and $B_2$ ion pair at $m/z$ 223.1 and 251.1, respectively. There are a number of possible amino acid combinations that result in a $B_2$ ion of $m/z$ 251. These combinations are SY, YS, CF, FC, LH, HL, IH and HI. If the cysteine residues are modified, as in these samples, the combinations CF and FC no longer give a $B_2$ ion of 251 and are therefore excluded. The two N-terminal residues were therefore either serine and tyrosine or histidine with leucine or isoleucine, but the order of these residues is unknown from the ion pair. The ion at $m/z$ 380.2 differs from the $B_2$ ion at $m/z$ 251.1 by 129, which corresponds to glutamic acid (E). The difference between 493.3 and 380.2 corresponds to leucine or isoleucine (L/I). Searching the Swiss Prot database using MS-Seq of Protein Prospector with the three ions at $m/z$ 251.1, 380.2 and 493.3 resulted in 132 retrieved proteins, all of which were different entries of the protein actin. The single peptide sequence from these proteins was SYELPDGQVITIGNER. Other ions from the spectrum in Figure 4.26 matched the retrieved sequence. The ion at $m/z$ 1299.8 corresponds to the second isotope (i.e. the isotope containing one $^{13}$C atom) of the $Y_{12}^-$ ion. The monoisotopic peak of the $Y_{12}^-$ ion was also visible at $m/z$ 1298.7. The ion at $m/z$ 1087.7 corresponded to the second isotope of the $Y_{10}^-$ ion. The presence of the second isotope rather than the monoisotopic ion is attributed to the selection of the precursor. As mentioned above, the second isotope of the precursor was chosen for MS/MS, rather than the monoisotopic ion. The remaining two abundant ions, $m/z$ 650.4 and 690.4 are also $Y$ ions. The ion at 650.4 is actually doubly charged, which can be seen from the smaller monoisotopic ion at $m/z$ 649.8. This corresponds to the doubly charged $Y_{12}^-$ ion. The ion at $m/z$ 690.4 corresponds to $Y_6^-$. Hence, even with what appears to be very little information, peptide identification can be achieved and confirmed with high confidence using this technique.
Figure 4.25 MS/MS spectrum of a precursor ion at m/z 719.9 from an LC-MS separation of fraction 5 from an anti-CD4 preparation.

Figure 4.26 MS/MS spectrum of a precursor ion at m/z 895.9 from an LC-MS separation of fraction 5 from an anti-CD4 preparation.
4.2.6 Preliminary results with a different anti-CD4 column

The development of a second anti-CD4 column and its use in the search for CD4-associating proteins has commenced. The antibody, OKT4, unlike Q425 used for previous preparations, does not require calcium for binding to CD4. Initial elutions from the OKT4 column used guanidine rather than the EDTA or glycine buffers used to elute proteins from the Q425 anti-CD4 column. As for previous samples, after enzymatic cleavage of the proteins in the eluates, the peptides were subjected to anion exchange chromatography (in this case performed by Oliver Bernhard, Millennium Institute, Westmead). Fractions 4-6 from the ion exchange chromatography of the eluate from the anti-CD4 column indicated the presence of CD4 (data not shown). After desalting the fractions with C18ZipTips™, the samples were analysed by nanoESI-MS. Figure 4.27 shows the ESI-MS spectra of fraction 4 from the IgG1 control digest and the OKT4 anti-CD4 digest. A number of multiply charged ions were present in both the spectra, and most of the ions in the anti-CD4 preparation (Figure 4.27A) were also present in the IgG1 control preparation (Figure 4.27B). The insets in Figure 4.27 show an enlarged region from m/z 750-755, which would be expected to contain a doubly charged CD4 peptide ion at m/z 766.9. The spectrum from the anti-CD4 preparation did not show any obvious doubly charged ion at m/z 766.9. To investigate whether a CD4 peptide could be detected by MS/MS, m/z 766.9 was selected for MS/MS from both the anti-CD4 sample and the IgG1 control sample. The resulting MS/MS spectra are shown in Figure 4.28. MS/MS of m/z 766.9 from the anti-CD4 sample showed clear fragmentation characteristic of the presence of a peptide (Figure 4.28A). The MS/MS data matched the sequence of the CD4 peptide of mass 1531.7 (LTSGELWWQAER) and the Y ion series is labelled in Figure 4.28A. The MS/MS
Figure 4.27 ESI-MS spectra of fraction 4 from ion exchange chromatography of eluate from columns after tryptic cleavage. A) digest from anti-CD4 column (new antibody, OKT4), B) digest from IgG control column. Insets show enlarged region from m/z 750-775
Figure 4.28 MS/MS spectra of ion-exchange column fraction 4 from digests of elutions from the anti-CD4 column (new antibody, OKT4) and the IgG control column. The quadrupole was set to transmit m/z 766.9. The region m/z 50-760 is magnified by 4 and the region m/z 770-1450 is magnified by 8.

A) Digest from anti-CD4 (OKT4) column, The Y ion series from the peptide LTGSGELWWQAER is labelled. B) Digest from IgG control column
spectrum of the same precursor from the IgG1 control sample (Figure 4.28B) did not show
the peptide from CD4.

Fractions 5 and 6 from ion exchange chromatography of both the IgG1 control digest
and the anti-CD4 digest were also analysed by ESI-MS (data not shown). A large number
of multiply charged species were apparent in the spectra but it was very difficult to
establish differences between the spectra. Further development with the elution protocol
from this column is therefore necessary.

4.2.7 Importance of acidic residues

As mentioned in section 4.2.2, each of the identified peptides contained at least one acidic
residue (Table 4.2). Anion exchange chromatography retains negatively charged
components and it was therefore crucial for these analyses that the proteins contained
peptides with acidic amino acids. To evaluate further the potential of the method and the
risk of removing peptides without acidic residues, the sequences of the identified proteins
were compared.

Figure 4.29 shows the sequence of the cell surface glycoprotein, CD4. The tryptic peptides
in the sequence are labelled from T1 through T63. The acidic amino acids are shown in
boxes and the two peptides that were identified by MS/MS are in bold (T29 and T30). T29
has four acidic amino acids and T30 has two acidic amino acids. The sequence of the
protein tyrosine kinase, p56\textsuperscript{ck}, is shown in Figure 4.30. The seven peptides that were
identified, T6, T8, T13, T17, T37, T46 and T47, are shown in bold. Only one of these
peptides did not contain two or more acidic residues. T17 has only one acidic residue,
aspartic acid at position 3.
<table>
<thead>
<tr>
<th>MNR</th>
<th>GVPFR</th>
<th>HLLLVLQLALLPAATQGK</th>
<th>K</th>
<th>VVLGK</th>
<th>K</th>
<th>GDTVE</th>
<th>LTCTASQK</th>
<th>K</th>
<th>SIQFHWK</th>
<th>NSNQIK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
<td>T4</td>
<td>T5</td>
<td>T6</td>
<td>T7</td>
<td>T8</td>
<td>T9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ILGNQGSFLTK</td>
<td>GPSK</td>
<td>LNDR</td>
<td>ADSR</td>
<td>R</td>
<td>SLWDQGNFPLI</td>
<td>K</td>
<td>NLK</td>
<td>IEDSDTYICEDQK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T11</td>
<td>T12</td>
<td>T13</td>
<td>T14</td>
<td>T15</td>
<td>T16</td>
<td>T17</td>
<td>T18</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EEVQIYLVFGLTANSEDTHLLQGQSLTTL</td>
<td>ESPPGSSPSVQCR</td>
<td>SPR</td>
<td>GK</td>
<td>NQGGKG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T20</td>
<td>T21</td>
<td>T22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T23</td>
<td>T24</td>
<td>T25</td>
<td>T26</td>
<td>T27</td>
<td>T28</td>
<td>T29</td>
<td>LQMGK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LTGSQELWWQAER</td>
<td>ASSSSK</td>
<td>SWITFDLK</td>
<td>NK</td>
<td>ETVSK</td>
<td>R</td>
<td>VTQDPK</td>
<td>LQMGK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T30</td>
<td>T31</td>
<td>T32</td>
<td>T33</td>
<td>T34</td>
<td>T35</td>
<td>T36</td>
<td>T37</td>
<td>T38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LPLHLTLQPALPQYAGSNGLTLALE</td>
<td>AK</td>
<td>TGK</td>
<td>LHQE</td>
<td>VNLVVMR</td>
<td>ATQLQK</td>
<td>NLTCE</td>
<td>VWGPTSPK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T39</td>
<td></td>
<td></td>
<td>T40</td>
<td>T41</td>
<td>T42</td>
<td>T43</td>
<td>T44</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LMLSLK</td>
<td>LENK</td>
<td>EAK</td>
<td>VSK</td>
<td>R</td>
<td>EK</td>
<td>AVWWLNPEAGMWQCLLS</td>
<td>DSGQVLL</td>
<td>ESNIK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T44</td>
<td>T45</td>
<td>T46</td>
<td>T47</td>
<td>T48</td>
<td>T49</td>
<td>T50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>VLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFICVVR</td>
<td>CR</td>
<td>HR</td>
<td>R</td>
<td>R</td>
<td>OAER</td>
<td>MSQIK</td>
<td>R</td>
<td>LLSEEK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T52</td>
<td>T53</td>
<td>T54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCQCPHR</td>
<td>FQK</td>
<td>TCSP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T61</td>
<td>T62</td>
<td>T63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.29** The sequence of the T cell surface glycoprotein CD4 with peptides generated from a theoretical digest with trypsin labeled (T1 to T63). Acidic amino acids, aspartic acid (D) and glutamic acid (E) are highlighted by squares around these amino acids. The two peptides that were identified by MS/MS are shown in bold.
<table>
<thead>
<tr>
<th>GCGCSSHPEDDWMENIDVCENCYHPIVPLDGK</th>
<th>GTLLIR</th>
<th>NGSEVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPLVTYEGSNPPASPLQDNLVIALSEPSHDGDGFEK</td>
<td>GEQLR</td>
<td>ILEQSGEWWK</td>
</tr>
<tr>
<td>AQSLTTGQEGFIPFNVAK</td>
<td>ANSLEPEPWFFK</td>
<td>NLSR</td>
</tr>
<tr>
<td>ESEESTAGSFSLSVR</td>
<td>DFDPQNGGEWVK</td>
<td>HYK</td>
</tr>
<tr>
<td>HYTNASQGLCTR</td>
<td>LSRPCQTQKPQKPWWEDFWEVPR</td>
<td>ETLK</td>
</tr>
<tr>
<td>VAVK</td>
<td>SLK</td>
<td>QGSMSPDAFLAEANLMK</td>
</tr>
<tr>
<td>TPSGIK</td>
<td>LTINK</td>
<td>LLDDMAAQIAEGMAFEEER</td>
</tr>
<tr>
<td>LIEQNEYTAZ</td>
<td>EGAK</td>
<td>FPIK</td>
</tr>
<tr>
<td>IPYPGMTNPEVIONLER</td>
<td>GYR</td>
<td>MVRPNCPEDLEQLMR</td>
</tr>
<tr>
<td>SVLEDFFATGEQYQPQP</td>
<td>LCWK</td>
<td>ERPEPRFDYLR</td>
</tr>
</tbody>
</table>

Figure 4.30 The sequence of the protein tyrosine kinase p56^{}^{\text{lek}} with peptides generated from a theoretical digest with trypsin labeled (T1 to T63). Acidic amino acids, aspartic acid (D) and glutamic acid (E) are highlighted by squares around these amino acids. The seven peptides that were identified by MS/MS are shown in bold.
Table 4.5 shows a summary of the acidic residues in CD4 tryptic peptides of masses greater than 1 kDa. Singly charged species in a tryptic digest tend to be from background contamination and spectra were acquired from m/z 500 to avoid the contamination that is often dominant below this m/z value. The possibility of identifying peptides below 1 kDa is therefore unlikely. The first peptide listed in Table 4.5 is part of a signal peptide and would therefore not be detected from mature CD4 on a cell membrane. Of the 16 peptides over 1kDa, only 3 have no acidic amino acids and 6 have a single acidic residue. The remaining 7 peptides have two or more acidic amino acids. Therefore of the peptides with masses that are likely to be distinguished from the ESI background, 44% have two or more acidic residues and therefore have a high chance of being retained by the ion exchange column. These 7 peptides correspond to a total of 152 amino acids out of the total chain length of 458 amino acids, i.e. 33% of the protein. Therefore the maximum possible sequence coverage that could be achieved for CD4 with this method is 33% (Table 4.6).

Table 4.6 shows a summary of the potential sequence coverage from the identified proteins (see section 4.2.2). The number of peptides below, and above, 1 kDa are listed in separate columns, with the number of amino acids in each of the peptides included in brackets. The column listing the peptides with two or more acidic residues includes only those peptides that have masses over 1 kDa. The potential sequence coverage is calculated from the number of amino acids in peptides over 1 kDa that contain two or more acidic residues out of the total number of amino acids in the protein chain (as for the CD4 protein mentioned above).
<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Mass</th>
<th>Acidic residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLLLVLQALLPAATQGK</td>
<td>1898.2</td>
<td>0</td>
</tr>
<tr>
<td>GDTVELTCTASQK</td>
<td>1408.7</td>
<td>1D 1E</td>
</tr>
<tr>
<td>ILGNQGSFLT</td>
<td>1176.7</td>
<td>0</td>
</tr>
<tr>
<td>SLWDQGNFPLIK</td>
<td>1529.8</td>
<td>1D</td>
</tr>
<tr>
<td>IEDSDTYICEVEDQK</td>
<td>1842.8</td>
<td>3D 3E</td>
</tr>
<tr>
<td>EEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCR</td>
<td>4408.2</td>
<td>1D 3E</td>
</tr>
<tr>
<td>TLSVSQLELQDSGTWTCTVLQNQK</td>
<td>2735.3</td>
<td>1D 1E</td>
</tr>
<tr>
<td>IDIVVLAFQK</td>
<td>1144.7</td>
<td>1D</td>
</tr>
<tr>
<td>EGEQVEFSFPLAFTVEK</td>
<td>1956.0</td>
<td>4E</td>
</tr>
<tr>
<td>LTGSGELWWQAER</td>
<td>1531.7</td>
<td>2E</td>
</tr>
<tr>
<td>SWITFDLK</td>
<td>1008.5</td>
<td>1D</td>
</tr>
<tr>
<td>LPLHLTLQPALPYAGSGNLTLALEAK</td>
<td>2828.6</td>
<td>1E</td>
</tr>
<tr>
<td>LHQEVNLLVMR</td>
<td>1336.7</td>
<td>1E</td>
</tr>
<tr>
<td>NLTCEVWGPTSPK</td>
<td>1487.7</td>
<td>1E</td>
</tr>
<tr>
<td>AVWVLNPEAGMWQCLLSDSGQVLLESNIK</td>
<td>3256.6</td>
<td>1D 2E</td>
</tr>
<tr>
<td>VLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVR</td>
<td>3854.2</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 4.5** The sequences of CD4 tryptic peptides of mass >1 kDa with their monoisotopic mass and the number of aspartic acid (D) and glutamic acid (E) residues in the sequence.

Of the 11 proteins identified, the B cell receptor-associated protein has the lowest potential sequence coverage. Tubulin has 12 peptides with 2 or more acidic residues and a potential sequence coverage of 60%. Regulatory light chain myosin has a 70% potential sequence coverage based on amino acids. This light chain myosin, however, has only 9 peptides over 1000 Da and 8 of these contain two or more acidic residues. The anion exchange column may therefore retain 89% of the detectable peptides. Heavy chain myosin
is much larger than the other proteins and has many peptides below 1 kDa. This lowers the potential sequence coverage (43%). Of the heavy chain myosin peptides over 1 kDa, 76% have two or more acidic amino acids, which is a much more promising figure. The protein tyrosine kinase, p56^{ck}, has a potential coverage of 58%, significantly higher than for CD4. These percentages illustrate that acidic residues appear to be quite widespread in these proteins and are encouraging examples that demonstrate the suitability of anion exchange for this type of analysis.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Chain length</th>
<th>Peptides &lt;1000Da</th>
<th>Peptides &gt;1000Da</th>
<th>2+ acidic residues</th>
<th>Potential sequence coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4</td>
<td>458</td>
<td>47 (159)</td>
<td>16 (299)</td>
<td>7 (152)</td>
<td>33%</td>
</tr>
<tr>
<td>p56&lt;sub&gt;lk&lt;/sub&gt;</td>
<td>508</td>
<td>23 (97)</td>
<td>24 (411)</td>
<td>16 (295)</td>
<td>58%</td>
</tr>
<tr>
<td>Tubulin (alpha)</td>
<td>451</td>
<td>20 (71)</td>
<td>20 (380)</td>
<td>12 (271)</td>
<td>60%</td>
</tr>
<tr>
<td>Myosin-regulatory light chain</td>
<td>172</td>
<td>14 (41)</td>
<td>9 (131)</td>
<td>8 (121)</td>
<td>70%</td>
</tr>
<tr>
<td>Myosin-light chain alkali</td>
<td>150</td>
<td>7 (38)</td>
<td>8 (112)</td>
<td>6 (88)</td>
<td>59%</td>
</tr>
<tr>
<td>Myosin heavy chain (type A)</td>
<td>1960</td>
<td>253 (890)</td>
<td>74 (1070)</td>
<td>56 (844)</td>
<td>43%</td>
</tr>
<tr>
<td>Actin - beta</td>
<td>375</td>
<td>20 (75)</td>
<td>17 (300)</td>
<td>10 (194)</td>
<td>52%</td>
</tr>
<tr>
<td>B cell receptor-associated protein</td>
<td>299</td>
<td>29 (140)</td>
<td>12 (159)</td>
<td>5 (77)</td>
<td>26%</td>
</tr>
<tr>
<td>40S ribosomal protein</td>
<td>69</td>
<td>11 (37)</td>
<td>3 (32)</td>
<td>2 (21)</td>
<td>30%</td>
</tr>
<tr>
<td>60S acidic ribosomal protein P2</td>
<td>115</td>
<td>8 (22)</td>
<td>5 (93)</td>
<td>4 (74)</td>
<td>64%</td>
</tr>
</tbody>
</table>

Table 4.6 Summary of the potential sequence coverage of the tryptic peptides from the identified proteins.
4.3 DISCUSSION

4.3.1 Method development

The method shown in Figure 4.3 was developed through collaboration with Dr Andrew Sloane, Mr Oliver Bernhard, Dr Tim Hochgrebe and Dr Garry Lynch.

The IgG1 column served two purposes: to remove components that bound non-specifically to the column packing material, and to provide control samples. By passing the lysate over the IgG1 column prior to the anti-CD4 column, the majority of non-specific binding proteins bound to the IgG1 column and therefore non-specific binding to the anti-CD4 column was minimised. Components that were apparent in the mass spectra of samples from the IgG1 column were considered to be from non-specific binding and were not selected for MS/MS when they were also detected in the anti-CD4 samples.

The anti-CD4 affinity column that contained the immobilised monoclonal antibody Q425 was used for all experiments except those presented in Section 4.2.6. The Q425 antibody requires Ca2+ to bind CD4 and therefore the initial elution protocols used EDTA (a divalent cation chelator) to elute the bound complexes from the column (Sloane, 2000). The EDTA elution protocol minimised elution of proteins that were bound non-specifically to the columns. After further development, glycine was used to elute the components from the column (Oliver Bernhard, Centre for Virus Research, Westmead Millennium Institute, personal communication) and this resulted in the identification of peptides that were not apparent in the samples eluted with EDTA. The mass spectra of samples from glycine elutions also showed more components in the IgG1 samples, as would be expected since glycine is more likely to disrupt non-specific interactions. Despite the elution of more non-
specific binding proteins, there were many components in the anti-CD4 column eluate that were not in the IgG\textsubscript{1} column eluate (Figure 4.8 and 4.9).

Preliminary work with a different antibody, OKT4, has commenced. This antibody does not require Ca\textsuperscript{2+} for CD4 binding and therefore elution with EDTA was not applicable. The first attempts of elution from this column used a buffer containing guanidine (Oliver Bernhard, personal communication). The ESI-MS spectra of samples from this elution showed many multiply charged ions that were present in the IgG\textsubscript{1} samples and the anti-CD4 samples. The presence of the same peptides in samples from both columns suggested that the elution protocol caused a large number of non-specifically-bound proteins to be eluted from the columns. The high concentration of peptides from the non-specific binding proteins masked the less abundant peptides from CD4-binding proteins. The elution protocol for the OKT4 column therefore requires additional development for future identification of CD4-associating proteins. It was interesting that a CD4 peptide was identified in a sample from the OKT4 column but not from the IgG\textsubscript{1} column (Figure 4.28). This identification relied upon previous results and would not have been possible without prior knowledge. The identification of the CD4 peptide did, however, demonstrate that with further development, the OKT4 anti-CD4 column should also provide interesting results, which may complement the results obtained from the Q425 anti-CD4 column.

The vast majority of mass spectrometry publications have involved water-soluble proteins (Schaller, 2000). The poor solubility and low abundance of hydrophobic proteins has impeded their analysis. The development of methods to enable recovery and detection of hydrophobic proteins from their native environment will help identify protein binding
partners and elucidate possible protein functions. A further challenge in the analysis of membrane or hydrophobic proteins is introduced when protein complexes need to be maintained for analysis. In this work detergents were required to solubilize the hydrophobic proteins and to maintain higher order structures that were crucial for protein interactions. The deleterious effect of detergents on ESI mass spectra is well known (Loo et al., 1996) and removal of the detergent is crucial for MS analysis. The non-ionic detergents used in the sample preparation were removed by anion exchange chromatography. A consequence of the anion exchange chromatography, however, was that the peptides detected by MS were predominantly those containing acidic amino acids.

Initial preparations used iodoacetamide in the cell lysis buffer and the resulting samples did not show any evidence of the protein tyrosine kinase, p56\textsuperscript{ck}. When iodoacetamide was excluded from the cell lysis buffer, numerous p56\textsuperscript{ck} peptides were detected. Similar results were obtained by 1-dimensional gel electrophoresis and MALDI analysis of the digested gel bands (Sloane, 2000). P56\textsuperscript{ck} is known to associate with CD4 through two cysteine residues on each protein and a Zn\textsuperscript{2+} ion (Turner et al., 1990; Gratton et al., 2000). Iodoacetamide disrupts the non-covalent association between these two molecules. Therefore the use of iodoacetamide in the lysis buffer can be used to investigate whether co-elution of particular proteins is dependent on the association of p56\textsuperscript{ck} and CD4.

The high background observed in the ESI mass spectra and the presence of abundant singly charged ions often made detection of multiply charged peptide ions difficult. To assist in detection of multiply charged ions, reducing the voltage applied to the MCP detector of the quadrupole-ToF instrument was found to be an advantage. Lower voltages on the MCP detector discriminate against ions of lower charge because of differences in the
momentum of ions with the same m/z but different masses (Loo and Orgorzalek Loo, 1995). For MS/MS acquisitions the MCP was increased for maximum detection of the product ions.

Numerous peptides from human keratins and trypsin were identified in the samples. The contamination of samples with keratins from human skin and hair has often been observed in conventional proteomics applications (as in Chapter 3). Peptides from the autolysis of trypsin are also frequently detected in tryptic digests of proteins.

The cell lysis and preparation in the method presented in Figure 4.3 was designed to capture any CD4 that may be present. This may include mature CD4 on the cell membrane as well as CD4 that is still within the cell. Future refinement may include steps to isolate membrane CD4 associations. Application of the method to the analysis of monocytes and macrophages has commenced (Stephen Watt, Honours student, University of Wollongong, Wollongong). Other future directions include application of the technique to investigate interactions with CD4 in activated T cells, and T cells that have been exposed to gp120. The inclusion of a second affinity column following the anti-CD4 column would also be of interest. An anti-p56$^{leek}$ or anti-gp120 column may provide further information and verify associations between co-eluting proteins. The development of a control preparation with CD4-negative T cells is also being considered (Oliver Bernhard, personal communication).

### 4.3.2 Significance of identified proteins

**CD4 and p56$^{leek}$**

The identification of CD4 in the samples from the anti-CD4 column but not from the IgG$_1$ control column provided initial confidence in the use of the affinity chromatography
method for the identification of potential CD4-associated proteins. The two sequenced CD4 peptides were observed in the early stages of method development and in following preparations were sought as indicators of useful samples. Both the CD4 peptides that were sequenced were from the extracellular domain of CD4.

Following the adaptation of the method to exclude iodoacetamide from the cell lysis buffer, p56\(^{\text{lck}}\) peptides were observed in the samples. The detection of the CD4-associating protein, p56\(^{\text{lck}}\), provided further evidence that the method was applicable to the identification of proteins that associated with CD4. Importantly, p56\(^{\text{lck}}\) was not apparent in the samples from the IgG\(_1\) column, indicating that its identification was not a result of non-specific binding to the column packing material.

**Tubulin, actin and myosin**

The tubulin peptide, TIGGGDDSFNFTFFSETGAGK (residues 41-60), was first identified in samples that had been prepared with iodoacetamide-containing cell lysis buffer. The same peptide and two additional tubulin peptides were later identified in samples that had been prepared from cell lysis buffer that did not contain iodoacetamide. The detection of tubulin in samples from lysis buffer containing iodoacetamide suggested that co-elution of CD4 and tubulin does not depend on the association between p56\(^{\text{lck}}\) and CD4. Immunoprecipitation studies from cell lysates and Western blotting of the eluates from the IgG\(_1\) and anti-CD4 columns confirmed that tubulin associated with CD4 (Sloane, 2000). Furthermore, immunoprecipitation studies were also performed in the presence of iodoacetamide, confirming that the CD4-tubulin association was not dependent on the association of p56\(^{\text{lck}}\) and CD4 (Sloane, 2000).
Myosin peptides were identified in samples both from cells that were lysed in buffer containing iodoacetamide, and from those lysed in the absence of iodoacetamide. As was the case for tubulin, this suggests that co-elution of myosin was not dependent on the association of p56\textsuperscript{ck} and CD4. Many different forms of light and heavy chain myosins have been characterised. Non-muscle myosin is known as myosin-I and is often membrane-bound (Alberts \textit{et al.}, 1994). The myosin molecule consists of two heavy chains, two alkali light chains and two regulatory light chains (Swiss Prot Entry P35579, http://au.expasy.org/sprot,) each of which was identified in the samples from the anti-CD4 column.

Tubulin and actin both form important components of the cytoskeleton. Tubulin forms microtubules and actin forms microfilaments (or actin filaments) (Alberts \textit{et al.}, 1994). Myosin (a motor protein) binds to actin filaments and this complex is important for cell movement (Alberts \textit{et al.}, 1994). Only a single actin peptide was identified in the samples from the anti-CD4 column, but its presence can be rationalised because it is known to associate with myosin.

\textit{Other proteins}

The significance of the two ribosomal proteins that were identified from the anti-CD4 preparations is not clear. The identification of only a single peptide from two different ribosomal proteins does not provide confident identification. Non-specific binding of ribosomal proteins to the column would not be surprising considering the high abundance of ribosomes in cells (Alberts \textit{et al.}, 1994). The identified peptides were not, however, apparent in the control samples. It is possible that immature CD4, which was still associated with ribosomes, was isolated by the affinity purification method. As mentioned
above, the cell preparation method is not specific for membrane-bound CD4. To test this hypothesis, a method that allowed the specific isolation of membrane-associated CD4 would be required.

Identification of the B cell receptor-associated protein is also difficult to explain. The B cell receptor-associated protein is also known as D-prohibitin. A sequence from a different prohibitin protein was also retrieved from a database search but the sequence was unable to be confirmed from the MS/MS data (Table 4.4, precursor at m/z 722.5 with doubly and triply charged ions). Further work is required to establish whether the identification the B cell receptor-associated protein is significant.

4.3.3 Database searching
A variety of different databases and programs for searching the databases are available to the public directly via the Internet or using email based queries. Table 1.2 (Chapter 1) gives a summary of the databases available and the information they contain. The programs that are available to search these databases are shown in Table 1.3. As discussed in Chapter 3, the databases used in this work were searched using Peptide Search programs (http://pepsea.protana.com/) and/or Protein Prospector programs (http://prospector.uscf.edu). Peptide Search programs use the combined entries from the Swiss Prot and Trembl databases and do not have an option for searching other databases. Protein Prospector programs allow searching of Swiss Prot, OWL, Genpept, NCBI nr, Ludwig nr and dbEST. The programs available in Protein Prospector have more options for the input of additional search information. The option to restrict the search to a particular species is very useful because retrieved sequences from unrelated organisms can be
Chapter 4: Identification of CD4 and co-associating proteins

excluded. The ability to enter additional information is crucial when the sequence tag (or stretch of sequence) entered has many possible matches. By specifying the origin of the sample, many extraneous matches can be avoided. Conversely, by introducing constraints, a true identification can be missed altogether. The possibility of a high degree of sequence homology between proteins from different species should also not be ignored. If a sequence is not retrieved when species and/or other constraints are imposed, it is important to attempt the same search with less specific parameters.

The ability to access the vast number of protein and translated oligonucleotide sequences that are available is crucial for proteomic applications. Rapid identifications are possible from proteins that have previously been sequenced and related proteins can be retrieved from these databases. There is significant potential for automation (section 4.3.7) and many laboratories use automated search procedures. With the vast amount of information available, often very little sequence information is required for identification of a protein. There are, however, limitations associated with database searching, particularly with the use of search programs that are available on the Internet. Furthermore, some MS/MS spectra do not lead to protein identification, even when the product ion spectrum is highly informative (such as some of the examples in Table 4.4). These problems are addressed in the following sections.

4.3.4 Peptide fragmentation

The fragmentation of peptides in an MS/MS experiment has been well documented (Papayannopoulos, 1995) and many trends have been observed as outlined in Chapter 1 (Section 1.1.5). Some of these trends were apparent in the mass spectra of peptides in this
work. These characteristic fragment ions are helpful for obtaining information from MS/MS spectra when only a small number of product ions are apparent.

**Characteristic product ions from tryptic peptides**

Tryptic peptides, unless they are the C-terminal peptide of the protein, have an arginine or lysine residue at their C-terminus. The $Y_1$ ion of a tryptic peptide with a C-terminal lysine has an m/z of 147 and the $Y_1$ ion of a peptide with a C-terminal arginine has an m/z of 175. These characteristic ions are often helpful in determination of the C-terminal amino acid. The peptide EGEQVEFSFPLAFTVEK (Figure 4.14) is an example where the $Y_1$ ion from lysine is clear, whilst the peptide LIEDNEYTAR (Figure 4.15) has a C-terminal arginine. When an MS/MS spectrum is derived from more than one peptide of the same nominal mass, there may be evidence of both an arginine and lysine $Y_1$ ion, as in the example shown in Figure 4.16 for the two peptides LTGSGELWWQAER and CamDFTEDQTAEFK. Assignment of the $Y_1$ ion may also assist in assignment of the $B_{n-1}$ ion if present.

The presence of the C-terminal basic residues has a substantial effect on the MS/MS spectra of peptides (Hunt *et al.*, 1986; Tang *et al.*, 1993). The fragmentation typically results in charge retention on the C-terminus and the predominance of Y-type sequence ions. This same trend was clearly observed here for the peptides presented in Figures 4.10 through 4.17.

**A$_2$/B$_2$ product ions**

The $B_1$ ion is rarely observed in MS/MS spectra but the $B_2$ ion is usually present and is generally observed with the corresponding $A_2$ ion (Kinter and Sherman, 2000). The $B_2/A_2$ product ions can be useful tools for identification of the N-terminus. The possible amino
acids in the B_2 ion can be calculated from their residue masses. The complementary Y_{n-2}^- ion may be present in the MS/MS spectrum. If the Y_{n-1}^- ion is also present, the identity of the N-terminus can be confirmed. The peptide LIEDNEYTAR (Figure 4.15) provided an example where these ions enabled confirmation of the N-terminus. The B_2 ion was apparent at m/z 227.2, its identity confirmed by the A_2 ion 28 Da below the B_2 ion. This B_2 ion could have arisen from a combination of proline and glutamic acid, or from 2 leucine/isoleucine residues. The molecular ion mass was 1224.8 Da (612.4 x 2) and the corresponding Y ion at 1224.8-227.2 = 997.6 was clearly present. The Y_{n-1}^- ion was also apparent, at an m/z 113 higher than the Y_{n-2}^- ion. The Y ions confirmed that the two N-terminal amino acids were isoleucine or leucine residues (since these two amino acids are indistinguishable by low energy CAD).

Although the B_2 and A_2 product ions were clearly evident for the peptide LIEDNEYTAR (Figure 4.15) and provided information on the peptide sequence, their presence was not essential for the identification of the peptide in this case. In contrast, the peptide SYELPDPQVITIGNER (Figure 4.26) was an example where the B_2 and A_2 ions were critical for identification of the peptide sequence. Recognising that the ion at m/z 251.2 was a B ion lead to the generation of a sequence tag from the B ions in the low m/z range and subsequent identification of a peptide from actin.

**Fragmentation at proline**

Preferential fragmentation at the N-terminal side of proline residues (the "proline effect") has been extensively documented (Hunt et al., 1986; Barinaga et al., 1989; Tang et al., 1993). Owing to the preferential cleavage on the N-terminal side of proline, cleavage on the C-terminal side is far less frequent and consequently ions from cleavage on the C-terminal
side of proline are often of low abundance or not present. Examples of this effect were apparent for the peptides EGEQVEFSFPLAFTVEK (Figure 4.14) and IPWFQYPIIYDIR (Figure 4.17). For EGEQVEFSFPLAFTVEK the Y$_8$" ion, which results from cleavage N-terminal to the proline residue, was of high abundance. The Y$_7$" ion from cleavage on the C-terminal side of proline was not apparent in the spectrum. A similar observation was seen for the peptide IPWFQYPIIYDIR where the Y$_7$" ion (PIIYDIR) was of high abundance and the Y$_6$" ion (IIYDIR) was of low abundance.

A second observation with proline-containing peptides is the dominance of product ions that result from internal fragments. Following cleavage at the proline amide bond, producing a typical Y-type ion, a second cleavage of the type that would normally produce a B ion can occur (Hunt et al., 1986). An example of abundant internal fragments resulting from cleavage adjacent to proline was apparent from the peptide IPWFQYPIIYDIR (Figure 4.17).

*Other fragmentation observed in this study*

Product ions resulting from cleavage at a glycine residue are typically lower in abundance than other product ions, particularly if the cleavage is between GG or AG (Staudenmann and James, 2001). This trend was apparent in a number of the spectra in this work. An example was in the spectrum from the peptide TIGGGDDSFTFFSETGAGK (Figure 4.13) where cleavage between isoleucine (I) and glycine (G) to give Y$_{18}$" produced a more abundant ion than the Y$_{17}$" ion produced from cleavage between two glycine residues. The Y$_{15}$" and Y$_{16}$" ions were also of lower abundance owing to cleavage at glycine. The Y$_{14}$" ion, produced by cleavage between two aspartic acid residues, was significantly more abundant than either Y$_{15}$" or Y$_{16}$".
The loss of characteristic neutral fragments from product ions can also assist in assignment of amino acids. Loss of ammonia (17 Da) typically occurs from peptides that contain asparagine, glutamine, lysine or arginine but may also occur from the N-terminus (Kinter and Sherman, 2000). The loss of water (18 Da) often occurs from peptides that contain serine or threonine but may also occur from peptides that contain the acidic amino acids, glutamic acid and aspartic acid (Kinter and Sherman, 2000). When the N-terminal amino acid is glutamic acid, the loss of 18 Da is observed owing to the formation of pyroglutamic acid. The peptide EGEQVEFSFPLAFTVEK (Figure 4.14) provided an example of a peptide with an N-terminal glutamic acid. The loss of 18 from the B$_5$ ion resulted in the ion at m/z 525.3. Likewise, loss of 18 from the B$_4$ ion resulted in the ion at m/z 426.2.

Another point to note, which is not a direct observation of a fragmentation pattern but has important implications on the appearance of the MS/MS spectra, is the charge on the peptide ions. The vast majority of the multiply charged ions in the MS spectra of the digests were doubly charged. Tryptic peptides have at least two basic sites for proton attachment, the C-terminal lysine or arginine and the N-terminus. The location of the charge on peptides in MS/MS experiments controls the fragmentation of the peptide (see Chapter 1). The dominance of multiply charged peptides and the characteristic spectra produced from MS/MS of tryptic peptides has resulted in the prevalent use of trypsin for enzymatic cleavage prior to MS/MS. Of the peptides that resulted in protein identification (Table 4.2), only three were not doubly charged. The peptide consisting of residues 588-608 from heavy chain myosin (DFSALESQDQDQLQELLQEENR), the peptide from residues 63-78 of myosin light chain alkali (VLDFEHFLPMLQTVAK), and the peptide from residues 478-
490 of p56<sup>lck</sup> (ERPEDRPTFDYLR) were identified from triply charged ions. The sequence from the heavy chain myosin peptide did not contain any additional basic sites that would result in an additional proton to those expected on the C-terminal arginine and the N-terminus. The doubly charged ion of the same peptide was also apparent, suggesting that the triply charged ion was not necessarily the predominant ion in the spectrum. The sequence of the myosin light chain alkali peptide contained a histidine residue, which has significant gas phase basicity and might also be expected to be protonated. The three protons would therefore most likely reside on the C-terminal lysine, the N-terminus and the histidine residue. The sequence of the triply charged peptide from p56<sup>lck</sup> has two internal arginine residues as well as a C-terminal arginine. The internal arginine residues each arise from the presence of a proline residue C-terminal side to the arginine. Trypsin is unable to cleave the peptide bond between a lysine or arginine and a proline residue (Kinter and Sherman, 2000). The presence of the internal arginine residues gave rise to a triply charged peptide ion.

The location of protons, and the ability of the protons to mobilise and locate at other peptide bonds, is a fundamental aspect of peptide fragmentation (Wysocki et al., 2000). The presence of basic amino acids in the peptide, such as the arginine residues in the p56<sup>lck</sup> peptide and the histidine in the myosin light chain alkali peptide, may have significant effects on the MS/MS spectra. The localisation of protons on basic residues can result in a decrease in the abundance of other ions, impeding sequencing of the full peptide. These effects may be of importance in the interpretation of some of the peptides that did not result in protein identification (Table 4.4). The analysis of peptide fragmentation is, however, beyond the scope of this thesis and will not be expanded further here.
Options for database searching

As mentioned in Chapter 3, the Protein Prospector programs enable more information to be entered for database searching compared to Peptide Search. Searching Swiss Prot with Peptide Search has advantages because less time is required to enter the information. For MS/MS spectra with clear product ion series (usually Y ions) Peptide Search is very easy to use and the Swiss Prot/Trembl database is often sufficient for protein identification. For spectra that do not have clear ion series, but do contain significant information, the programs in Protein Prospector may be more useful. An example of this can be drawn from the spectrum of the peptide SYELPDGQVITIGNER from actin (Figure 4.26). The B ions at m/z 251.1, 380.2 and 493.3 gave a sequence tag that resulted in the identification of actin using either Peptide Search or the program MS-Seq in Protein Prospector. Alternatively, MS-Seq searching of the same database with only the two Y ions above the precursor, m/z 1087.7 and 1298.7, resulted in the same matches as for the B ion search with only two additional proteins (134 matches, 132 of which were actin, the correct match). To enter the same Y ion information into Peptide Search is much more difficult. The mass difference between m/z 1087.7 and 1298.7 is 211 Da, which could correspond to either the combinations GGP (in any order) or NP (or PN). Furthermore, the Y ion of m/z 1087.7 was actually the second isotope (i.e. contained one $^{13}$C atom). Hence, the true mass difference is 212 Da, which could correspond to DP or VL (VI) (in any order). If all of the possibilities were taken into account, a number of tags would have to be separately entered into Peptide Search. Hence, MS-Seq provided a more convenient search tool for this example.

With increased understanding of peptide fragmentation and improved ability to control the fragmentation through derivatisation, there is a need for more versatile search
programs. Many research groups working in the field of proteomics have in-house software programs to cope with individual requirements. With highly specific fragmentation and programs that are able to utilise the information to its full potential, a relatively small amount of information is required for identification of a protein that is already present in a database.

4.3.5 Unidentified peptides

Table 4.4 outlined the large number of precursor ions that were selected for MS/MS but did not result in identification of the parent protein. There were a number of reasons that may explain why these peptides remain unidentified. One possibility is that the protein from which the peptide originated is not in the database that was searched. A more likely explanation, however, is that there is some inconsistency between the information entered into the search program and the information contained in the database. Molecular mass changes in a peptide may arise from post-translational modifications or modifications introduced through sample handling. An example of the effect of changes in the peptide mass was observed with the peptide from residues 103-122 of regulatory light chain myosin. As stated in section 4.2.2, this peptide was observed with three different molecular masses. The two peptides that did not correspond to the correct molecular masses were from a sodiated peptide ion and a peptide ion that had lost ammonia. The extensive sequence information from the MS/MS spectra enabled the identification of each of these peptides through “sequence-only” searches. If a sequence tag had been used rather than “sequence-only” searching, the incorrect peptide mass would have prevented identification of these peptides. The use of “sequence-only” searching to avoid any mass discrepancies is
advantageous but there is not always sufficient information in the MS/MS spectrum to obtain a stretch of amino acids of an ideal length for this type of search (i.e. approximately 5 or more amino acids).

Other problems encountered with identification of peptides from database searching arose when minimal fragmentation was apparent in the MS/MS spectrum, which prevented confirmation or exclusion of sequences that were retrieved. Erroneous sequence tags may also account for some of the unidentified peptides.

4.3.6 Capillary LC-MS

The development of capillary LC-MS has vastly improved the analysis of complex mixtures by mass spectrometry. The capability of detecting numerous components eluting at the same time removes the need to fully resolve the components in a mixture by other means. Even with short chromatography gradients and minimal separation, there is a noticeable improvement in the detection of peptides in mixtures such as the anti-CD4 samples in this work.

The CapLC™ used for this work is capable of delivering low flow rates without flow splitting. A flow rate of 2 μL/min with a 0.32 mm column was used for the experiments presented in section 4.2.4. Lower flow rates and smaller columns could also be used with this system to further improve the sensitivity. Although the TIC chromatogram did not clearly show the elution of peptides, the BPI chromatogram often provided an indication of the retention time at which the peptides began to elute from the column. The most valuable information was obtained from the individual chromatograms of particular m/z ratios.
The most notable advantages in the use of LC-MS were an increase in the signal-to-noise ratios of individual peptides, and a corresponding increase in sensitivity. Components that were not apparent in the spectra from nanoESI-MS were apparent in the spectra from LC-MS. The separation of different peptides with the same m/z was also beneficial, simplifying interpretation of MS/MS spectra and improving the ability to assign the charge on the precursor ions that otherwise overlapped.

A further aspect of LC-MS is the capability of automated sample handling and data processing (see section 4.3.8).

4.3.7 Data dependent MS to MS/MS switching

The option for “function switching” during an LC-MS experiment allowed MS/MS data to be acquired when suitable precursor ions were detected by MS. The data generated from these experiments was highly dependent on the precursor selection parameters and the data acquisition parameters. For this type of analysis, time is a crucial factor that must be taken into account as components that elute while the mass spectrometer is still in MS/MS mode rather than MS mode will remain undetected.

The selection of an ideal precursor involves a number of important user-defined parameters. The first of these is the threshold, which is the minimum ion count at which the instrument will switch into MS/MS mode, providing other criteria are also met. The threshold is set to avoid MS/MS acquisitions on background noise. As noted in section 4.2.5, the threshold in this work was set to between 3 and 5 ion counts. Other optional parameters can also be used to improve the selection of peptide ions and avoid contaminants. These parameters include: charge recognition, lists of excluded and included
masses, a time window for selection of the same precursor and a mass window for precursor selection. In the current work, only precursors with 2, 3 or 4 charges were selected for MS/MS. A list of masses to include was not used but masses of previously identified precursors and known contaminating peptides were entered into exclude lists in some experiments. The time window offers the option to exclude a precursor that has been selected once in the LC-MS acquisition from further selection in the remainder of the experiment. The same precursor can be excluded for the whole experiment or for a specified period of time. In this work, selected precursors were excluded for only a short time, merely to prevent repeated selection of the same precursor in an eluting peak. Importantly, if the same m/z eluted with a different retention time it was likely to be from a different peptide and was able to be selected for MS/MS. The mass selection window prevents the selection of more than one isotope of the same ion. Various detection windows were used for different acquisitions and these ranged from m/z 0.6 to 1.5.

It was found that the crucial parameters for obtaining clear product ions in the MS/MS spectra were the length of time each precursor was acquired and the voltage on the collision cell. The nanoES-MS/MS data acquired previously provided very useful information for ideal collision cell voltages. There are a number of available options for setting the collision cell voltage in a “function switching” experiment: i) charge recognition, which calculates an ideal collision cell voltage dependent on the charge of the ion selected for MS/MS; ii) an option for setting the collision cell voltage manually on the tuning page of the standard MS software or iii) a “collision energy profile”, which is set by the user. The “collision energy profile” selects the collision cell voltage based on the m/z of the precursor and a list of voltages. The profile can contain a number of collision cell
voltages for particular m/z ranges. For example, if the m/z range 500-700 had three different voltages, 28 V, 29 V and 30 V, a precursor in this m/z range would be subjected to MS/MS at each of these voltages, time permitting. Each voltage requires the acquisition of at least one spectrum; therefore with a 1 s integration time and a 0.1 s delay, acquisition of a spectrum at each collision cell voltage would take 1.1 s. To obtain a single spectrum at each voltage would take 3.3 s. The consideration of time factors is therefore crucial to obtain sufficient data.

The switch from MS/MS back to MS can also be set in a number of ways. For this work the instrument was set to return to MS after a specified time in MS/MS. Importantly, the time set was dependent upon the number of precursors that were selected for MS/MS from a particular mass spectrum. As noted in section 4.2.5, MS/MS data can be acquired on up to eight components simultaneously but in most experiments this was limited to four components. If a time of 3.3 s was set and four components were selected for MS/MS the instrument would not return to MS mode for a total of 13.2 s. If only a single component was selected, the switch would occur after only 3.3 s. A balance between improved data quality (by increasing the number of spectra acquired in MS/MS) and the chance of missing newly-eluting precursors must be achieved. Various times were used in different experiments to explore the possibilities for these particular mixtures of peptides. Acquisitions with longer MS/MS times resulted in superior spectra but often missed suitable precursors. Shorter MS/MS acquisition times resulted in the selection of many precursors but some MS/MS spectra did not have sufficient data to produce sequence tag information.
Another important point to make about the selection of precursors is that as components elute from the column and are detected by MS, the instrument will switch into MS/MS as soon as the threshold is reached for a single or multiple components. At that point in time, MS/MS is performed on whichever precursor(s) reached the threshold. There is a possibility that at any particular retention time range, more suitable precursors eluted immediately after those selected for MS/MS and these could be missed entirely. This means that even the most abundant peptides may be missed in a particular LC-MS/MS experiment. A number of LC-MS/MS experiments on the same sample, with small changes in MS/MS acquisition time or LC elution gradients may dramatically influence the acquisition of data and should be included as standard protocol in these types of experiments.

4.3.8 Potential for automated sample handling and data processing

The requirement for high throughput analyses in proteomic applications has increased the need for at least partial automation in the analysis of peptides and proteins. As discussed in Chapter 3, the processing of spots from gels often uses robotic gel cutters and sampling handling robots. Large laboratories may have automated sample handling, data acquisition and data processing. With small laboratories, available resources limit even partial automation of procedures. The potential for automation in the procedure outlined in Figure 4.3 in our laboratory is in the LC-MS acquisitions and the processing of data from nanoES-MS/MS and LC-MS/MS.
Automation of LC-MS acquisitions

The instrumentation used for LC-MS (CapLC™ and QTof™2 with MassLynx™ software) enables processing of large numbers of samples with automatic injection and acquisition of samples. The CapLC™ can also be used for addition of reagents in protocols such as enzymatic cleavages. In this project, minimal automated sample handling was used for a number of reasons. First, the small number of samples did not necessitate automatic sample handling. Secondly, the LC-MS procedures were not well established and changes were made to each experiment to investigate the most suitable settings for particular samples. Furthermore, the option for automatic injection and LC-MS of samples requires that the instrument always be prepared for data acquisition. A significant amount of time is required for column conditioning and data need not be acquired in the initial part of an LC-MS experiment (prior to the elution of peptides). This results in a large wastage of gas. A small change in the automated control to include a switch for the nebuliser and desolvation gas is all that is required for inclusion of gas control in the automated acquisition.

Automation of data processing

The software used for the acquisition of MS data, MassLynx™, includes data processing and database searching software. With the large amount of data generated from a “function switching” experiment, data processing can be very time consuming. The ProteinLynx™ software enables automatic summing of spectra from the same precursor and the option for performing a background subtract and data smoothing. The peak profiles are converted to single points corresponding to the peak centre and a peak list is generated. The peak list can then be entered into a search program for database searching. It was found that the success of this automatic processing was highly dependent on the quality of the data. Although
manual processing was time consuming, more useful information was obtained through visual examination of summed spectra of each precursor and critical assessment of the data. Software to assist in the assignment of stretches of sequence in an MS/MS spectrum was useful, but was again dependent upon the quality of the data. For confidence in the assignment of any peptide sequence to a particular MS/MS spectrum, manual assessment was deemed necessary. The potential for automation in small-scale projects may be weighed up against the possibility of compromising the accuracy of outputs. For high throughput projects, manual processing of the large amounts of information is not practical.

4.3.9 Conclusions: Summary and future directions

The development of a method for the isolation of proteins that co-elute with CD4 has resulted in the identification of proteins that may associate with CD4. The method has proved to be a rapid and sensitive way of identifying proteins for further immunological investigations. Further studies with one of the identified proteins, alpha tubulin, showed co-immunoprecipitation of CD4 and tubulin (Sloane, 2000). Additional investigations of the significance of other identified proteins are yet to be undertaken. Application of the same method to the study of monocytes and macrophages is currently underway (Stephen Watt, Honours project, University of Wollongong) and future directions for this work include application of the method to identify proteins that co-elute with CD4 in activated T cells, and in T cells that have been exposed to the HIV-1 envelope glycoprotein, gp120.

The method used here also provides an example of the use of proteomics for identifying possible binding partners, which in turn suggests potential functions of the identified proteins. Examples of functional proteomics approaches are increasing in the
literature and are of significant value in understanding protein function. Developments in biological mass spectrometry have led to widespread improvements in the field of proteomics. The use of LC-MS has enabled the analysis of complex mixtures of peptides with low limits of detection. Procedures for automation have also advanced many high-throughput applications and are likely to increase in importance, even in small-scale laboratories. The combination of different chromatography methods and advancements in the field of chromatography may result in less dependence on the time-consuming and labour-intensive methods of two-dimensional gel electrophoresis.
CHAPTER FIVE: NONCOVALENT COMPLEXES OF GLUTATHIONE S-TRANSFERASE

5.1 INTRODUCTION

5.1.1 Noncovalent Interactions

Noncovalent interactions are critical for the maintenance of the secondary, tertiary and quaternary structure of proteins. The four main types of interactions important in protein-protein and protein-ligand complexes are electrostatic interactions, hydrogen bonding, van der Waals forces and hydrophobic effects.

Early methods that were used to obtain information on protein size and the components present in a protein-protein or protein-ligand complex included electrophoresis, size exclusion chromatography, and ultracentrifugation. Analytical ultracentrifugation is still a powerful tool for investigating stoichiometry and equilibrium constants (Loo, 1997). The ultraviolet (UV) absorption of a mixture of a protein and ligand is measured as the sample is centrifuged. The higher molecular weight complex settles faster and is apparent from the change in absorbance through the centrifugation process (Fersht, 1999). Electrophoresis and size exclusion chromatography as techniques for molecular weight determination do not provide mass accuracy comparable to that provided by modern mass spectrometry. Equilibrium gel filtration, however, is useful for determination of equilibrium constants (Fersht, 1999). In this method a size exclusion column is equilibrated with buffer containing the ligand of interest. The protein is then applied to the column in the equilibrating buffer. Free ligand is able to penetrate the beads and is therefore retained longer by the column. The beads do not retain ligand that is bound to the protein and the complex travels through
the column more quickly. The resulting peak areas of the free ligand and the complex with the protein are used to calculate the binding properties of the system (Fersht, 1999).

Spectroscopic methods such as circular dichroism (CD), UV and fluorescence are not able to provide high resolution structural information but are useful for monitoring protein folding or denaturation and also to observe changes in structure upon binding of ligands. Tryptophan, and to a lesser extent tyrosine, absorb light at around 275-295 nm and re-emit at around 330-340 nm (Fersht, 1999). The environment around these amino acids impacts the intensity of their fluorescence (Mathews and van Holde, 1990). Tryptophan only fluoresces weakly in an aqueous environment but when folded into a hydrophobic core it fluoresces more strongly (Fersht, 1999). Quenching or enhancement of fluorescence may be observed when a protein undergoes conformational changes owing to ligand binding or denaturation.

The peptide backbone of proteins absorbs UV light in the 180-220 nm region, whilst the aromatic side chains absorb in the 275-295 nm region (Fersht, 1999; Mathews and van Holde, 1990). Changes in the conformation of proteins also affect the UV absorption spectra of proteins.

CD examines the difference in absorption of left and right circularly polarised light by optically active molecules (Fersht, 1999; Pelton and McLean, 2000). The peptide backbone of a protein is optically active and the secondary structure produces different characteristic absorbances. CD can be used to estimate the amount of α-helices and β-sheets within a protein (Mathews and van Holde, 1990). It is important to remember that when used alone, this technique does not provide detailed structural information. It is most useful when comparing two different structures, such as the change in protein conformation upon binding of a ligand to a protein.
Surface plasmon resonance (SPR) is a relatively new technique for examining the binding of ligands to an immobilised protein. This technique measures the change in refractive index as a ligand in solution is passed across an immobilised protein and a complex forms between the protein and ligand (Loo, 1997). A problem with this technique is the possibility that the binding properties of the protein are altered when it is immobilised on the surface. There is also a limit to the molecular weight of the ligand that can be used since small differences in the molecular mass are not easily detected.

Differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC) are the two most widely used forms of calorimetry (Larive et al, 1999). DSC measures the heat absorbed by a protein as it undergoes denaturation (Fersht, 1999). The denaturation of a protein is highly endothermic and therefore a high uptake of energy occurs during this process (Fersht, 1999). The reversible binding of ligands can be monitored by the change in the heat absorption curve caused by the presence of ligands. ITC determines the heat of binding by measuring the electrical current required to maintain a constant temperature whilst ligand is added to the protein (Darby and Creighton, 1993; Fersht, 1999). The information obtained from this type of experiment is dependent on how close the concentration is to the dissociation constant. There are three possible measurements that may be obtained. These are: the stoichiometry of the complex; the enthalpy of binding; and the dissociation constant. The enthalpy of binding can be measured regardless of the size of the difference between the dissociation constant and the protein concentration. The dissociation constant can only be calculated when the concentration is less than or equal to the dissociation constant. The stoichiometry can be obtained when the concentration greater than or equal to the dissociation constant. (Fersht, 1999). Calorimetry techniques are important for
measurement of the overall binding enthalpies but they do not distinguish binding to potentially different quaternary structures (van Dongen and Heck, 2000).

Nuclear magnetic resonance (NMR) and X-ray crystallography are the two techniques that are able to provide high-resolution structures of biomolecules and biomolecular complexes. Multi-dimensional NMR, where the sample is irradiated with two RF fields, can be used predict secondary and even tertiary structure of a protein through the generation of a series of possible structures (Fersht, 1999; Pelton and McLean, 2000). These experiments usually involve measurement of proteins that have been labelled with $^{13}$C or $^{15}$N (Darby and Creighton, 1993) and knowledge of the primary sequence of the protein is required. Relatively large sample amounts are required for NMR and the high concentrations of protein necessary may result in protein precipitation (Loo, 1997). Another drawback with NMR is that it is generally limited to masses below 30 kDa (Fersht, 1999; Pelton and McLean, 2000).

X-ray crystallography can determine structures in crystals to a resolution of approximately 1Å and the technique has provided unsurpassed information on the structure of biomolecules. The biggest obstacle with X-ray crystallography, however, lies in the production of suitable crystals. This requires a large amount of very pure protein and crystals of some proteins may be unattainable.

Mass spectrometry has enormous potential for the analysis of noncovalent complexes. The introduction of electrospray ionisation has enabled the transfer of intact complexes into the gas phase. The speed and sensitivity of MS analyses is an advantage for protein analyses. In addition, the high mass accuracy achieved with MS enables clear assignment of the stoichiometry of complexes. There is also mounting evidence that demonstrates agreement between binding constants calculated from MS data and constants acquired from condensed phase techniques such as calorimetry (see section
5.1.2 below). These data must be carefully interpreted, though, as some systems do not show such agreement.

**5.1.2 Mass spectrometry of noncovalent complexes**

The analysis of noncovalent complexes by mass spectrometry has been an area of interest for over a decade. Research has been directed to the understanding of the processes occurring in the gas phase as well as to answering specific questions on protein structure. Many examples of protein structural investigations are now available in the literature. These include DNA/protein interactions (Veenstra, 1999), protein/carbohydrate complexes (van Dongen and Heck, 2000), large protein/protein complexes (van Berkel et al., 2000; Benjamin et al., 1998) and protein interactions with ligands (Robinson et al., 1996; Wu et al., 1997). Complexes involving a vast array of different proteins that have been explored by MS include: HIV proteins (Loo et al., 1998b; Sannes-Lowery et al., 1997), ribosomes (Benjamin et al., 1998), enzymes (van Berkel et al., 2000; Wu et al., 1997; Chung et al., 1999), carrier proteins such as transthyretin and retinol-binding protein (Rostom et al., 1998; Nettleton et al., 1998) and hemoglobins (Green et al., 1999; Lippincott et al., 2000). There have been a number of reviews on the application of mass spectrometry for the study of noncovalent complexes, which provide more details on these studies (Przybylski and Glocker, 1996; Loo, 1997; Smith and Light-Wahl, 1993; Smith and Zhang, 1994).

*Instrumentation*

It has been well documented that analyses under non-denaturing conditions result in ion series of lower charge than those seen under standard denaturing conditions (Loo, 1997; Przybylski and Glocker, 1996; Smith and Light-Wahl, 1993). The lower charges and higher molecular weights of noncovalent complexes require a higher m/z range for
detection. Most available mass analysers have also been used for the detection of noncovalent complexes (Loo, 1997; Smith and Light-Wahl, 1993) but with complexes of very high molecular weights, the m/z range of these complexes may be outside the limit of detection (Murayama et al., 1997). Orthogonal ToF analysers can routinely detect m/z up to 20 000, whilst conventional quadrupoles are usually limited to ca. m/z 4000 and for this reason ToF analysers are increasingly being used for these applications. There are examples of quadrupole mass filters that have been modified to increase this limit to over m/z 10 000 specifically for these applications (Loo and Sannes-Lowry, 1998). Further advantages of ToF analysers are the high resolution and sensitivity. The hybrid quadrupole-ToF instrument is therefore an ideal instrument for the analysis of noncovalent complexes. The quadrupole provides a first stage mass analyser, which offers the option for MS/MS. In the work presented here both the Micromass QTof and QTof2 were used.

The transfer of noncovalent complexes to the gas phase requires careful optimisation of a range of instrumental parameters (Smith and Light-Wahl, 1993; Przybylski and Glocke, 1996; Loo, 1997; Loo and Sannes-Lowry, 1998). These include the solution containing the analyte and the parameters of the electrospray interface. Standard ESI-MS analysis of proteins and peptides generally uses a solution of the analyte in aqueous organic phase (usually 50% acetonitrile or methanol). The presence of the organic solvent improves the stability of ion signal and also affords increased sensitivity (Loo, 1997). A weak acid such as 0.1-0.2% formic acid or acetic acid is also included in the solution. The organic solvent and reduced pH disrupt noncovalent interactions and causes protein unfolding. These conditions are obviously unsuitable for analysis of noncovalent complexes by ESI-MS. To maintain the complexes most analyses are performed in an aqueous buffer system at a pH close to physiological
conditions. ESI-MS is sensitive to salts and there are a limited number of buffers that can be used. Ammonium acetate and ammonium bicarbonate are useful buffers that are tolerated by ESI-MS at low concentrations (millimolar range). Alternatively the analyte can be dissolved in unbuffered water but this generally provides less reproducible results. These solutions do not disrupt the noncovalent interactions that are crucial for higher order structures in proteins. The ease with which a stable spray is achieved is reduced compared to conventional ESI solutions but most samples can still be analysed from these aqueous systems at suitable temperatures and source conditions.

Nanoscale ESI, or nanoESI (Wilm and Mann, 1996), has similar properties to standard ESI but the droplets produced are much smaller (see Chapter 1). NanoESI therefore requires less desolvation and does not use nebuliser gas. The possibility of collision-activated dissociation (CAD) occurring in the nanoESI source is therefore reduced compared to the standard ESI source. Furthermore, the capillary voltage is much lower on the nanoESI source. For these reasons nanoESI is thought to provide much more gentle conditions for the transfer of noncovalent complexes to the gas phase (Fligge et al., 2000; Chung et al., 1999; Rostom et al., 1998; Nettleton et al., 1998).

There are a number of source parameters that are critical for the transfer of noncovalent complexes to the gas phase. The source temperature, capillary voltage, desolvation gas, cone voltage, collision gas, capillary position and pressure in the source region all have pronounced effects on the abundance of noncovalent complexes seen in the gas phase. Changing one parameter often has an effect on the other parameters. This is particularly noticeable for the optimisation of pressure and voltage (Tahallah et al., 2001). If conventional ESI is used, the nebuliser gas also requires optimisation. If nanoscale ESI is used, the capillary may need special attention to achieve a stable spray. This involves a balance between capillary voltage, back-pressure on the capillary and
size of the capillary orifice. The retention of noncovalent complexes in the gas phase
requires gentle conditions that still achieve effective desolvation. Using more stringent
settings for one parameter may allow milder conditions in another region (Smith and

Behaviour of complexes in the gas phase

The question at the forefront of the analysis of noncovalent complexes by mass
spectrometry is whether this gas phase analysis can be correlated with the condensed
phase behaviour of molecules. A number of observations have raised specific issues in
this area. An issue of particular relevance to the complexes of GSTs in this study is the
observation of non-specific complexes in the gas-phase. These have been observed for a
number of systems and have also been documented under conventional electrospray
conditions, particularly the formation of complexes from charge-carrying components
present in the solution (Smith and Light-Wahl, 1993). The reasons for these non-
specific complexes are not always clear but it is possible that they reflect both
condensed phase complexes and complexes that are introduced by the electrospray
process (Smith and Light-Wahl, 1993; Przybylski and Glocker, 1996). Decreasing the
concentration of the analytes may reduce the occurrence of these non-specific gas phase
aggregates (Loo, 1997). To ensure that a complex seen in the gas phase is from specific
interactions, a number of simple investigations should be undertaken. First, the
spectrum obtained of the components in denaturing solution should be compared to the
spectrum obtained under non-denaturing conditions. This reveals associations that are
disrupted with organic solvent and acidic conditions and thereby identifies any covalent
modifications that may be present. Even slight changes in pH or organic solvent content
of the solution may reveal that the complex is sensitive to the solution conditions and
that covalent modification has not occurred.
Once the possibility of any covalent adduct formation has been ruled out, investigations to assist in the assignment of the complex as either specific or non-specific can be pursued. The stoichiometry of the complex is an important feature of the spectrum. The dominance of particular complexes in spectra may indicate specific binding interactions. Conversely, a large number of complexes from the components in varying combinations, with little variation in abundance, may suggest non-specific aggregation rather than specific interactions. To further investigate the binding stoichiometry, the concentrations of individual components in the solution can be varied. Experiments involving changes in the source parameters (such as temperature or voltage) may also provide information on the stability and specificity of a particular complex. Competition experiments can also be useful to support the specificity of a complex (Przybylski and Glocker, 1996). The use of site-directed mutants is also a valuable tool for investigating the specificity of a binding interaction, particularly with competition experiments between mutants and the native protein (Sannes-Lowery et al., 1997). Increasing the cone voltage causes an increase in the acceleration of the ions through the intermediate pressure region of the source and thereby promotes CAD with gas molecules in this region (Loo et al., 1998b; Green et al., 1999; van Dongen and Heck, 2000). Gentle CAD conditions may dissociate non-specific interactions more readily than specific interactions but it must be considered that weak noncovalent interactions may also be disrupted this way.

The majority of evidence suggests that ESI-MS does have the ability to detect specific noncovalent complexes that are also present in the condensed phase. It appears that even complexes of relative low affinity are amenable to analysis by this technique (Baczynskyj et al., 1994). Careful consideration of complexes observed and their behaviour under the changes mentioned above provides a powerful tool for discovery of
the preferred stoichiometry of components that form complexes under non-denaturing conditions. Further questions are posed beyond the simple task of observing the specific complexes. First, does the behaviour of these complexes reflect their condensed phase binding properties? Secondly, does the gas phase complex retain some or all of the condensed phase structure?

Electrostatic interactions are thought to be strong in the gas phase whilst hydrophobic effects are not expected to be maintained with the loss of solvent molecules (Robinson et al., 1996; Loo, 1997; Wu et al., 1997; Loo and Sannes-Lowry, 1998; Last and Robinson, 1999; Chung et al., 1999). A study of acyl CoA binding protein (ACBP) and various acyl CoA derivatives of different hydrophobic chain length illustrated this effect. The increasing chain length did not appear to affect the proportion of complexes in the gas phase, even though the condensed phase dissociation constants are significantly different (Robinson et al., 1996). In a study of HIV protein complexes, Loo and colleagues reported that complexes relying on electrostatic interactions were more stable in the gas phase than complexes held together by hydrophobic interactions (Loo et al., 1998b). Importantly, however, the condensed phase behaviour of the complexes that did contain hydrophobic interactions was reflected in the MS data, even though the complexes were more sensitive to elevated source voltages and temperatures (Loo et al., 1998b). Another study of inhibitors of carbonic anhydrase found that, although hydrophobic interactions were critical in solution, the gas-phase complexes relied upon polar interactions from electrostatic groups and hydrogen bonding (Wu et al., 1997). In contrast, interactions between two dimeric structures were retained in the gas-phase when the interaction was known to be largely hydrophobic (Nettleton et al., 1998). Other studies have shown a close correlation between the behaviour of complexes in solution and their gas phase behaviour (Smith and Light-Wahl, 1993; Loo,
Przybylski and co-workers found a correlation between solution stability and gas-phase stability of oligomeric tail domains of the p24 family of transmembrane proteins (Fligge et al., 2000). The gas phase behaviour of HIV-1 Tat peptide binding to TAR RNA also reflected condensed phase interactions (Sannes-Lowery et al., 1997). The interactions between Tat and TAR are primarily electrostatic and were unable to be dissociated by CAD. This exemplified that electrostatic interactions may be stronger in the gas phase owing to the absence of solvent-mediated interactions (Sannes-Lowery et al., 1997). These observations emphasize the importance of careful consideration of the results obtained by mass spectrometry and their relation to the behaviour of complexes in the condensed phase. Continuing studies on different systems will further our understanding of the behaviour of complexes in the gas phase.

The observation of a lower number of charges and a smaller range of charges in mass spectra that are acquired under non-denaturing conditions suggests that there is a difference in the structure of a protein from denaturing conditions to non-denaturing conditions (Smith and Zhang, 1994; Loo, 1997; Benjamin et al., 1998). The extent to which the complex retains condensed phase structure is unknown. The observation of protein multimers in the gas phase also supports the retention of at least some secondary and tertiary structure that is required for quaternary structure (Smith and Light-Wahl, 1993; Przybylski and Glocker, 1996). This observation, however, does not provide evidence that the condensed phase structure is maintained in the gas phase ions. There is also the possibility that the structures able to form in the gas phase are more diverse than in solution. The importance of solvent molecules in the condensed phase structure of biomolecules is well established and the effect of desolvation is yet to be clarified. Kaltashov and Fenselau examined the stability of the α helical conformation of a protein in the gas phase and the results suggested that the hydrogen bonds of the helix
enabled the structure to be maintained (Kaltashov and Fenselau, 1997). Further studies with β-pleated sheets indicated that multistrand β-sheets retained a structure somewhat like the native structure but single β-strands collapsed in the absence of solvent (Li et al., 1998a). Loo et al. used H/D exchange and CAD to probe the structures of a small native peptide and its synthetic counterpart (Loo et al., 1998a). They reported retention of the unique higher order structure of each of the peptides. Although these peptides were quite small (16 amino acids in length) and possibly displayed exceptionally stable structures, the results showed the potential of these techniques for the investigations of higher order structures.

In summary, there remains conflicting data and opinions as to whether ESI-MS results are useful for the interpretation of condensed phase structure and mechanisms. It is clear, however, that ESI-MS has an important role in structural protein chemistry and the knowledge gained from the behaviour of molecules in the gas phase is valuable, regardless of whether it precisely reflects condensed phase behaviour. The continued analysis of model protein systems will assist in the understanding of mechanisms that operate in the gas phase. The establishment of a library of information on the behaviour of different types of complexes under ESI-MS versus solution conditions would also be beneficial. This information could assist in the interpretation of ESI-MS data from systems for which there is little structural information from condensed phase studies.

5.1.3 Glutathione S-Transferases

Discovery and nomenclature

The purification of an enzyme from rat liver that catalysed the conjugation of glutathione with various compounds was reported in 1961 (Booth et al., 1961). It was previously recognised that this type of conjugation was critical for the excretion of
xenobiotics through urine. These discoveries lead to many years of research on, what were later named the glutathione S-transferases or GSTs.

GSTs have been found in both cytosolic and microsomal (membrane-bound) forms but the cytosolic forms have been characterised more extensively. At least seven different classes of cytosolic GST exist, based upon gene structure, sequence homology and substrate specificity. These classes are named alpha (A), pi (P), theta (T), sigma (S), mu (M), kappa (K) and zeta (Z) (Oakley et al., 1999; Gustafsson et al., 1999; Nieslanik and Atkins, 2000). Different isoenzymes also exist within a class of GST and these are designated by Arabic numerals.

The active structure of the cytosolic enzymes is dimeric and the monomeric units may be identical or from proteins within the same class (Armstrong, 1997; Cameron et al., 1995). Nomenclature for denoting the identification of a particular dimer has been developed (Armstrong, 1997). A dimer formed between isoenzyme type 1 and isoenzyme type 2 from the pi class is denoted by P1-2. A homodimer is denoted in the same way. For example, a homodimer of isoenzyme 1 from the alpha class is designated A1-1. Four different GST sequences were used in this work. Two native proteins were used, A1 and P1. The remaining two proteins were the result of single amino acid substitutions in the sequence of the isoenzyme A1. The first of these substitutions was a cysteine to serine at position 112 (C112S). The second was a tyrosine to phenylalanine substitution at position 9 (Y9F). These proteins would normally be denoted A1(C112S) and A1(Y9F). For simplicity GST A1 will be denoted A1, GST A1(C112S) will be denoted as A1CS and GST A1(Y9F) will be denoted A1YF. The A1CS protein was used owing to its improved stability against oxidation. This substitution has no apparent effect on the properties of the enzyme (Professor Lu-Yun Lian, Biomolecular Science Department, UMIST, Manchester, U.K. personal communication). The A1YF modified
protein has reduced catalytic capabilities and has been widely used to probe the importance of the tyrosine residue at position 9 (see below).

Function of GSTs

Early work with GSTs recognised their importance in the production of mercapturic acids in the excretion of unwanted compounds from the body (Booth et al., 1961; Al-Kassab et al., 1963; Boyland and Williams, 1965; Habig et al., 1974). In this reaction, shown in Scheme 5.1, the conjugation of glutathione (γ-glutamylcysteinylglycine; GSH) to an electrophile is catalysed by GST. The glutamyl and glycyl portions are then enzymatically cleaved and acetylation by acetyl-CoA produces a mercapturic acid (Mathews and van Holde, 1990). Mercapturic acids are much more soluble, and in some circumstances less toxic, than the initial electrophiles (Cameron et al., 1995).

Scheme 5.1

In addition to the conjugation of glutathione to xenobiotic substrates, GST has other functions. For example, GST is able to catalyse the release of glutathione from a glutathione conjugate (Dietze et al., 1998). Examples of these types of reactions are, however, limited.

It is also known that GST binds to non-substrate lipophilic molecules such as bilirubin, steroids, fatty acids and drugs (Oakley et al., 1999; Dirr and Wallace, 1999). The binding of proteins to these types of anionic compounds was first recognised in proteins named ligandins. It was later noted that GSTs and ligandins were identical.
(Meister and Anderson, 1983). This ligandin function is believed to be important for storage and transport of these molecules (Sinning et al., 1993; Cameron et al., 1995). The GSTs also form covalent bonds with a variety of other reactive species but this covalent modification destroys the catalytic ability of the protein (Meister and Anderson, 1983).

**Structure of cytosolic GSTs**

Each monomer in a GST dimer has a molecular mass of between 23 and 28 kDa (Oakley et al., 1999). Each of these subunits in the dimer has two domains. The dimer contacts are primarily between domain I of one subunit and domain II of the other subunit (Armstrong, 1997). In the A, M and P classes these interactions form a lock-and-key motif (Sayed et al., 2000). The monomeric subunits from different classes do not have compatible interfaces and therefore do not form dimers (Armstrong, 1997).

There are two glutathione binding sites per GST dimer. These sites are termed the G-sites and are formed by domain I of each subunit (Allardyce et al., 1999). There are also two sites for the binding of the hydrophobic substrates, termed the H-sites. The H-sites are formed mainly by domain II. The G-site in different isoenzymes is similar, whereas there is much more variation in the H-site. Presumably the variation in H-sites improves the ability to bind to different substrates. The H-site of the P class enzymes is relatively open. The C-terminal helix found only in A class enzymes appears to cover the substrate in the H-site when glutathione is present in the G-site (Allardyce et al., 1999; Dirr and Wallace, 1999; Nieslanik et al., 1999). The H-site in these enzymes is therefore not nearly as open as in the P class enzymes. This C-terminal helix is also thought to increase the specificity for more hydrophobic substrates (Allardyce et al., 1999; Dirr and Wallace, 1999).
The position of glutathione in the G-site is very similar in the different classes of GSTs, even though individual interactions vary (Armstrong, 1997). The G-site relies heavily on electrostatic interactions and virtually all of the hydrogen bonding sites of glutathione are utilised (Sinning et al., 1993). The majority of H-site interactions are hydrophobic, but protein hydrogen bonds and hydration effects do enable interactions with polar atoms of a substrate (Micaloni et al., 2000; Cameron et al., 1995).

Another site in GSTs that is less well characterised is the ligandin site. It has been reported that there are two distinct non-substrate binding sites (Dirr and Wallace, 1999). The most widely recognised site is in the cleft formed between the two subunits of the dimer (Dirr and Wallace, 1999; Wallace and Dirr, 1999). The second site exists on each subunit and is close to the tryptophan residue at position 20 of GST A1 (Dirr and Wallace, 1999). The same group also reported that stabilisation of the C-terminal helix of A class enzymes appears to impact on both these non-substrate binding sites (Dirr and Wallace, 1999).

**Catalytic mechanism of GSH conjugation**

It has been well documented that the G-site of GSTs contains a serine or tyrosine residue that is critical for the catalysis of glutathione conjugation (Armstrong, 1997). In A, M, P and S classes a conserved tyrosine is utilised. The T class uses a serine residue. The hydroxyl group from these residues is believed to activate the thiol group of glutathione to promote nucleophilic attack on the hydrophobic electrophile (Allardyce et al., 1999). This activation occurs through a decrease in the pKa of the thiol group (Nieslanik and Atkins, 2000; Dietze et al., 1996b; Armstrong, 1997). In solution, the pKa of glutathione is approximately 9.3 but at the active site of GST the pKa is reduced to between 6.5 and 6.9 (Dietze et al., 1996b), such that the predominant species at physiological pH is the thiolate anion, GS⁻ (Dietze et al., 1996b; Armstrong, 1997). The
hydroxyl group stabilises the thiolate anion through a hydrogen bond (Cameron et al., 1995; Widersten et al., 1996; Oakley et al., 1999; Dietze et al., 1996a; Nieslanik and Atkins, 2000). Site-directed mutagenesis has shown that substitution of a phenylalanine in place of the tyrosine at position 9 in A1 results in a decrease in the rate of catalysis for the substrate chloro-2,4-dinitrobenzene ($k_{cat} 58s^{-1}$ for wild-type and $0.34s^{-1}$ for the Y9F mutant)(Allardyce et al., 1999).

In addition to this tyrosine residue, the A class enzymes also have a conserved arginine residue that further stabilises the glutathione thiolate (Widersten et al., 1996; Armstrong, 1997; Sinning et al., 1993). These two interactions are known as first sphere interactions since they are directly co-ordinated to the sulphur atom of glutathione (Armstrong, 1997). Additional interactions that are distal to the sulphur atom (so-called second sphere interactions) also stabilise the thiolate anion. An example of a second sphere interaction could include electrostatic effects introduced by dipoles from $\alpha$ helices (Armstrong, 1997). These types of interactions have not been as well documented as the first sphere interactions.

*Inhibition of GSTs*

GSTs have the ability to metabolise alkylating agents, which are used in cancer therapy, and have been implicated in the development of drug resistance to other therapeutics (Armstrong, 1997; Oakley et al., 1999). For this reason there has been significant interest in GST inhibitors. The drug sulfasalazine inhibits A, M and P class GSTs. It has also been shown to increase the cytotoxic effects of an alkylating agent in cancer cells that exhibited elevated levels of GST P1 (Oakley et al., 1999). A similar enhancement has been seen with the GST inhibitor ethacrynic acid (Armstrong, 1997). The administration of GST inhibitors has shown a cytotoxic effect on healthy cells and
therefore their therapeutic use is limited. Inhibitors have, however, been crucial in the establishment of the structure of these enzymes.

Glutathione conjugates (GSR) with hydrophobic R groups are potent inhibitors of GSTs. These types of inhibitors can occupy both the G-site and the H-site of GSTs, depending on the size of the R group. S-alkylglutathiones have been used characterise structures of GST. Examples include, X-ray crystallography of the S-nonylGSH in complex with P1 (Oakley et al., 1999), S-hexylGSH in complex with P1 (Oakley et al., 1997) and S-benzylGSH in complex with A1 (Sinning et al., 1993), NMR spectroscopy of S-hexylGSH in complex with M2 (McCallum et al., 2000) and fluorescence studies with S-alkylGSHs of increasing chain lengths (Dietze et al., 1996b). Three S-alkylGSHs were chosen for examining the behaviour of these complexes by mass spectrometry. These were S-propylGSH (GS-prop), S-hexylGSH (GS-hex) and S-decylGSH (GS-dec).

Previous mass spectrometric studies of glutathione S-transferase

A previous study of GST A1 by non-denaturing ESI-MS revealed the formation of a dimer of GST that was apparent in the gas phase (Ishigai et al., 2000). The binding of reduced glutathione (GSH), oxidised glutathione (GSSG) and two different inhibitors was also investigated. Complexes formed between dimeric A1 and the ligands was observed by ESI-MS. It was interesting that these investigators found the addition of GSH resulted in a mass increase of 340 Da rather than the molecular weight of GSH (307.3 Da). It was proposed that the additional mass was from the specific retention of water molecules with the ligand. Observations of water molecules in gas phase noncovalent complexes have been reported previously (Benjamin et al., 1998; Green et al., 1999; Fabris and Fenselau, 1999) but this water retention does not appear to be consistent in all complexes and may be dependent on how rapidly the water molecules are exchanged in the condensed phase (Loo et al., 1998b). The GST enzyme has been
widely investigated in solution and crystal structures and ESI-MS has revealed GST complexes in the gas phase. This system is therefore suitable for use as a model in the investigation of the gas phase behaviour of noncovalent complexes.

5.1.4 Aims

The aims of this project were to extend the results of the earlier study by Ishigai et al. on the non-covalent complexes of GST and seek to clarify some of the questions raised by that study. Furthermore, the GST enzyme has been well characterised by spectroscopic methods (including high resolution NMR), X-ray crystallography and kinetic studies and is therefore an ideal model system for investigating the behaviour of noncovalent complexes in the gaseous environment of a mass spectrometer.
5.2 RESULTS

5.2.1 Molecular mass determination, conventional electrospray conditions

The GST proteins used in this study were first analysed under denaturing nanoESI conditions (i.e. 50% aqueous acetonitrile with 0.2% formic acid) to confirm their molecular masses and assess the purity of the samples. The resulting spectra are shown in Figure 5.1 and the measured molecular masses are given in Table 5.1. The A class GSTs (Figure 5.1A-C) each showed a major series of peaks corresponding to protonated molecular ions with charges ranging from approximately 17+ to 39+. For the P class GST (Figure 5.1D) the charge distribution ranged from 16+ to 26+. In A1 and A1\textsuperscript{YF} an extra series of molecular ions was detected. The experimental mass of A1 was 25500.6 ± 0.5 Da, which was in good agreement with the theoretical mass of 25499.9 Da. The extra component in this sample was of lower abundance and it was therefore difficult to establish an accurate mass. The mass calculated was 25845 ± 0.5 Da, i.e. 345 Da above the native protein\textsuperscript{4}. The experimentally determined mass for A1\textsuperscript{YF} was 25483.3 ± 0.9 Da, very close to the theoretical mass of 25483.9 Da. The extra component in this spectrum was also of low abundance. The calculated mass of this extra component was 25336 ± 1 Da, which corresponds to a mass 147 Da lower than A1\textsuperscript{YF}. The spectrum of A1\textsuperscript{CS} did not appear to contain either of these extra components and the experimentally determined mass (25483.6 ± 0.3 Da) was once again in close agreement with the theoretical mass of 25483.9 Da.

The two modified forms of A1 have the same molecular mass, because both the cysteine-to-serine substitution and the tyrosine-to-phenylalanine substitution result in a difference in mass of 16 Da lower than the native protein. The modifications could

\textsuperscript{4} The origin of this extra component or adduct is not immediately obvious but the high abundance of the native protein suggests that the unknown adduct was a minor component of the sample and may be of little relevance.
therefore not be confirmed by mass alone, so the presence of these modifications was confirmed by tryptic digestion and MS/MS analyses of the resulting peptides (see 5.2.2 below).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Theoretical mass</th>
<th>Experimental mass</th>
<th>Other components</th>
</tr>
</thead>
<tbody>
<tr>
<td>GST A1(wild type)</td>
<td>25499.9</td>
<td>25500.6 ± 0.5</td>
<td>25845.3 ± 0.5</td>
</tr>
<tr>
<td>GST A1(C112S)</td>
<td>25483.9</td>
<td>25483.6 ± 0.3</td>
<td>-</td>
</tr>
<tr>
<td>GST A1(Y9F)</td>
<td>25483.9</td>
<td>25483.3 ± 0.9</td>
<td>25336 ± 1</td>
</tr>
<tr>
<td>GST P1</td>
<td>23224.7</td>
<td>23225.8 ± 0.3</td>
<td>23356.8 ± 0.2</td>
</tr>
</tbody>
</table>

Table 5.1 Theoretical and experimental molecular masses of glutathione S-transferases. The third column lists extra components that were detected in the spectra.

The experimentally determined mass of the most abundant species in the P1 spectrum (Figure 5.1D) was 23356.8 ± 0.3 Da. A second component was determined to have a mass of 23225.8 ± 0.2 Da, which was in closer agreement with the theoretical molecular mass of 23224.7 Da. The difference between the two components was 131 Da. The residue mass of the amino acid methionine is also 131 Da. The sequence of GST P1 has an N-terminal initiator methionine, which is normally cleaved to give the native protein sequence (Swiss Prot, accession number P09211, http://au.expasy.org/sprot). To determine whether the component of mass 23356.8 Da was due to the presence of an initiator methionine, tryptic digestion of the protein and MS/MS analysis of the resulting peptides was also performed (see 5.2.2 below). The 20+ charge state in Figure 5.1 is labelled with ‘P1’ and ‘P1Met’ to indicate the two different components.
Figure 5.1 ESI-MS spectra of GST under denaturing conditions (50% CH₃CN (aq) with 0.1% formic acid). Protein concentration was between 1 and 5 pmol/µL.

A) wild-type A1, cone voltage 40V; B) A1^{CS}, cone voltage gradient of 25-120V from m/z 600-2200; C) A1^{YF}, cone voltage 40V; D) P1, cone voltage gradient of 25-120V from m/z 600-2200.
5.2.2 Confirmation of modifications in glutathione S-transferases

The three proteins AlCS, AlYF and P1 were cleaved with trypsin and the resulting peptides were analysed by nanoESI-MS and MS/MS procedures. The sequences of Al and P1 are given in Figures 5.2 and 5.3, respectively. The tryptic peptide T15 from the digest of Al (Figure 5.2) contained cysteine 112, which is replaced with a serine residue in the AlCS protein. The mass of this peptide is 3084.58 Da for the wild-type Al protein but with the C112S modification the peptide molecular mass is 3068.61 Da. From the ESI-MS spectrum of the digest of A1CS (data not shown), a triply charged precursor at m/z 1023.86 was chosen for MS/MS. This ion had a measured molecular mass of 3068.56 Da, which is in close agreement with the expected mass for T15. The MS/MS spectrum of the m/z 1023.86 precursor is shown in Figure 5.4. The ion at m/z 686.36 corresponded to Y6" (the last six amino acids of the peptide, i.e. SPPEEK). The ion at m/z 599.32 corresponded to Y5", i.e. the C-terminal sequence PPEEK. These ions confirmed the presence of a serine in place of the cysteine residue at position 112.

The substitution of a phenylalanine for tyrosine 9 in Al resulted in a mass shift of the T2 tryptic peptide from Al (Figure 5.2). T2 from the wild-type protein has a molecular mass of 919.47 Da, whereas T2 from the protein AlYF has a molecular mass of 903.47 Da. ESI-MS of a tryptic digest of A1YF showed a doubly charged ion a m/z 453.2, which corresponds to a molecular mass of 903.4 Da (data not shown). The MS/MS spectrum of this precursor is shown in Figure 5.5. This small peptide had an expected sequence of LHFFNAR (Figure 5.2). The full sequence was confirmed with the Y ion and complementary B ion series. Figure 5.5 shows the Y ion series and the mass difference between Y5" and Y4", which corresponds to phenylalanine, is labelled ‘F’.
Figure 5.2 The sequence of GST A1 with peptides generated from a theoretical digest with trypsin labelled (T1 to T35). The initiator methionine, which is considered amino acid 1 in the literature, is not pictured so the first amino acid is labelled 2. The tyrosine at position 9 and the cysteine at position 112 are labelled as these are the two positions which were modified to give A1^{YF} and A1^{CS}. These modifications are contained in tryptic peptides T2 and T15, respectively.
**Figure 5.3** The sequence of GST P1 with peptides generated from a theoretical digest with trypsin labelled (T1 to T19). The initiator methionine, which is considered amino acid 1 in the literature, is not pictured so the first amino acid is labelled 2.
Figure 5.4 MS/MS spectrum of m/z 1023.86 (3+) (T15) from a tryptic digest of Al\textsuperscript{CS}. Only the region below the precursor is displayed. The Cys→Ser modification at position 112 was confirmed from the Y ion series shown. The difference between Y\textsubscript{5}'' and Y\textsubscript{6}'' corresponded to the serine residue of interest.

Figure 5.5 MS/MS spectrum of m/z 453.2 (2+) (T2) from a tryptic digest of Al\textsuperscript{YF}. The Tyr→Phe modification at position 9 was confirmed from the Y ion series shown. The difference between Y\textsubscript{4}'' and Y\textsubscript{5}'' corresponded to the phenylalanine residue of interest.
Figure 5.6 MS/MS spectra for confirmation of the P1 N-terminal initiator methionine.
A) MS/MS spectrum of m/z 669.4 (2+) from a tryptic digest of P1.
B) MS/MS spectrum of m/z 734.9 (2+) from a tryptic digest of P1.
The region above m/z 1250 is enlarged by a factor of 10 in both spectra. See text for details.
The main species in the spectrum of P1 acquired under denaturing conditions was proposed to be from the intact protein with the initiator methionine present at the N-terminus. The mass of the N-terminal tryptic peptide (T1, Figure 5.3) of P1 without the initiator methionine is 1336.72 Da. With the initiator methionine present, T1 has a mass of 1467.76 Da. ESI-MS of a digest of the P1 sample (data not shown) showed doubly charged ions at m/z 734.9 and m/z 669.4, which correspond to peptides of masses 1467.8 and 1336.8 Da, respectively. The MS/MS spectra of these two precursors are shown in Figure 5.6. The region above m/z 1250 in Figure 5.6B clearly shows the ion at m/z 1337.8, which corresponds to Y_{11}" of the sequence MPPYTVVYFPVR. The ion at m/z 1240.7 in both spectra corresponds to Y_{10}" , which is the same for each peptide. The immonium ion of methionine (m/z 104) was present only in the MS/MS spectrum of the m/z 734.9 ion. Another difference between the spectra was the peak at m/z 229.1 from the precursor of m/z 734.9, which corresponded to the B_2 ion (MP) of MPPYTVVYFPVR. The presence of the initiator methionine was therefore clearly evident from these spectra.

5.2.3 Non-denaturing electrospray ionisation mass spectrometry of glutathione S-transferases

The ability to transfer noncovalent complexes to the gas phase for analysis required careful control of the mass spectrometry ionisation conditions. Specific tuning parameters on the QTof, such as the focus voltage and pusher time were set the same each time (see Chapter 2, Materials and Methods). An elevated pressure in the source was achieved by partially closing the in-line isolation valve of the rotary pump at the start of each new session on the instrument. To determine the appropriate level to which the pumping needed to be reduced, a solution of caesium iodide (2 mg/mL) was sprayed
and monitored in the peak window of the MassLynx™ software. The rotary pumping
was fine-tuned to maximise the peaks visible in the region m/z 3000-3600 without
introducing excessive background noise. Once the pressure was set, caesium iodide was
used to calibrate the instrument.

NanoESI⁵ mass spectra of the each of the GSTs under non-denaturing conditions
(i.e. 10mM ammonium acetate, pH 7.4) are shown in Figure 5.7 over the m/z range
2800 to 3700. For the A class GSTs protonated dimers with charges from 14+ to 17+
were observed. For the P class enzyme the 13+ ion was more abundant and the 17+ ion
was not observed (Figure 5.7D). The ESI-MS spectrum of A1 (Figure 5.7A) showed the
most abundant species was the dimer (A1-1). An additional series of ions was apparent,
which yielded a mass 337 Da higher than the GST dimer. A third series, which appeared
to be heterogeneous, was also evident but the low abundance and heterogeneity meant it
was difficult to accurately measure the mass of this species; however, this series most
likely arose from the addition of two molecules of mass 337 Da. The mass spectrum
previously reported for this protein also showed an additional component of
approximately 340 Da (Ishigai et al., 2000).

The mass spectra of the modified forms of GST A1 showed fewer components
than the wild-type enzyme. In the spectrum of A1⁵F (Figure 5.7C) a small amount of
another dimeric species was apparent with a mass 147 Da below the dimer of A1⁵F.
This may be attributed to the formation of a dimer between A1⁵F and a truncated form
of the same protein, seen in the ESI-MS spectrum under denaturing conditions (Figure
5.1, section 5.2.1). Dimers formed between two molecules of the truncated protein were

---

⁵ NanoESI was used throughout this work. From herein ESI therefore refers to nanoESI. ‘Standard ESI’ refers to
nanoESI performed under non-denaturing conditions.
Figure 5.7 Non-denaturing ESI-MS spectra of A) wild-type A1, B) \( A1^{CS} \), C) \( A1^{YF} \) and D) P1. Protein concentration 5pmol/mL in 10mM NH₄CH₃COO. The charge on the ions is shown above each group of peaks.
not apparent, presumably because of the relatively low abundance of this form of the protein.

The presence of two proteins that differed only by an N-terminal methionine in the PI protein sample was evident from ESI-MS spectra obtained under standard (denaturing) conditions (Figure 5.1 and 5.4). With these two forms present, three possible dimers could be formed. The spectrum of PI acquired under non-denaturing conditions (Figure 5.7D) clearly showed three dimers of different masses, namely 46449 ± 1 Da, 46583.0 ± 0.7 Da and 46713 ± 2 Da, which correspond to the PI/PI, PI/PIMet and PIMet/PIMet dimers successively. The abundance of each of these dimers in the non-denaturing spectrum reflected the abundance indicated by denaturing ESI-MS (Figure 5.1D). The PIMet protein was the most abundant, and dimers of this protein gave rise to the most abundant ion series in Figure 5.7D. The second most abundant series was from dimers formed between PI and PIMet, whereas the PI/PI dimer was of lowest abundance.

5.2.4 Addition of glutathione to GSTs

To simplify the interpretation and discussion, the complexes reported herein are represented by the notation [dimer + ligand]. Spectra from ESI-MS under non-denaturing conditions of a mixture of glutathione (GSH) with A1 in molar ratios of 1:1 and 1:5 are shown in Figure 5.8B and C. Without added GSH, the spectrum of A1 showed three ion series (Figure 5.8A), which corresponded to the dimer and two additional adducts. The presence of the additional adducts in this spectrum (b and c) was addressed above and was also noted by Ishigai et al (Ishigai et al., 2000). The addition of GSH to A1 resulted in the disappearance of the ions from b and c (Figure 5.8B) and the appearance of two new ion series (Figure 5.8C). These new ion series
corresponded to the addition of one and two molecules of GSH to the A1 dimer (series d and e respectively, Figure 5.8C). The mass added to the dimer from the addition of GSH was 307 Da (Figure 5.8C). Hence, the GSH effectively displaced the adduct of 337 Da, providing evidence that this ion series arose from an adduct with a small molecule rather than a modified protein.

The addition of the reduced form of glutathione (GSH) to A1 has been investigated previously and it was found that when GSH was added three complexes were apparent (Ishigai et al., 2000). The dimeric form of A1 was the most abundant species. The second component had a molecular mass that was approximately 340 Da above that of the dimer. The third component appeared to be heterogenous, and was possibly partly from the addition of two 340 Da entities to the dimer. As noted above, the mass of GSH is 307.3 Da, therefore to account for the difference in mass they observed, Ishigai et al. proposed that the additional mass arose from the presence of water molecules (Ishigai et al., 2000).

The spectra shown in Figure 5.9 are from non-denaturing ESI-MS of GSH mixed with each of the A class GSTs in a five molar excess of GSH. In each of the spectra, the dimer was apparent, along with the addition of multiple GSH molecules (307.3 Da). The most abundant component in the spectrum of A1 with GSH (Figure 5.9A) was the dimer of A1 without any bound GSH. The most abundant component in the spectrum of the mixture of A1CS and GSH was the complex [A1CS dimer + GSH] (series b, Figure 5.9B). The spectrum of the mixture of A1YF and GSH revealed [A1YF dimer + 2GSH] (series c, Figure 5.9C) to be the most abundant complex. The presence of the second dimer in the spectrum of A1YF was also apparent as a small ion series in-between the abundant ions from the major series.
Figure 5.8 ESI-MS spectra of A1 (wild-type) + GSH under non-denaturing conditions
A) wild-type A1 without added GSH, B) A1 + GSH (molar ratio 1:1), C) A1 + GSH (molar ratio 1:5).
In each of the spectra the lower case letters correspond to the following: a: [A1 dimer], b: [A1 dimer + 337 Da], c: [A1 dimer + approximately 600 to 700 Da], d: [A1 dimer + GSH], e:[A1 dimer + 2GSH].
Figure 5.9 ESI-MS spectra of GST + GSH (molar ratio 1:5) under non-denaturing conditions
A) wild-type A1 + GSH, B) A1<sup>CS</sup> + GSH, C) A1<sup>YF</sup> + GSH
In each of the spectra the lower case letters correspond to the following: a: [dimer], b: [dimer + GSH], c: [dimer + 2GSH], d: [dimer + 3GSH], e: [dimer + 4GSH]
The differences in the ion abundances seen in the spectra in Figure 5.9 may be attributed to different concentrations of protein in the samples. The concentrations of A1, A1\textsuperscript{YF} and P1 were estimated via different methods (see Chapter 2) prior to desalting. A1\textsuperscript{CS} was received as a lyophilised powder and did not require desalting. For this reason, the concentration was most accurate for A1\textsuperscript{CS} whereas for the other proteins the concentration may have varied from that estimated. Nevertheless, it is interesting to note that, with a molar excess of GSH, the dimer with two molecules of GSH was not generally the most abundant complex.

Spectra from mixtures of GSH and P1 were complicated by the presence of the three different dimers, P1/P1, P1/P1Met and P1Met/P1Met, seen in Figure 5.7. The ESI mass spectrum of P1 and GSH in a five molar excess of GSH showed complexes with each of the three dimers and GSH (data not shown). The two most abundant ion series were from the dimers of P1/P1Met and P1Met/P1Met. The next most abundant complex was [P1Met/P1Met + GSH]. Even with the five-fold excess, the complexes formed between the dimer and only one molecule of GSH remained the most abundant components.

5.2.5 Addition of S-alkylglutathione inhibitors to GSTs

Three different S-alkylglutathiones were used to investigate the binding of inhibitors to GSTs. S-alkylglutathiones were chosen to prevent covalent adduct formation through the free thiol group of the GSH moiety. The ESI-MS spectra of a mixture of A1\textsuperscript{CS} and S-propylglutathione (GS-prop) under non-denaturing conditions are presented in Figure 5.10. With a molar ratio of 1:1 (A1\textsuperscript{CS}:GS-prop) the three ion series present resulted from the complexes [A1\textsuperscript{CS}dimer] (series a, Figure 5.10A), [A1\textsuperscript{CS}dimer + GS-prop] (series b, Figure 5.10A) and [A1\textsuperscript{CS}dimer + 2GS-prop] (series c, Figure 5.10A). The spectrum
from the mixture with a five-fold excess of GS-prop (Figure 5.10B) showed an increase in the abundance of $[\text{A}1^{\text{CS}}\text{dimer} + 2\text{GS-prop}]$ and the appearance of a fourth component, which corresponded to the complex $[\text{A}1^{\text{CS}}\text{dimer} + 3\text{GS-prop}]$ (series d, Figure 5.10B). The dimer without any bound GS-prop was of very low intensity in Figure 5.10B and the most abundant complex was $[\text{A}1^{\text{CS}}\text{dimer} + 2\text{GS-prop}]$.

Figure 5.11 shows ESI-MS spectra from the addition of GS-hex to the A class enzymes in a molar ratio of 1:5 (A1:GS-hex). In each of the spectra, the most intense series of peaks was from the complex $[\text{dimer} + 2\text{GS-hex}]$. A smaller series was also apparent from the complex $[\text{dimer} + 3\text{GS-hex}]$. In the spectrum of GSH with A1$^{\text{YF}}$ (Figure 5.11C) a third component of relatively low abundance was also visible. This component corresponds to the dimer between A1$^{\text{YF}}$ and truncated A1$^{\text{YF}}$ with two bound molecules of GS-hex. The addition of GS-hex to the GSTs revealed a very clear ion series, with $[\text{dimer} + 2\text{GS-hex}]$ proving to be the most abundant complex, even at low molar ratios of ligand to GST.

The ESI-MS spectra of GS-dec with A1$^{\text{CS}}$ and GS-dec with A1$^{\text{YF}}$ acquired under non-denaturing conditions are presented in Figure 5.12. The complex $[\text{dimer} + 2\text{S-dec}]$ was by far the most abundant component in both spectra, with only a relatively small amount of the complex $[\text{dimer} + 3\text{GS-dec}]$ apparent. Ion series from the dimer alone and the complex of dimer with one molecule of GS-dec were not apparent in these spectra.
Figure 5.10 ESI-MS spectra of GST A1^{CS} + GS-prop under non-denaturing conditions. A) A1^{CS} + GS-prop (molar ratio 1:1), B) A1^{CS} + GS-prop (molar ratio 1:5)
In each of the spectra the lower case letters correspond to the following: a: [A1^{CS}dimer], b: [A1^{CS}dimer + GS-prop], c: [A1^{CS}dimer + 2GS-prop], d: [A1^{CS}dimer + 3GS-prop]
Figure 5.11 ESI-MS spectra from the addition of GS-hex to each of the alpha class enzymes under non-denaturing conditions. The ion charge is displayed at the top of the figure and the 16+ molecular ions are labelled with the identity of the complex.

A) wild-type A1 + GS-hex (molar ratio 1:5)
B) A1CS + GS-hex (molar ratio 1:5)
C) A1YF + GS-hex (molar ratio 1:5)
Figure 5.12 Non-denaturing ESI-MS from the addition of GS-dec to the two modified A class enzymes. The most abundant ion series in both spectra correspond to the complex [dimer + 2GS-dec].

**Spectrum A:** A1^YF + GS-dec (molar ratio 1:5)

**Spectrum B:** A1^CS+ GS-dec (molar ratio 1:5)
5.2.6 Competition experiments

The binding properties of S-alkylGSH inhibitors were further assessed in competition experiments in which two different inhibitors were added to GST and ESI-MS analysis of the mixtures was performed under non-denaturing conditions. Spectra acquired from a mixture in which GS-hex was added to a solution already containing A1CS and GSH are shown in Figure 5.13. These spectra were obtained from mixtures with increasing molar ratios of GS-hex over the protein. The spectrum in Figure 5.13A was acquired when the enzyme (A1CS) was in excess over the inhibitor (GS-hex) and the substrate (GSH) was in excess over the enzyme. Even with this deficiency of GS-hex, the most abundant ion series was from the complex \([A1CS\text{dimer} + 2\text{GS-hex}]\) (series c, Figure 5.13A). The complexes \([A1CS\text{dimer} + \text{GS-hex}]\) (series a) and \([A1CS\text{dimer} + \text{GSH} + 2\text{GS-hex}]\) (series d) were also significantly abundant with the deficiency of GS-hex (Figure 5.13A). The ion \([A1CS\text{dimer} + \text{GSH} + \text{GS-hex} + 16\text{H}^+]^{16+}\) was the only ion apparent from the complex \([A1CS\text{dimer} + \text{GSH} + \text{GS-hex}]\) (series b, Figure 5.13A).

When the molar ratio of GS-hex was increased, the spectra showed a decrease in the relative abundance of series a and b (Figure 5.13B and C). The spectrum from a molar ratio of 1:10:10 (A1CS:GSH:GS-hex) showed that the most abundant component was again \([A1CS\text{dimer} + 2\text{GS-hex}]\). Additional ion series were visible in this spectrum that were not apparent at the lower molar ratios of GS-hex. These were from the complexes \([A1CS\text{dimer} + 3\text{GS-hex}]\) (series e), \([A1CS\text{dimer} + \text{GSH} + 3\text{GS-hex}]\) (series f) and \([A1CS\text{dimer} + 4\text{GS-hex}]\) (series g) (Figure 5.13C). In each of the spectra in Figure 5.13 the relatively high abundance of \([A1CS\text{dimer} + 2\text{GS-hex}]\) over the other complexes was very clear and suggested preferential binding of GS-hex over GSH.
Figure 5.13 ESI-MS spectra from mixtures of $A1^{CS}$, GSH and GS-hex acquired under non-denaturing conditions.

A) $A1^{CS}$ + GSH + GS-hex (molar ratio 2:20:1)
B) $A1^{CS}$ + GSH + GS-hex (molar ratio 1:10:1)
C) $A1^{CS}$ + GSH + GS-hex (molar ratio 1:10:10)

In each of the spectra the lower case letters correspond to the following: a: $[A1^{CS}\text{dimer} + \text{GS-hex}]$, b: $[A1^{CS}\text{dimer} + \text{GSH} + \text{GS-hex}]$, c: $[A1^{CS}\text{dimer} + 2\text{GS-hex}]$, d: $[A1^{CS}\text{dimer} + \text{GSH} + 2\text{GS-hex}]$, e: $[A1^{CS}\text{dimer} + 3\text{GS-hex}]$, f: $[A1^{CS}\text{dimer} + \text{GSH} + 3\text{GS-hex}]$, g: $[A1^{CS}\text{dimer} + 4\text{GS-hex}]$. 
Figure 5.14 shows spectra from three different mixtures of GS-prop, GS-hex and A1\textsuperscript{CS}. The relatively high abundance of the complex [A1\textsuperscript{CS}dimer + 2GS-hex] was again apparent (ion series e), even with a deficiency of GS-hex (Figure 5.14A). The complexes [A1\textsuperscript{CS}dimer + GS-prop] (series b) and [A1\textsuperscript{CS}dimer + GS-hex] (series c), [A1\textsuperscript{CS}dimer + GS-prop + GS-hex] (series d) and [A1\textsuperscript{CS}dimer + 2GS-prop + GS-hex] (series f) were apparent at a molar ratio of 2:15:1 (A1\textsuperscript{CS};GS-prop:GS-hex) (Figure 5.14A). These complexes were not evident in the spectra acquired from the mixtures with higher molar ratios of GS-hex (Figure 5.14B and C).

The complex [A1\textsuperscript{CS}dimer + GS-prop + 2GS-hex] (series g) was seen in the spectra of each of the mixtures of GS-prop, GS-hex and A1\textsuperscript{CS}. A second complex with three ligands bound to the dimer was [A1\textsuperscript{CS}dimer + 3GS-hex], which was apparent at a molar ratio of 1:5:5 (A1\textsuperscript{CS};GS-prop:GS-hex) (series h, Figure 5.14C). Since there are only two sites for ligand binding on the A1\textsuperscript{CS} dimer, the presence of a third ligand may arise from non-specific binding.

Figure 5.15 shows the ESI-MS spectra of a mixture of GS-hex, GS-dec and A1\textsuperscript{YF} (Figure 5.15A) and a mixture of GS-hex, GS-dec and A1\textsuperscript{CS} (Figure 5.15B). Both the spectra were acquired with a molar ratio of 1:5:5 (protein:GS-hex:GS-dec). The most abundant ion series in these spectra was from the complex [dimer + 2GS-dec] (series c). Also of high relative abundance was the ion series from the complex [dimer + GS-hex + GS-dec] (series b). The complex [dimer + 2GS-hex] was also apparent (series a) but this complex was not nearly as abundant as in the previous competition experiments (Figures 5.13 and 5.14). The complexes [dimer + 2GS-hex + GS-dec] (series d), [dimer + GS-hex + 2GS-dec] (series e) and [dimer + 3GS-dec] (series f) were also apparent and were possibly a result of non-specific binding of a third inhibitor molecule to the dimer.
Figure 5.14 ESI-MS spectra from mixtures of $A1^{CS}$, GS-prop and GS-hex acquired under non-denaturing conditions.

A) $A1^{CS}$ + GS-prop + GS-hex (molar ratio 2:15:1)
B) $A1^{CS}$ + GS-prop + GS-hex (molar ratio 1:5:1)
C) $A1^{CS}$ + GS-prop + GS-hex (molar ratio 1:5:5)

In each of the spectra the lower case letters correspond to the following: a: [$A1^{CS}$ dimer], b: [$A1^{CS}$ dimer + GS-prop], c: [$A1^{CS}$ dimer + GS-hex], d: [$A1^{CS}$ dimer + GS-prop + GS-hex], e: [$A1^{CS}$ dimer + 2GS-hex], f: [$A1^{CS}$ dimer + 2GS-prop + GS-hex], g: [$A1^{CS}$ dimer + GS-prop + 2GS-hex], h: [$A1^{CS}$ dimer + 3GS-hex].
Figure 5.15 ESI-MS spectra of the mixture resulting from the addition of GS-hex and GS-dec to the modified A-class enzymes (A1^{CS} and A1^{VF}) under non-denaturing conditions. The charge on the ions are displayed at the top of the figure.

A) A1^{VF} + GS-hex + GS-dec (molar ratio 1:5:5)
B) A1^{CS} + GS-hex + GS-dec (molar ratio 1:5:5)

In each of the spectra the lower case letters correspond to the following: a: [dimer + 2GS-hex], b: [dimer + GS-hex + GS-dec], c: [dimer + 2GS-dec], d: [dimer + 2GS-hex + GS-dec], e: [dimer + GS-hex + 2GS-dec], f: [dimer + 3GS-dec].
5.2.7 MS/MS of glutathione S-transferase noncovalent complexes

MS/MS is useful for a number of purposes in the analysis of noncovalent complexes, particularly when the identity of components in a complex may not be clear from the mass of the complex alone. Using MS/MS, the complex can be dissociated into its component parts to identify the separate binding partners. MS/MS can also be used to exclude any covalent modifications to proteins. If a covalent modification were present, gentle CAD would be likely to disrupt noncovalent interactions first, leaving the covalent bonds intact. An example of this is presented below.

Spectra from MS/MS of the 16+ ion of the A1CS dimer (i.e. [A1CS dimer + 16H+]16+) are displayed in Figure 5.16. With a voltage of 40 V applied to the collision cell there was no apparent fragmentation of the precursor (Figure 5.16A). Increasing the collision cell voltage resulted in the appearance of two different ion series. The first, consisting of the peaks at m/z 2125, 2318 and 2550, corresponded to the monomeric form of A1CS. The second series was from a fragment of monomeric A1CS (peaks at m/z 1627, 1860 and 2170) that had a molecular mass of 13012.8 ± 0.5 Da. Ishigai et al. also reported a fragment of approximately 13 kDa in the MS/MS spectrum of A1 (Ishigai et al., 2000). Allowing for a 5 Da uncertainty in the mass, there are 11 possible regions of A1CS to which the fragment could correspond. The identity of this fragment could not be confirmed by the molecular mass alone but known properties of peptide fragmentation allowed the probable identity of the fragment to be proposed. One of the possible fragments was from amino acids 109-221 with a molecular mass 13010.3 Da. This corresponds to the C-terminal portion of the protein after cleavage adjacent to the proline residue at position 109. Preferential cleavage adjacent to proline residues has been documented previously (Loo et al., 1993; Loo et al., 1998b). The fragment seen by Ishigai et al. was also proposed to be a result of fragmentation at proline 109 (Ishigai et
Figure 5.16 Spectra from MS/MS of $[\text{Al}^{\text{CS}} \text{ dimer} + 16\text{H}^+]^{16+}$ (m/z 3187).
Collision offset A) 40V, B) 60V, C) 80V, D) 100V
This region includes the C112S modification and therefore the fragment mass was different to that observed by Ishigai et al.

Figure 5.17 shows MS/MS spectra of the 16+ ion of the complex \([\text{Al}^{\text{YF}}\text{dimer} + 2\text{GS-hex}]\) at different collision cell voltages. Each inset displays an enlarged region containing the peaks observed in the region above the precursor. The peak at \(m/z\) 3425 was from the precursor with one less proton and the loss of one GS-hex molecule. The peak at \(m/z\) 3399 was from the precursor with one less proton and the loss of two GS-hex molecules, i.e. the 15+ ion of the dimer stripped of ligands. As can be seen from the spectra, the loss of one GS-hex (\(m/z\) 3425) was apparent at a collision cell voltage of 40 V (Figure 5.17A). At 50 V the loss of one (\(m/z\) 3425) and two (\(m/z\) 3399) GS-hex molecules was apparent (Figure 5.17B). In the spectra acquired at collision cell voltages above 50 V, the 15+ ion from the loss of only one GS-hex (\(m/z\) 3425) was no longer evident. The spectra also showed a similar trend for the 16+ ions. At a collision cell voltage of 50 V, the loss of both one and two GS-hex molecules was apparent (labelled w and v, respectively, Figure 5.17B). Increasing collision cell voltages resulted in the disappearance of the peak corresponding to the loss of one GS-hex from the complex and an increase in the \(\text{Al}^{\text{YF}}\) dimer stripped of ligands (v, Figure 5.17C-E). In the spectra acquired with collision cell voltages of 50 V and higher, peaks from the monomeric form of \(\text{Al}^{\text{YF}}\) were apparent (\(m/z\) 2125, 2318, 2550 and 2833). These were equivalent to those seen in the MS/MS spectra of the \(\text{Al}^{\text{CS}}\) dimer (Figure 5.16), which is of the same mass as \(\text{Al}^{\text{YF}}\). It is important to note that there was no evidence of the monomer with GS-hex bound. This suggested that the GS-hex only binds to the protein in its dimeric form. The spectrum acquired at a collision cell voltage of 80 V also showed peaks from a fragment of \(\text{Al}^{\text{YF}}\) (\(m/z\) 1645 and 1862, Figure 5.17E). The measured molecular mass of the fragment was 13028.9 ± 0.6 Da, which was of higher mass than the fragment.
Figure 5.17 Spectra from MS/MS of [AlF\textsuperscript{Y}dimer + 2GS-hex + 16H\textsuperscript{+}]\textsuperscript{16+} (m/z 3235.5). Collision offset A) 40V, B) 50V, C) 60V, D) 70V, E) 80V. The insets show an enlarged region above the precursor. The labels v and w correspond to [Al\textsuperscript{Y}F\textsubscript{dimer}] and [Al\textsuperscript{Y}F\textsubscript{dimer} + GS-hex].
observed in the MS/MS spectra of the $A1^{CS}$ dimer but very similar to that seen by Ishigai et al. (13026.4 Da) (Ishigai et al., 2000). This is in agreement with the proposed fragmentation at proline 109. The $A1^{YF}$ protein has the same sequence as the wild-type $A1$ from position 109 to the C-terminus, whereas the $A1^{CS}$ protein contains the modification at position 112. To confirm this proposed sequence, further MS/MS experiments would be required.

MS/MS was also used to confirm the identity of complexes in mixtures of two ligands with the dimer. The MS/MS spectra in Figure 5.18 were acquired from the precursor at m/z 3255 that was apparent in the spectrum of a mixture of $A1^{CS}$, GSH and GS-hex (Figure 5.13). The insets show an enlarged section from m/z 3170 to m/z 3275. In each of the spectra, except at a collision cell voltage of 50 V, the precursor was clearly visible. The first peak to become apparent below the precursor was from the loss of GSH (307 Da) from the precursor (z, Figure 5.18A). This suggested that, in the gas phase, GSH was dissociated from the complex more readily than the other ligands. At collision cell voltages of 30 V and 40 V the following components were apparent: [$A1^{CS}$ dimer], [$A1^{CS}$ dimer + GSH], [$A1^{CS}$ dimer + GS-hex], [$A1^{CS}$ dimer + GSH + GS-hex] and [$A1^{CS}$ dimer + 2GS-hex]. The predominant peak in the spectrum at a collision cell voltage of 50 V was from the $A1^{CS}$ dimer without any attached ligands (m/z 3187, labelled v, Figure 5.18D). The masses of the fragments observed in the spectra and knowledge of the ligands present in the sample enabled assignment of each of the peaks and confirmation of the precursor identity, [$A1^{CS}$ dimer + GSH + 2GS-hex].

Figure 5.19 shows MS/MS spectra from a precursor of m/z 3239 that was assigned as the complex [$A1^{CS}$ dimer + GS-hex + GS-dec]. The insets show an expanded section including the precursor and a region below the precursor. The fragments produced from MS/MS were most apparent at a collision cell voltage of 50 V and
Figure 5.18 Spectra from MS/MS of $[\text{Al}^{1\text{CSdimer}} + \text{GSH} + 2\text{GS-hex} + 16\text{H}^+]^{16+}$ (m/z 3255). Collision offset A) 20V, B) 30V, C) 40V, D) 50V
The insets show an enlarged region from m/z 3170 to m/z 3275. In each of the spectra the lower case letters correspond to the following: v: $[\text{Al}^{1\text{CSdimer}}]$, w: $[\text{Al}^{1\text{CSdimer}} + \text{GSH}]$, x: $[\text{Al}^{1\text{CSdimer}} + \text{GS-hex}]$, y: $[\text{Al}^{1\text{CSdimer}} + \text{GSH} + \text{GS-hex}]$, z: $[\text{Al}^{1\text{CSdimer}} + 2\text{GS-hex}]$
Figure 5.19 Spectra from MS/MS of \([\text{A}1^{\text{CS}}\text{dimer} + \text{GS-hex} + \text{GS-dec} + 16\text{H}^+]^{16+}\) (m/z 3239). Collision offset A) 35V, B) 45V, C) 50V

The insets show an enlarged region around m/z 3200. In each of the spectra the lower case letters correspond to the following: v: \([\text{A}1^{\text{CS}}\text{dimer}]\), w: \([\text{A}1^{\text{CS}}\text{dimer} + \text{GS-hex}]\), x: \([\text{A}1^{\text{CS}}\text{dimer} + \text{GS-dec}]\)
corresponded to $[\text{Al}^{\text{CS}}\text{dimer}]$ (v), $[\text{Al}^{\text{CS}}\text{dimer} + \text{GS-hex}]$ (w) and $[\text{Al}^{\text{CS}}\text{dimer} + \text{GS-dec}]$ (x). The dimer in the absence of ligands was the most abundant component at a voltage of 50 V (Figure 5.19C). It is interesting that the complex $[\text{Al}^{\text{CS}}\text{dimer} + \text{GS-dec}]$ was more abundant than $[\text{Al}^{\text{CS}}\text{dimer} + \text{GS-hex}]$ at collision cell voltages of 45 V and 50 V suggesting that, in the gas phase, GS-hex was removed more readily from the complex than GS-dec. Also visible in the spectrum at 50 V in the lower m/z region were peaks from the monomeric form of $\text{Al}^{\text{CS}}$.

To investigate the gas phase properties of complexes with three ligands, further MS/MS experiments were undertaken. The S-alkylGSH inhibitors have been shown to simultaneously occupy the G and H-site of GST, resulting in the stoichiometry 1:2 (dimer:inhibitor). The presence of additional ligands may therefore be the result of non-specific binding. Figure 5.20 shows MS/MS spectra of the 15$^+$ ion of complex $[\text{Al}^{\text{CS}}\text{dimer} + 3\text{GS-hex}]$ at collision cell voltages of 35 V and 45 V. At a collision cell voltage of 35 V the loss of a single molecule of GS-hex was apparent (Figure 5.20A). There was no evidence of the loss of any further ligands in this spectrum. Increasing the voltage by 5 V resulted in the appearance of peaks corresponding to the loss of two and all three of the GS-hex molecules. These results suggested that displacement of the third molecule of GS-hex in the complex $[\text{Al}^{\text{CS}} + 3\text{GS-hex}]$ required less energy than that required for the displacement of the first two molecules of GS-hex. Figure 5.21 shows a comparison of the spectra from MS/MS of the 15$^+$ ions of the complexes $[\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-hex}]$ and $[\text{Al}^{\text{CS}}\text{dimer} + 3\text{GS-hex}]$ at a collision cell voltage of 35 V. Under the same conditions, MS/MS of the complex $[\text{Al}^{\text{CS}}\text{dimer} + 3\text{GS-hex}]$ revealed the loss of one molecule of GS-hex whereas MS/MS of the complex $[\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-hex}]$ did not show the loss of any molecules of GS-hex. The results suggested that less energy was required to displace a single molecule of GS-hex from the complex $[\text{Al}^{\text{CS}}\text{dimer} +$
loss of S-hexylGSH

Figure 5.20 Spectra from MS/MS of $[\text{Al}^{15}\text{C}^{6}\text{dimer} + 3\text{GS-hex} + 15\text{H}^+]^{15+}$ (m/z 3477). Collision offset A) 35V, B) 40V
The loss of only one molecule of GS-hex was apparent in spectrum A whilst in spectrum B the loss of one, two and three molecules of GS-hex were apparent.
Figure 5.21 Spectra from MS/MS of A) [A1CS dimer + 2GS-hex + 15H+]15+ (m/z 3451) and B) [A1CS dimer + 3GS-hex + 15H+]15+ (m/z 3477). Collision offset 35V
3GS-hex] than was required to displace a single molecule from the complex [A1\textsuperscript{CS}dimer + 2GS-hex].

### 5.2.8 In-source collision activated dissociation of noncovalent complexes

Collision-activated dissociation in the source (in-source CAD) is achieved by increasing the voltage on the sampling cone. A major difference between this process and an MS/MS experiment is that mass selection is not performed prior to fragmentation. Figure 5.22 shows the effect of in-source CAD on the spectra of a mixture of A1\textsuperscript{CS} and GSH in a molar ratio of 1:10 (A1\textsuperscript{CS}:GSH). The spectra displayed were acquired at cone voltages of 65, 85, 105, 125 and 145 V. At a cone voltage of 65 V, three ion series were apparent and these were from the complexes [A1\textsuperscript{CS}dimer] (a), [A1\textsuperscript{CS}dimer + GSH] (b) and [A1\textsuperscript{CS}dimer + 2GSH] (c). The most abundant series was from [A1\textsuperscript{CS}dimer + GSH]. With increasing cone voltages, the abundance of the complexes [A1\textsuperscript{CS}dimer + GSH] and [A1\textsuperscript{CS}dimer + 2GSH] decreased and the dimer with no bound ligands was the predominant species (Figure 5.22C-E). In the spectra acquired at cone voltages of 125V and 145V the ion series from [A1\textsuperscript{CS}dimer + GSH] and [A1\textsuperscript{CS}dimer + 2GSH] were of very low intensity. At cone voltages of 85 V and above monomeric A1\textsuperscript{CS} was also apparent (labelled M, Figure 5.22B-E).

The effect of in-source CAD on the spectra of a mixture of A1\textsuperscript{CS} and GS-dec (molar ratio 1:5) is shown in Figure 5.23. In each of the spectra the most abundant series of ions was from the complex [A1\textsuperscript{CS}dimer + 2GS-dec]. This further reinforces the notion of a strong, selective complex being formed between the dimer and two molecules of GS-dec. At a cone voltage of 125 V other series were apparent in increasing relative abundance. These series were from the complexes [A1\textsuperscript{CS}dimer] and [A1\textsuperscript{CS}dimer + GS-dec]. Monomeric A1\textsuperscript{CS} (m/z 2318, 2549, 2833) was of higher
abundance with increasing cone voltage and the monomer with one molecule GS-dec bound was also apparent (Figure 5.23E). The presence of monomer with GS-dec bound suggested that the interactions between the inhibitor and an individual monomer were sufficiently strong to maintain a complex even when the dimeric structure of the protein was disrupted.

5.2.9 Effect of increased collision cell voltage

In MS/MS spectra, the collision cell voltage is used to further accelerate the precursor ion and cause more energetic collisions with the collision gas. By increasing the collision cell voltage without prior mass selection the entire m/z range of ions can undergo CAD in the gas cell. Figure 5.24 shows the effect of increasing the collision cell voltage on the spectra of a mixture of A1CS, GSH and GS-hex in a molar ratio of 2:20:1. The spectrum acquired at a collision cell voltage of 30 V was similar to that acquired at a voltage of 10 V (data not shown).

The most abundant ion series in the spectrum acquired at 30 V was from the complex [A1CSdimer + 2GS-hex] (series c, Figure 5.24A). This component was also the most abundant at a voltage of 40 V. The abundance of [A1CSdimer + GS-hex] (series b) and [A1CSdimer] (series a) increased when the voltage was increased from 30 V to 40 V (Figure 5.24B). At a voltage of 50 V, the dimer with no bound ligands (series a) was the most abundant series. The complex [A1CSdimer + 2GS-hex] was, however, still quite prominent. When the collision cell voltage was increased to 60 V the complex [A1CSdimer + GSH + 2GS-hex] (series d) was no longer apparent. Monomeric A1CS was also visible in the spectra acquired at collision cell voltages of 50 V and 60 V. The increase in collision cell voltage showed a general trend of increasing loss of the GS-hex molecules bound to the A1CS dimer and loss of dimeric structure.
Figure 5.22 Spectra from in-source CAD of a mixture of A1\textsuperscript{CS} + GSH (molar ratio 1:10)
Cone voltage A) 65V, B) 85V, C) 105V, D) 125V, E) 145V
In each of the spectra the lower case letters correspond to the following: a: [A1\textsuperscript{CS}dimer], b: [A1\textsuperscript{CS}dimer + GSH], c: [A1\textsuperscript{CS}dimer + 2GSH]
The peaks labelled M correspond to molecular ions of the monomer A1\textsuperscript{CS}
Figure 5.23 Spectra from in-source CAD of a mixture of Al$^{1\text{CS}}$ + GS-dec (molar ratio 1:5)
Cone voltage A) 105V, B) 115V, C) 125V, D) 135V, E) 145V
In spectrum E the lower case letters correspond to the following: a: [Al$^{1\text{CS}}$dimer], b: [Al$^{1\text{CS}}$dimer + GS-dec]
The peaks labelled M and M* in spectrum E correspond to molecular ions of the monomer Al$^{1\text{CS}}$ and the molecular ions of [monomeric Al$^{1\text{CS}}$ + GS-dec] respectively.
Figure 5.24 ESI-MS spectra from a mixture of $\text{Al}^{\text{CS}}$, GSH and GS-hex (molar ratio 2:20:1) under non-denaturing conditions with increasing collision offset. Collision offset A) 30V, B) 40V, C) 50V, D) 60V.

In each of the spectra the lower case letters correspond to the following: a: $[\text{Al}^{\text{CS}}\text{dimer}]$, b: $[\text{Al}^{\text{CS}}\text{dimer} + \text{GS-hex}]$, c: $[\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-hex}]$, d: $[\text{Al}^{\text{CS}}\text{dimer} + \text{GSH} + 2\text{GS-hex}]$.

The peaks labelled M correspond to molecular ions of the monomer $\text{Al}^{\text{CS}}$. 
5.2.10 Titrations of GST A1-1(C112S) with various ligands

To further characterise the binding properties of the ligands, titrations were performed with Alc\textsuperscript{CS} and the ligands GSH, GS-prop and GS-hex. The results from these titrations are displayed in Figures 5.25, 5.26 and 5.27. Each of the graphs shows the total peak height of a given component relative to the sum of peaks heights of all components.\textsuperscript{6} The relative abundances are shown for the Alc\textsuperscript{CS} dimer alone and the Alc\textsuperscript{CS} dimer with one, two, three and four ligands bound. Figure 5.25 shows the relative abundances of the components from the addition of GSH to Alc\textsuperscript{CS}. The data showed a decrease in the amount of the Alc\textsuperscript{CS} dimer with increasing molar excess of GSH. The relative abundance of the other components increased with molar excess. It is important to note that plotting the component peak height sum relative to the peak height sum of all ions does introduce a tendency in the graphs to slope downwards at higher molar excesses of ligands, owing to the appearance of complexes with multiple ligands bound at the higher molar ratios. Hence, even if a particular complex increases in abundance relative to the Alc\textsuperscript{CS} dimer, the presence of other complexes will reduce its abundance relative to the sum of peak heights of all ions. The peak height of the Alc\textsuperscript{CS} dimer could not be used to normalise the data, as in some circumstances it was not present.

A comparison of the titrations of each of the different ligands is presented in Figure 5.26. There are notable differences between the three ligands. First, examining the Alc\textsuperscript{CS} dimer without any bound ligands showed that in each case the relative abundance of the dimer decreased with increasing molar excess of ligand. This decrease, however, occurred at lower molar excess of ligand in the case of GS-prop as

\textsuperscript{6} To obtain the ‘percentage of total intensity’, the sum of peak heights of all the ions for a given component was divided by the sum of peak heights of all ions. While recognising that the peak heights alone don’t correlate exactly with the abundance of complexes in solution owing to ionisation and detection differences, the percentage total intensity is herein referred to as the ‘relative abundance’.
opposed to GSH. Even more dramatic was the decrease of the A1\textsuperscript{CS} dimer in the plot of A1\textsuperscript{CS} with GS-hex. At a ten-fold molar excess of GS-hex the series from the A1\textsuperscript{CS} dimer was no longer apparent. This rapid decrease can be seen more clearly in Figure 5.27, which is an expanded portion of the same GS-hex curve seen in Figure 5.26.

Examining the data from the complex of the A1\textsuperscript{CS} dimer with one ligand showed a similar enhancement of the trend from GSH to GS-hex. In the titration with GSH the complex \([A1\textsuperscript{CS} \text{dimer} + \text{GSH}]\) (purple line) increased to 43% of total intensity and then began to decrease. The corresponding species in the GS-prop titration reached a maximum at a lower molar excess of ligand (2-fold as opposed to 15-fold) and then decreased more rapidly. This decrease appears not to be a consequence of the effect of extra complexes (mentioned above) as it was much more dramatic than the slow decrease seen in Figure 5.26. Presumably the decrease can be attributed to the higher molar excess allowing both the sites on the dimer to be accommodated, leaving only a low proportion with a single site occupied. This effect was even more dramatic in the titration with GS-hex. The expanded section in Figure 5.27 shows that the maximum relative abundance of \([A1\textsuperscript{CS} \text{dimer} + \text{GS-hex}]\) was reached at a molar excess of A1\textsuperscript{CS} rather than an excess of ligand, as was seen with the other ligands. The complex \([A1\textsuperscript{CS} \text{dimer} + \text{GS-hex}]\) seemed to maintain some baseline level throughout the titration (up to a forty-fold molar excess of ligand). The green line in each of the graphs shows the relative abundance of the dimer with two ligands. The titration with GSH showed a steady increase in the relative abundance of this complex. Figure 5.25 shows that at higher molar excesses of GSH this began to decrease. As mentioned above, this may have arisen from additional components from non-specific ligand binding. The equivalent complex in the titration of GST with GS-prop revealed a more rapid increase with maximum relative abundance at a ten-fold molar excess of GS-prop. Again the
slow decrease was seen in this curve, possibly owing to the increase in non-specific binding at these high molar excesses of ligand. The titration with GS-hex revealed that the maximum was reached at lower molar excess compared to the GS-prop titration. In this titration the highest relative abundance of \([A1^{CS}\text{dimer} + 2\text{GS-hex}]\) was seen at a two-fold excess of GS-hex over \(A1^{CS}\). The slow decrease was again seen with increasing proportions of the complexes with more than two GS-hex molecules bound to the \(A1^{CS}\) dimer.

The relative abundance from complexes with three and four molecules of ligands are displayed in blue and orange respectively. In all the titration experiments neither of these complexes were present initially. In each of the cases they showed an increase with increasing excess of the ligands. The most notable difference between the behaviour of these two species and the species with one or two ligands, was apparent in the titrations with GS-prop and GS-hex. With both these ligands, the binding of one and two molecules reached a defined maximum prior to a steady decrease. The complexes with three and four ligands did not display this behaviour across the range of molar ratios examined. The other notable point from the graphs of these two components was their increase in relative abundance as the relative abundance of the complexes with one and two ligands decreased. The presence of the complexes with three and four ligands bound did affect the graphs of the other components quite significantly and should be taken into account when interpreting the results.
Figure 5.25 Graph of relative intensity of complexes of Al^{CS} dimer and GSH versus the molar excess of GSH. The percentage total intensity on the y-axis is the intensity of the complex relative to the total intensity of all components. The coloured lines correspond to the following complexes: **Red**: [Al^{CS}dimer], **Purple**: [Al^{CS}dimer + GSH], **Green**: [Al^{CS}dimer + 2GSH], **Blue**: [Al^{CS}dimer + 3GSH], **Orange**: [Al^{CS}dimer + 4GSH]
Chapter 5: Noncovalent complexes of GST

Figure 5.26 Titrations of A1<sub>CS</sub> with A) GSH, B) GS-prop, C) GS-hex
The coloured lines represent the following: **Red**: [A1<sub>CS</sub>dimer], **Purple**: [A1<sub>CS</sub>dimer + ligand], **Green**: [A1<sub>CS</sub>dimer + 2ligands], **Blue**: [A1<sub>CS</sub>dimer + 3ligands], **Orange**: [A1<sub>CS</sub>dimer + 4ligands]
Figure 5.27: An enlarged portion of the titration of A1\textsuperscript{CS} with GS-hex. The coloured lines correspond to the following complexes: 
\textbf{Red}: [A1\textsuperscript{CS} dimer], \textbf{Purple}: [A1\textsuperscript{CS} dimer + GS-hex], \textbf{Green}: [A1\textsuperscript{CS} dimer + 2GS-hex], \textbf{Blue}: [A1\textsuperscript{CS} dimer + 3GS-hex], \textbf{Orange}: [A1\textsuperscript{CS} dimer + 4GS-hex]
5.2.11 Formation of GST dimers within a class but not between different classes

It has been noted that dimers of GST form between proteins of the same class but do not form between proteins from different classes. Hence, a dimer can form between A1 and A2 to give the dimeric structure noted as GST A1-2. Dimers should therefore not be able to form between P1 and A1 proteins. To determine whether mass spectrometry could confirm this a mixture of the P class enzyme and an A class enzyme was analysed under non-denaturing conditions. A typical spectrum from this experiment is shown in Figure 5.28. This spectrum was processed using MaxEnt™ software to convert the data to a mass scale from a mass-to-charge scale. The presence of the A1\textsuperscript{CS} dimer, and of the P1 dimer is clear. P1 showed the three different dimers as in Figure 5.7D. These were P1/P1 (A), P1/P1Met (B) and P1Met/P1Met (C) (Figure 5.28). Heterodimers of P1 and A1\textsuperscript{CS} would have masses of 48709 Da (A1/P1) and 48804 Da (A1/P1Met). There was no evidence of these masses in the spectrum, which reinforced the proposition that cross-class heterodimers are not formed between monomers of GST.
Figure 5.28 MaxEnt\textsuperscript{TM}1 spectrum of a mixture of A1\textsuperscript{CS} and P1. The three different dimers of P1 are labelled A: P1/P1, B: P1/P1Met and C: P1Met/P1Met. There was no evidence of heterodimers of A1\textsuperscript{CS} and P1, which would have masses of 48709 Da and 48840 Da.
5.3 DISCUSSION

5.3.1 ESI-MS spectra of proteins acquired under non-denaturing conditions

For electrospray mass spectrometry analyses of proteins, a solution of the analyte in 50% aqueous organic solvent (usually acetonitrile or methanol) that has been acidified with a weak acid such as acetic or formic acid is generally used. In these solvents a protein cannot normally maintain its secondary or tertiary structure and is therefore present in an unfolded or partially unfolded form (Loo, 1997). Protons that are readily available from the acidified solution attach to the accessible sites of the unfolded protein and the resulting spectra show an envelope of a large number of multiply charged ions. The molecular mass can then be calculated from the m/z of each ion and an average taken. This provides a more accurate mass than would be achieved with a single molecular ion.

To transfer noncovalent complexes from the condensed phase into the gas phase the secondary and tertiary elements of structure must be preserved in the solution from which the analyte is sprayed. The solutions used are chosen to represent conditions close to physiological conditions (i.e. aqueous solutions of neutral pH). With the maintenance of structure, the sites where protons would be likely to attach are generally much less accessible. In neutral, or weakly basic solutions there are fewer protons available for protonation. Therefore the spectra characteristically have a smaller m/z range of ions and the charges observed are lower. The resulting m/z range at which the protein molecular ions are seen is usually higher than under denaturing conditions. If the quaternary structure involves complexes of more than one protein, the increase in molecular mass also results in complexes being observed at higher m/z ranges. Under denaturing conditions the m/z range in which ions of the GSTs were observed was m/z 700-1800, with charges ranging from 17+ to 30+. The spectra of GSTs acquired under
non-denaturing conditions showed a smaller range (m/z 3000-3600) with charges ranging from 14+ to 17+. The spectra of GST proteins therefore demonstrated the expected characteristic differences of ESI-MS spectra of proteins acquired under denaturing and non-denaturing conditions (Przybylski and Glocker, 1996; Loo, 1997).

5.3.2 Effect of gas and voltage on ESI-MS spectra acquired under non-denaturing conditions

It has been documented, and was observed in this work, that noncovalent complexes are sensitive to source voltages, source temperature and pressure (Loo, 1997; Smith and Light-Wahl, 1993; Przybylski and Glocker, 1996; Tahallah et al., 2001). The optimal source temperature for GST dimers was found to be approximately 50°C. At higher temperatures, the dimeric structure started to break down and monomeric protein increased in abundance. It was found that changes in the desolvation gas flow in the source, the collision gas pressure or in the backing or rough pumping influenced the spectra in a number of ways. These changes included a general increase or decrease in the abundance across the whole spectrum, changes in the relative abundance of one component compared to another, changes in the distribution of charges and in the amount of monomeric protein present. The magnitude of the changes varied and was also dependent on cone voltage and capillary voltage. The general trend was that harsher conditions, particularly in the high-pressure source region, resulted in a decreased abundance of the noncovalent complexes. Restricting the in-line isolation valve on the rotary pump of the QTof™ influenced the pressure in different regions of the mass spectrometer and was difficult to monitor accurately. The interdependence of voltage and pressure made it difficult to establish identical conditions each time the instrument was tuned for these experiments. Therefore comparisons between
experiments performed on different days had to be interpreted with caution. For experiments that required comparisons between different mixtures, the samples were run under the same conditions on the same day. This minimised variations in experimental conditions and ensured reproducibility of the results obtained.

5.3.3 Complexes formed from the addition of GSH to GST

Previous studies with glutathione bound to GST showed that the complex exists with two molecules of glutathione bound to the dimer (Armstrong, 1997; Sinning et al., 1993; Cameron et al., 1995; Oakley et al., 1997). A similar result was expected to be apparent in ESI-MS spectra acquired under non-denaturing conditions. Ishigai et al. reported that the most abundant complex from the addition of glutathione to A1 was [A1dimer + GSH] (Ishigai et al., 2000). The results presented in this chapter also showed this complex to be the most abundant with A1, A1CS and P1. The only protein that did show the expected complex [dimer + 2GSH] to be the most abundant was the modified A class enzyme A1YF. The relative abundances of complexes in the spectra from a mixture of GSH and A1YF did not show a dramatic preference for the binding of two ligands, as seen with the addition of GS-hex to the same protein. It is possible, however, that the difference between A1YF and the other proteins was a result of uncertainty in the protein concentrations.

Owing to the confusing results obtained from the addition of GSH to the three A class proteins, and the contrast with the results from the addition of S-alkylGSHs, titrations were performed to attempt to better understand the binding characteristics of these ligands. These titration experiments are discussed in section 5.3.6.

The addition of GSH to GST A1 in the results presented by Ishigai et al. (Ishigai et al., 2000) showed a different increase in the mass than was seen in similar
experiments presented in the current work. Ishigai et al. reported the addition of 340 Da to the dimer when GSH was added to A1 (Ishigai et al., 2000). The molecular mass of glutathione is 307.3 Da and in the current work an addition of 307 Da to the dimer of A1 was observed. Furthermore, the same additional mass was also seen for the proteins A1CS, A1YF and P1 and components arising from multiple additions of 307 Da were also detected. In contrast to the reproducible masses obtained from the addition of GSH, the spectrum of A1 without added GSH acquired under non-denaturing conditions (Figure 5.7) showed a series of ions from the addition of 337 Da. A similar adduct was also evident in the spectra from Ishigai et al. (Ishigai et al., 2000). Under denaturing conditions the spectrum of A1 also showed a second component, which was 345 Da above the mass of A1, but it was of very low relative abundance (Figure 5.1). The presence of a glutathione complex prior to the addition of any free glutathione has been attributed to the preparation of the A1 protein because affinity chromatography with glutathione as the eluant was used (Oxford Biomedical Research, Michigan, U.S.A.) (Ishigai et al., 2000). Therefore residual glutathione may still have been present in the solution. A1CS, A1YF and P1 were prepared without an affinity column and did not show the additional component in the spectra without added GSH, which further indicates that the presence of the extra component in the spectrum of A1 arose from the purification technique. The reason for the addition of 337 Da in this spectrum and not in the spectrum with additional free GSH is unclear. Ishigai et al. observed an increase in the abundance of the complex [A1dimer + 340] when GSH was added to A1. The results presented here show that the adduct of 337 Da was effectively displaced by GSH alone (307 Da).

To account for the higher mass observed (i.e. 340 Da instead of 307 Da), Ishigai et al. proposed that the complex of A1 with GSH included additional small molecules,
possibly two molecules of water (Ishigai et al., 2000). The data presented here do not support this conclusion. Further work is required to identify the precise origin of this component. The inclusion of water molecules in gas phase complexes seen by ESI-MS has been reported previously (Benjamin et al., 1998; Green et al., 1999; Fabris and Fenselau, 1999) but the reproducibility of spectra with complexes containing water is as yet unknown and this an area that requires further investigation.

5.3.4 Effect of the tyrosine to phenylalanine modification

The tyrosine-to-phenylalanine substitution has been shown to have a dramatic effect on the catalytic activity of GST but does not appear to change the binding of GSH to GST (Allardyce et al., 1999). In contrast, the substitution has been shown to impact the binding of ligands such as S-dinitrophenylglutathione (Allardyce et al., 1999). The effect of the Y9F substitution on the ESI-MS spectra presented was not obvious. The addition of GSH to A1YF resulted in a different spectrum to the spectra obtained from a mixture of GSH and A1 (wild-type) or A1CS (Figure 5.9). It was, however, possible that the difference may have arisen from concentration differences in the initial protein solutions. The A1YF protein was the only GST that showed the complex [dimer + 2GSH] to be the most abundant component in the spectra from a mixture of GSH and GST. This may be a consequence of a decrease in the dissociation of GSH from this protein but further experiments would be required to draw conclusions from these data about the binding of GSH to A1YF under ESI-MS conditions.

Previous results from the binding of hydrophobic ligands to the wild type A1 and A1YF indicated that loss of the hydroxyl group improved the binding of these ligands (Allardyce et al., 1999; Nieslanik and Atkins, 2000). The S-alkylGSH ligands may have been expected to bind to A1YF with a lower dissociation constant than the complexes
with Al and Al^{CS}. The binding of GS-hex was used to probe any apparent differences. The spectra from the addition of GS-hex to the A class GSTs (Figure 5.11) showed the binding of GS-hex to each of the proteins clearly favoured the formation of the complex [dimer + 2GS-hex]. The spectra from these experiments did not reveal any obvious difference in the binding of GS-hex between these proteins. Similarly the formation of complexes between GS-dec and dimers of A^{CS} or A^{YF} (Figure 5.12) indicated strong, selective binding of two GS-dec molecules to the dimers. The selectivity and strength of the interactions may have concealed any small differences in the binding of the ligands to these proteins. MS alone was not sufficient to reveal differences between the binding characteristics of hydrophobic ligands to GSTs. To further examine whether differences between the complexes [A^{CS}dimer + 2GS-hex] and [A^{YF}dimer + 2GS-hex] could be detected by mass spectrometry, MS/MS was performed on these complexes. The spectra in Figure 5.17 show the results from MS/MS of the 16+ ion of the complex [A^{YF}dimer + 2GS-hex]. The results from MS/MS of the equivalent ion from [A^{CS}dimer + 2GS-hex] acquired on the same day did not show any difference to those seen in Figure 5.18 (data not shown). It is probable that the gas phase behaviour of these complexes is not identical to their condensed phase behaviour, which may explain why a difference in binding of GS-hex to A^{YF} was not apparent. MS/MS was also performed on the complexes [A^{YF}dimer + GSH] and [A^{CS}dimer + GSH] (data not shown) and no apparent difference was observed in the spectra of these complexes either. Although it is speculative to infer that MS/MS results reflect the condensed phase binding characteristics, there is some merit in examining the behaviour of these complexes by MS/MS. The absence of differences between the MS/MS spectra does not rule out different binding properties in the complexes. Notable differences in the spectra could,
however, suggest differences in binding that might also be relevant in the condensed phase.

### 5.3.5 Effect of increasing chain length of S-alkylglutathiones

It has been well documented that glutathione conjugates of the form GSR (where R is a hydrophobic group) inhibit GSTs (Armstrong, 1997). Over three decades ago, S-alkylglutathiones were reported to inhibit the enzyme glyoxalase and increasing chain length resulted in stronger inhibition (Vince and Wadd, 1969). The results from competition experiments with S-alkylGSH inhibitors indicated that binding of GST proteins to longer chain S-alkylGSH inhibitors was preferred. With a mixture of GS-prop and GS-hex the most abundant complex was \([\text{dimer} + 2\text{GS-hex}]\), even when GS-prop was fifteen times more concentrated than GS-hex and the protein concentration was twice the GS-hex concentration (Figure 5.14). When GS-hex and GS-dec were added to A1$^{\text{CS}}$ and A1$^{\text{YF}}$, ESI-MS of the mixture showed the complex \([\text{dimer} + 2\text{GS-dec}]\) was the most abundant component.

The S-alkylGSH inhibitors have the ability to occupy the G-site and the H-site simultaneously (Armstrong, 1997; Oakley et al., 1997). The GSH moiety is contained in the G-site whilst the hydrophobic alkyl chain lies in the H-site. Presumably the longer chain inhibitors are able to better occupy the H-site as well as the G-site and thereby utilise binding contacts more efficiently. GS-prop did not show the clear spectra that GS-hex and GS-dec revealed. This may reflect a lower ability of this ligand to use both the sites. Docking simulations by Koehler et al. have shown that the hydrophobic cavity of the H-site was only partially occupied with the alkyl chain of S-butylGSH (GS-but) whereas GS-hex filled the site (Koehler et al., 1997). This could account for the lower affinity of the shorter chain S-alkylGSH inhibitors. Using the fluorescent properties of
tyrosine, it has been shown that with increasing chain length of S-alkylGSH inhibitors, the environment around tyrosine 9 in the dimer of A1 becomes less exposed to solvent and more hydrophobic (Dietze et al., 1996b). This also suggests that the active site is more fully occupied by the longer alkyl chains.

The crystal structures of S-nonylGSH in complex with the dimer of GST P1 revealed that the nonyl chain was found in the same position as the hexyl chain of GS-hex from a previously published structure (Oakley et al., 1999). The three extra carbon units in the nonyl chain occupied a hydrophobic pocket. The presence of additional contacts with the extra units on a hydrophobic chain would presumably increase the binding of the ligand to the dimer. Oakley et al. reported that the S-nonylGSH inhibitor made 87 van der Waals contacts, 3 salt links and 24 hydrogen bonds with the enzyme (Oakley et al., 1999).

It has also been hypothesised that the H-site would be able to accommodate approximately nine carbon units (Sinning et al., 1993). With chains longer than 9 units, the cleft in the dimer interface could be partially filled as well, possibly increasing the binding affinity by additional contacts in this cleft.

The increased binding affinity of GST for the longer chain S-alkylGSH inhibitors was also apparent from the MS/MS spectra of the complex [AlCSdimer + S-hex + S-dec] (Figure 5.19). The loss of GS-hex was more abundant than the loss of GS-dec, particularly at a collision cell voltage of 45 V. This suggested that the GS-hex was bound less strongly than the GS-dec. With the absence of solvent molecules in the gas phase, hydrophobic interactions are not expected to be as relevant as they are in condensed phase. The strong binding of these S-alkylGSH inhibitors has been attributed to hydrophobic interactions between the H-site of the GST dimer and the hydrophobic alkyl chain. That MS/MS of these complexes showed differences in the energy required
to displace the ligands, suggests that there is still some effect from the difference in hydrophobicity. Alternatively, other interactions may also be important in the binding of these S-alkylGSH inhibitors to GST. These may include van der Waals interactions (such as those reported for S-nonylGSH bound to GST P1 (Oakley et al., 1999)) and steric effects. The A class GSTs possess a C-terminal helix that is thought to cover the H-site whilst the G-site is occupied (Allardyce et al., 1999; Dirr and Wallace, 1999; Nieslanik et al., 1999). This structural element, if preserved in the gas phase, may stabilise the S-alkylGSH molecules in the gas phase complexes. With GS-dec it is also possible that the alkyl chain is partially bound in the cleft between the two monomers, increasing the binding strength through interactions at that point.

5.3.6 Non-specific binding

The possibility of non-specific binding was evident for a number of the complexes seen in the ESI-MS spectra. The spectra of complexes formed between GSH and GST did not show the same specific binding that was seen in spectra of complexes between GST and GS-hex. With the addition of ligands to the GSTs, complexes with more than two molecules of the ligand in complex with the dimer were frequently observed. GST has only two catalytic sites, each made up of a G-site and an H-site. The binding of a third ligand therefore is presumed to be the result of non-specific binding.

To examine further the possibility of non-specific binding in the complexes formed between GSH and GST, titrations were performed with Al\textsuperscript{CS} and GSH. These were compared to titrations of GS-prop with Al\textsuperscript{CS}, and of GS-hex with Al\textsuperscript{CS}. The changes evident in the graphs were more obvious in the titration of GS-hex with Al\textsuperscript{CS} (Figure 5.26). Trends that were apparent in the GS-hex titration were also apparent in the GS-prop titration but occurred more gradually with increasing concentration of GS-
prop. It was more difficult to see the same trends in the titration with GSH. The GSH titration (Figure 5.25) did reveal some trends that were apparent in the titrations with GS-prop and GS-hex (Figure 5.26). The decrease in the relative intensity of the $\text{Al}^{\text{CS}}$ dimer was clearly evident up to an eighty-fold molar excess of GSH. The relative intensity of the complex [Al$^{\text{CS}}$dimer + GSH] also showed a similar trend in reaching a maximum at fifteen-fold molar excess and thereafter decreasing. This trend was seen over a higher range of ligand molar excess compared to the titrations of GS-prop and GS-hex. It was difficult to see the same trend with the complex [Al$^{\text{CS}}$dimer + 2GSH]. It should be noted, however, that in the titration with GS-prop it was also difficult to assess the trend in the complex [Al$^{\text{CS}}$dimer + 2GS-prop] even though it did appear to follow behaviour similar to [Al$^{\text{CS}}$dimer + 2GS-hex]. From these titrations it was difficult to conclude that the binding of GSH to GST was specific but the behaviour of the complex [Al$^{\text{CS}}$dimer + GSH] did suggest that there was at least some specific binding present, albeit weaker than that seen for the S-alkylGSH ligands.

The titrations also provided information on the behaviour of the complexes with more than two ligands bound. Over the range of molar ratios examined, the graphs of the relative intensities of [dimer + 3ligands] and [dimer + 4ligands] did not reflect the trends seen for the complexes with one or two ligands. The titrations alone could not lead to the conclusion that the binding of the third ligand was due to non-specific interactions. Other experiments were undertaken to investigate differences between complexes with two ligands versus complexes with three ligands. MS/MS, in-source CAD and increases in the collision cell voltage under otherwise normal non-denaturing MS conditions were experiments that were used to explore the specificity of the complexes.
MS/MS of the complex \([\text{Al}^{\text{CS}}\text{dimer} + 3\text{GS-hex}]\) revealed the loss of a single molecule of GS-hex occurred at a lower collision cell voltage than that required to displace two and three molecules of GS-hex (Figure 5.20). This suggests that one ligand is displaced more easily from the complex \([\text{Al}^{\text{CS}}\text{dimer} + 3\text{GS-hex}]\) than the remaining two. A comparison of the complexes \([\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-hex}]\) and \([\text{Al}^{\text{CS}}\text{dimer} + 3\text{GS-hex}]\) at the same collision cell voltage (Figure 5.21) showed the loss of one molecule of GS-hex from the latter complex but not from the complex with two GS-hex molecules. These data indicate that in the gas phase a single ligand is displaced from the complex \([\text{Al}^{\text{CS}}\text{dimer} + 3\text{GS-hex}]\) more easily than from the complex \([\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-hex}]\). It appears, therefore, that the third GS-hex molecule was bound less tightly than the first two molecules, consistent with the notion that this ligand was not bound at a specific site or sites.

In-source CAD of the complexes formed from the addition of GSH to \(\text{Al}^{\text{CS}}\) showed a general decrease in the complexes formed between GST and GSH over the range of increasing cone voltages. The dimer stripped of ligands was the only remaining complex at a voltage of 145 V (Figure 5.22). The appearance of monomeric GST also increased with increasing voltage. These results indicated that the loss of glutathione from the complexes occurred prior to dissociation of the dimer into monomers. This observation was also reported by Ishigai et al. (Ishigai et al., 2000). Differences in the loss of glutathione from \([\text{Al}^{\text{CS}}\text{dimer} + 2\text{GSH}]\) and \([\text{Al}^{\text{CS}}\text{dimer} + 3\text{GSH}]\) were not apparent in the spectra.

In contrast, the strong binding of GS-dec to dimeric \(\text{Al}^{\text{CS}}\) was evident in the spectra from in-source dissociation of these complexes (Figure 5.23). Even at a cone voltage of 145 V the complex \([\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-dec}]\) was the most abundant complex. There was very little change in the spectra from cone voltages up to 125 V (except for
the appearance of monomeric $\text{Al}^{\text{CS}}$). The binding of one and two molecules of GS-dec was obviously very strong and specific. The presence of monomeric $\text{AlCS}$ with GS-dec bound also suggested that the binding was very strong. It is unlikely that the monomer/GS-dec complex was from a covalent modification, as MS/MS of a complex containing GS-dec did not show the monomer/GS-dec adduct (Figure 5.19).

An interesting effect was seen in the spectra from a mixture of $\text{Al}^{\text{CS}}$, GSH and GS-hex when the collision offset was increased in an MS experiment under non-denaturing conditions (Figure 5.24). The spectrum changed dramatically upon changing the collision cell voltage from 40 V to 50 V since the most abundant complex was $[\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-hex}]$ at 40 V and $[\text{Al}^{\text{CS}}\text{dimer}]$ at 50 V. The most interesting feature of the spectrum at 50 V was that even though the dimer in the absence of ligand was the most abundant complex, the complex $[\text{Al}^{\text{CS}}\text{dimer} + 2\text{GS-hex}]$ was still of high abundance, and was actually the most abundant complex of the 17+ ions. The intermediate series from the complex $[\text{Al}^{\text{CS}}\text{dimer} + \text{GSH}]$ was of much lower abundance. This suggests that, once conditions were sufficiently energetic for ligands to be stripped from the complex, both GS-hex molecules were removed, rather than just one. The relative abundance of the complex $[\text{Al}^{\text{CS}}\text{dimer} + \text{GSH} + 2\text{GS-hex}]$ did not appear to change significantly on changing the collision cell voltage from 30 V to 50 V. At a voltage of 60 V this complex was no longer apparent. If the third ligand in the complex $[\text{Al}^{\text{CS}}\text{dimer} + \text{GSH} + 2\text{GS-hex}]$ were the result of non-specific binding it would be expected to be lost from the spectrum at quite low collision energies. The data suggest that the third ligand was displaced at similar energies to the first two ligands, which may indicate that the third ligand was also present in a specific binding mode.
The possibility of three ligands bound in specific sites on GST may be attributed by the ligandin function of the enzyme. The GSTs are known to possess the ability to bind larger hydrophobic molecules (usually >400 Da) in the cleft between the two dimers (Oakley et al., 1999). This ligandin function is believed to be involved in sequestering, storing and transporting lipophilic molecules. The site for ligandin binding is close to the active site and possibly partially overlaps this site (Oakley et al., 1999). It is possible, however, that the ligandin binding-site can be occupied even when the G-site and H-site are occupied. This means that with the larger S-alkylGSHs, such as GS-hex and GS-dec, additional molecules could bind in the ligandin site whilst the catalytic sites (G-site and H-site) of the dimer were occupied. Furthermore, with competition experiments between GSH and GS-hex there are a number of possibilities that could account for the addition of more than two ligands. First, GSH could be present in one catalytic site, GS-hex in the other catalytic site and another molecule of GS-hex in the ligandin site. Alternatively, GS-hex does not require both the G-site and H-site of the catalytic site so GSH could be bound in the G-site whilst GS-hex was bound in the H-site. This would allow for the binding of two molecules of GSH (the G-sites), two molecules of GS-hex (the H-sites) and a further molecule of GS-hex in the cleft between the two molecules. The large number of bound ligands was not observed in the competition experiment so it is unlikely that this option for multiple ligand binding accounts for the unexpected high relative abundance of the complex [Al dimer + GSH + 2GS-hex] at low collision offsets (Figure 5.24).

5.3.7 Dimer formation revealed by mass spectrometry

The ability of ESI-MS to clearly show that spectra from a mixture of GSTs from different classes did not form heterodimers was a simple but important result. The
ability to detect dimers of GST with a method that is rapid and sensitive could prove to be very useful in investigation of interactions that are important for dimer formation.

The three A class GSTs were all of the isotype A1 and therefore were effectively homodimers in terms of the binding interfaces between monomeric units. This was also true for dimers between PI and PI Met. ESI-MS could, however, be used very effectively to provide fast, and detailed information on any heterodimers that form between different isotypes within a class. If a particular combination of isotypes from the same GST class did not form heterodimers, sequence comparisons might suggest residues that are critical for interactions between the subunits. Furthermore, site-directed mutagenesis could be used to investigate whether particular residues are crucial for dimer formation. ESI-MS provides a fast method for determining whether dimers are able to form between particular mutated proteins.

5.3.8 Conclusion: The applicability of ESI-MS to study non-covalent interactions of GST

The ESI-MS analysis of the glutathione S-transferase system showed varied results. The GS-hex and GS-dec inhibitors clearly showed that the most abundant complex arose from the binding of two molecules of inhibitor to the dimer, as was expected from the known structure of GST. The spectra from GSH did not show the same clear results although complexes of GSH with GST were detected. The contrasting results between the substrate and inhibitors reinforce the notion that the analysis of noncovalent complexes by ESI-MS should be interpreted with caution.

The spectra from complexes of GST with the S-alkylGSH inhibitors appeared to reflect the condensed-phase characteristics of these complexes. The binding of S-alkylGSH inhibitors relies on both electrostatic interactions in the G-site and
hydrophobic interactions, predominantly in the H-site. The increase in binding affinity with increasing chain length of S-alkylGSHs is attributed to the increased hydrophobic interactions. Hydrophobic interactions are thought to be diminished in the gas phase, whilst electrostatic interactions are thought to be strengthened (Wu et al., 1997; Robinson et al., 1996; Loo, 1997; Chung et al., 1999; Last and Robinson, 1999; Loo and Sannes-Lowry, 1998). The results obtained here for the GST enzyme system are not wholly consistent with this theory. The complexes between GST and GSH are known to be dependent on electrostatic interactions and yet ESI-MS spectra of these complexes suggest that the condensed phase abundance of the complexes is not maintained through transfer to the gas phase. The increase in binding affinity of longer chain S-alkylGSH inhibitors has been attributed to hydrophobic interactions and yet the abundances of the gas phase ions reflects the increase in binding affinity seen in the condensed phase, even though hydrophobic interactions are expected to have less impact in the gas-phase. The possibility exists that other interactions are also present that explain the apparent strong binding of S-alkylGSH inhibitors to GST in the gas-phase. Alternatively, the abundances of complexes in the condensed phase are reflected in the ESI-MS spectra, even though the complexes rely on hydrophobic interactions. The differences between the ion abundances seen for GSH and GST versus the S-alkylGSH inhibitors and GST suggest that important interactions between the inhibitors and GST enable these complexes to be transferred to the gas phase more efficiently.

The use of ESI-MS for the analysis of noncovalent complexes is very promising. Some systems have been shown to reflect the known behaviour of the condensed phase complexes, such as the GST/S-alkylGSH system presented here. There are, however, also examples of systems in which the condensed phase and gas-phase complexes behave differently, as seen here for the GST/GSH system. The observation of these
contrasts between ligands of the same enzyme highlights the caution required when ESI-MS is used for elucidating the structure of noncovalent complexes. Understanding of the process occurring when the complexes are transferred to the gas-phase may assist in interpreting changes from condensed phase behaviour to the gas-phase spectra. With improved understanding of these processes, ESI-MS may be a powerful tool for elucidation of the quaternary structure of protein complexes. As it stands, ESI-MS provides complementary information to that obtained from other techniques for structure determination and is a fast and sensitive method for obtaining high mass accuracy information on the stoichiometry of complexes. The gas-phase abundances of complexes are useful for obtaining information on the gas-phase binding properties and have often been found to reflect the condensed phase characteristics. Further analyses with model systems will continue to improve our understanding of the mechanisms involved and the increasing number of examples analysed by ESI-MS may assist in interpretation of spectra from protein complexes that have not been well characterised in solution.
Protein analysis by mass spectrometry has dramatically increased since the advent of MALDI and ESI. These two ionisation methods have resulted in the rapid expansion of biomolecular mass spectrometry and have had an important influence on fields such as immunology, biochemistry, virology, molecular biology, biotechnology and the new area of 'proteomics'.

The coupling of ESI to tandem mass spectrometers vastly increased the amount of information that could be obtained from proteins and peptides. With molecular mass and sequence information, the capability of identifying proteins from sequence databases is enormous. For novel proteins that have not yet been sequenced, ESI-MS also has a large potential for de novo peptide and protein sequencing, particularly with advances in automation.

The identification of proteins following 2-dimensional electrophoresis has come to rely on mass spectrometry. The speed, sensitivity and mass accuracy of MS are advantageous for the identification of proteins from the small amount of material that is displayed on a gel. An interesting application of proteomics, the identification of proteins from mitochondria before and after low conductance permeability (LC-PT), was presented in Chapter 3. A number of known mitochondrial-matrix proteins were identified in the supernatant from mitochondria that had not undergone LC-PT. Two possible explanations were proposed for the presence of these matrix proteins. These proteins either resulted from the loss of mitochondrial membrane integrity in a small population of the mitochondria, or from the controlled release of proteins from mitochondria that are functioning normally. Further work is required to determine the
reasons for the presence of matrix proteins in the supernatant. One of the problems encountered in this work was the identification of peroxisomal proteins in the mitochondrial preparations. The contamination of an organelle preparation with other cellular components is a common problem and careful methods must be used to avoid this problem in proteomics applications. Although the separation and identification of mitochondrial proteins was obtained in the work presented in Chapter 3, poor reproducibility made reliable assignment of interesting proteins extremely difficult. The laborious procedures, high probability of contamination and difficulty in obtaining equivalent staining intensity demonstrated some of the challenges associated with 2-dimensional electrophoresis. The introduction of highly automated procedures has improved gel reproducibility and minimised contamination introduced through human handling. Access to automated procedures, however, is generally limited to large, specialist laboratories. It is likely that conventional proteomics projects will become more confined to these types of laboratories.

Functional proteomics is an area of proteomics that has developed more recently and is well suited to laboratories with a lower capability for high throughput. A significant amount of information can be obtained through the identification of proteins that associate with a known protein. In the work presented in Chapter 4, affinity chromatography followed by ion exchange chromatography and nanoESI-MS/MS was used to identify proteins that associate with the T cell surface protein CD4. The identification of CD4 and p56<sup>lck</sup> established that the method was appropriate for the isolation of CD4 and co-associating molecules. The identification of other proteins such as tubulin and myosin was of interest and preliminary studies with tubulin have shown promising results (Sloane, 2000). The significance of these and other identified proteins is yet to be determined. Further development of the method may reveal whether these
identifications have biological relevance. In the longer term, the data presented here provides a basis for the identification of further binding partners of CD4, which may help to reveal potential targets in the treatment of HIV and other immune disorders.

The applicability of MS to the analysis of noncovalent complexes formed between biopolymers and ligands has been an area of much interest. A consequence of the gentle conditions of ESI-MS is that complexes of this type can be maintained intact through transfer to the gas phase. The degree to which the behaviour of gas phase complexes reflects condensed phase properties is not yet clear. Complexes of the protein glutathione S-transferase (GST) have been studied in detail by NMR, X-ray crystallography and other methods, and therefore provide a useful model for MS studies. ESI-MS revealed the formation of homodimers of GST, as expected from NMR, X-ray crystallography and other studies (this work and (Ishigai et al., 2000)). The formation of heterodimers between GST proteins from different enzyme classes was not observed in this work. This observation is in agreement with the literature. ESI-MS of a mixture of an S-alkylglutathione with GST showed the binding of two molecules of the S-alkylglutathione to the protein dimer was preferred. Furthermore, the strength of binding of S-alkylglutathione inhibitors to the GST dimers was shown to increase with increasing alkyl chain length. The binding of glutathione to GST, however, did not clearly show the expected stoichiometry of two ligands per dimer. The reasons for this discrepancy are not immediately apparent. Nonetheless, the results demonstrated that ESI-MS is a powerful technique for the analysis of noncovalent complexes and has the advantages of speed, sensitivity and the ability to measure molecular mass with high accuracy. The results also demonstrated that the data must be interpreted with care. With further understanding of the structure and behaviour of complexes as they are transferred to the gas phase, differences between results obtained from ESI-MS and
other techniques might be accounted for. Model proteins will continue to provide important data for understanding these mechanisms.

The value of ESI-MS in current protein chemistry has been clearly demonstrated. The sensitivity, mass accuracy, mass resolution and rapid generation of data are unsurpassed by other techniques for protein identification in proteomics. The use of ESI-MS to provide complementary information to that obtained from other techniques for quaternary structure determination has also proved to be valuable in many studies. Recent developments in data acquisition and processing, and in instrument design, have created an important role for mass spectrometry in many biological applications.
REFERENCES


Australian Proteome Analysis Facility (APAF) (1998) In 2-dimensional electrophoresis workshop, Sydney, Australia.


Courchesne, P. L. and Patterson, S. (1997) BioTechniques, 22, 244-250.


References


References


