FT-ICR Mass Spectrometry of Metalloproteins

Stefan K. Weidt, MChemPhys

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SCHOOL OF CHEMISTRY
FACULTY OF SCIENCE AND ENGINEERING
THE UNIVERSITY OF EDINBURGH

April 2008
Acknowledgements

I would like to thank my supervisors Dr. Pat R. R. Langridge-Smith and Prof. Peter J. Sadler for giving me the opportunity to work on this project. In particular Peter Sadler for his support and guidance throughout this project, without which this work would not have been possible. Pat Langridge-Smith for allowing me to work within SIRCAMS and make use of the FT-ICR MS instruments, as well as the opportunity to attend many international conferences and training courses. My experiences of Pat outside of the workplace are in stark contrast to those with Peter, and these unique experiences are fondly remembered. I am grateful for the time they spent proof reading this Thesis, and the feedback given in order to finish writing.

I would like to thank all the members of PJS group, past and present, who provided many an entertaining evening, many of which were for those leaving for new locations. In particular, I would like to mention Sarah Farley and Julie Ann Lough for their friendship over the years, and their support during my thesis write-up. A few group members also donated samples without which many of the experiments presented here would not have been possible; Dr Natalia Fowlis for expression and purification of YdaE, Dr Paul Bilton for expression and purification of FbpA, Dr Viv Alberts for providing the unlabelled and $^{15}$N-labelled cisplatin along with other platinum complexes not presented here and Dr Jingjing Xu for ruthenium samples, again this work is not presented here.

I would like to thank Dr Logan Mackay for his help in the SIRCAMS lab: His support and encouragement has helped me greatly throughout my PhD, along
with many distractions from it too! I would also like to thank the other members of the SIRCAMS lab, past and present, for their friendship and support, making it a friendly and fun place to work.

I would like to thank my flatmates, David and Michael, for their support and encouragement during my writing up, who, quite simply, have been waiting (rather impatiently) for the big party at the end.

This thesis was written using free and open-source software (other than the proprietary mass spectrometry software described elsewhere in this thesis). I would like to thank all the contributors who have donated their time to the open source community. I have made extensive use of the following software GNU/Linux OS, \LaTeX (\TeX), Gnuplot, Kile, KBib\TeX, KDVI, DVIPDFMx and XFig, without which I would have been inflicted with pain and suffering at the hands of the Microsoft corporation.

I would like to thank the BBSRC for funding my studentship, RASOR and the BMSS for providing financial assistance with travel and subsistence expenses while attending courses and conferences during my PhD.

And finally, special thanks go to my mother and sister for their support and encouragement over the years, who have made me who I am today.
Course And Conference Attendance

Courses attended:


2. Laser Safety

3. Advanced Proteomics Technologies, Turku, 2006 May 08-12

Conferences attended:

1. 2nd International Uppsala Symposium on ECD, 2004 August 25-28, North Berwick


4. Poster Presentation: ThP-172 Sequence-specific platination of peptides by plantium anticancer complexes studied by MS/MS fragmentation techniques,

Publications

1. Weidt, S. K.; Mackay, C. L.; Langridge-Smith, P. R. R.; and Sadler, P. J., Platination of superoxide dismutase with cisplatin: Tracking the ammonia ligands using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), Chem. Commun., 2007, 1719-21


Publication reprints are included in Appendix B at the end of this Thesis.
Abstract

In the work described in this Thesis, Fourier transform ion cyclotron mass spectrometry (FT-ICR MS) has been used to study a number of metalloproteins. The combination of high-resolution and mass-accuracy, along with fragmentation techniques such as CID, IRMPD and ECD, is shown to be a powerful method to probe metalloproteins and metallodrug interactions.

Metalloproteins require native conditions in order to preserve the conformation of the protein and hence maintain the metal binding. By using neutral pH and predominantly aqueous conditions, it is shown that the stoichiometry of binding in the gas-phase can be investigated using mass spectrometry. These conditions are also needed to study other non-covalent interactions, such as protein-ligand and protein-protein interactions. As shown here, the inherent high resolving power and mass measurement accuracy of the FT-ICR MS also enables the oxidation state of the metal present in a metalloprotein to be identified from the isotopically resolved spectra.

Cytochrome c (12.4 kDa) and myoglobin (17.0 kDa) contain a covalently and non-covalently bound heme group respectively, each containing one Fe$^{3+}$ ion. These were used as model proteins in order to investigate the calibration, mass-accuracy and resolving power needed to determine the oxidation state of the iron atom.

Carbonic anhydrase II (29 kDa) is an enzyme that catalyses the reversible hydration of carbon dioxide and contains one Zn$^{2+}$ ion. To investigate non-covalent protein-drug interactions, the inhibitor acetazolamide was added to a
sample of carbonic anhydrase II. Nozzle-skimmer CID, IRMPD and titration methods were used to probe the strength of interaction of the carbonic anhydrase-acetazolamide complex. It was found that lower charge-states bind more strongly than higher charge-states suggesting that the protein unfolds when more protons are bound.

The zinc-finger protein YdaE (6.5 kDa), from *Escherichia coli*, was found to bind one Zn$^{2+}$ ion and some dimer formation was also observed. The iron transport protein; Ferric binding protein A (34 kDa), holo-FbpA, from *Neisseria gonorrhoeae*, was found to bind one Fe$^{3+}$ ion. Reconstitution of FbpA with isotopically enriched $^{57}$Fe citrate led to an observed shift of 1 Da in the isotope distribution. With gentle source conditions, the mass spectra revealed that synergistic phosphate or citrate anions could also form complexes with FbpA.

Studies were also undertaken on superoxide dismutase (SOD), a protein which catalyses the dismutation of the superoxide radical. Samples from bovine and human erythrocytes were examined. Each monomer of beSOD (15.6 kDa) and heSOD (15.8 kDa) contains Zn$^{2+}$ and Cu$^{2+}$ ions together with a disulfide bridge. It was possible to observe the homo dimers of beSOD and heSOD provided that gentle source conditions were employed.

FT-ICR MS was used to investigate the nature of adducts between the platinum anti-cancer drug cisplatin, and SOD, due to the publication of an unusual and controversial X-ray crystal structure. The structure showed a cisplatin adduct at His19 of beSOD, which had retained the chlorido ligands; these are usually lost in aqueous conditions. After incubation with cisplatin, the mono-platinated species was the most prominent adduct observed. The use of $^{15}$N-labelled cisplatin led to unambiguous assignment of the two ammonia ligands being retained after binding to beSOD, supporting established cisplatin solution behaviour. Mass spectra of heSOD were more complex than those for beSOD due to phosphate buffer adducts and the loss or substitution of the chlorido ligands. ECD fragmentation spectra of the isolated mono-platinum adduct of beSOD were used in
order to localise the site of modification. Two new peaks that which not present in the unmodified spectrum, were assigned to a cisplatin adduct at His19.
# Table of Contents

Declaration i  
Acknowledgements ii  
Attendance iv  
Publications vi  
Abstract vii  
Table of Contents x  
List of Tables xv  
List of Figures xvi  
Glossary xix  

Chapter 1 Introduction 1  
1.1 Biological Mass Spectrometry 1  
1.2 Electrospray Ionisation 3  
1.3 FT-ICR MS 7  
1.3.1 Principles of FT-ICR MS 8  
1.3.1.1 Cyclotron Motion 8  
1.3.1.2 Trapping Motion 12  
1.3.1.3 Magnetron Motion 12  

x
1.3.2 Excitation and Detection .............................................. 13

1.4 Fragmentation of Ions (MS$^n$) ........................................... 17

1.4.1 Nomenclature Used in Peptide Fragmentation ..................... 18

1.4.2 Collision-Induced Dissociation ...................................... 19

1.4.3 Infrared Multi-Photon Dissociation ................................ 20

1.4.4 Electron Capture Dissociation ...................................... 21

1.5 Proteins ......................................................................... 22

1.5.1 Protein Structure ....................................................... 22

1.5.2 Protein Conformation Studies by MS ............................... 26

1.5.2.1 Charge-State Distributions ...................................... 28

1.5.2.2 Hydrogen/Deuterium Exchange .................................. 29

1.5.3 Metalloproteins ......................................................... 30

1.5.3.1 Isotope Distribution and Isotope Modelling .................. 32

1.6 Aims ............................................................................ 35

Chapter 2 Experimental: Materials and Instrumentation 36

2.1 Materials ........................................................................ 36

2.2 Techniques ..................................................................... 38

2.2.1 Gel Filtration Chromatography .................................... 38

2.2.1.1 Gel Filtration HPLC ............................................. 39

2.2.1.2 PD10 Column ...................................................... 39

2.2.2 Ultrafiltration Centrifugation ....................................... 39

2.2.2.1 Ultrafiltration Using Centrifugal Units ...................... 40

2.3 Instrumentation ............................................................ 40

2.3.1 9.4 Tesla Bruker Daltonics FT-ICR MS ............................. 41

2.3.2 12 Tesla Bruker Daltonics FT-ICR MS ............................. 43

2.3.3 The Advion TriVersa NanoMate ................................... 45

2.3.4 Ion Trapping and Detection ......................................... 47

2.3.5 Instrument Tuning ...................................................... 49
Chapter 3  Mass Spectrometric Studies of Carbonic Anhydrase II  

3.1 Carbonic Anhydrase II  

3.2 Mass Spectra of CAII  
3.2.1 Denaturing Conditions  
3.2.2 Native-Like Electrospray Conditions  

3.3 Carbonic Anhydrase II Inhibitor Studies  
3.3.1 In-Source Nozzle-Skimmer Dissociation  
3.3.2 IRMPD  
3.3.3 Titration of Acetazolamide With CAII  

3.4 Top-Down Analysis of CAII  

3.5 Summary  

Chapter 4  Mass Spectra of Recombinant Metalloproteins  

4.1 YdaE: A Zinc Finger Protein  
4.1.1 Sample Preparation of YdaE  
4.1.2 Studying YdaE by Mass Spectrometry  
4.1.2.1 Investigation of apo-YdaE  
4.1.2.2 Investigation of holo-YdaE  

xii
4.1.2.3 Formation of YdaE-Dimer ................. 105
4.1.3 The pH Dependence of Zinc Binding ......... 106
4.1.4 Cadmium Exchange Reactions ................. 110
4.2 Ferric-ion Binding Protein A .................. 111
4.2.1 Expression and Purification of the FbpA Protein .... 112
4.2.2 MS Study of Wild-Type holo-FbpA .......... 114
  4.2.2.1 MS of holo-FbpA after Gel Filtration .... 114
  4.2.2.2 MS of holo-FbpA after Ultrafiltration ..... 116
4.2.3 Reconstitution and Analysis of $^{57}$Fe-FbpA .... 118
4.3 Summary .................................... 122

Chapter 5 Superoxide Dismutase (SOD) ............. 124
5.1 Study of SOD using Mass Spectrometry .......... 128
  5.1.1 Studying SOD using Denaturing Conditions .... 128
    5.1.1.1 MS of beSOD ..................... 129
    5.1.1.2 MS of heSOD .................... 130
  5.1.2 Native-Like Conditions .................... 131
    5.1.2.1 MS of beSOD Homodimer ............ 132
    5.1.2.2 MS of heSOD Homodimer ............ 133
5.2 Incubation of SOD with Cisplatin .............. 134
  5.2.1 Studies of Intact SOD-Cisplatin Adducts .... 136
    5.2.1.1 Sample Preparation ................ 136
    5.2.1.2 beSOD-Cisplatin under Native Conditions .... 137
    5.2.1.3 heSOD-Cisplatin under Native Conditions .... 140
  5.2.2 Identifying the Sites of SOD-Cisplatin Adducts .... 142
    5.2.2.1 Sample Preparation ................ 142
    5.2.2.2 ECD of beSOD-Cisplatin Adducts ........ 143
    5.2.2.3 ECD of heSOD-Cisplatin Adducts ........ 147
5.3 Summary .................................... 151
References 154

Appendix A  Source Code  187

Appendix B  Publication Reprints  193
List of Tables

1.1 The 20 commonly occurring amino acids ........................................ 23
2.1 Materials .................................................................................. 37
2.2 Agilent ESI tuning mix composition ............................................ 53
2.3 Calibration table for Agilent ESI tuning mix ................................. 54
2.4 Calibration table for CsTFHA clusters ......................................... 57
2.5 Peptides and proteins used as standards ........................................ 60
2.6 Calibration table for peptide mixture ............................................ 62
3.1 Parameters for nozzle-skimmer CID of CAII-aza complex ............ 82
3.2 Parameters for IRMPD of CAII-aza complex ................................. 83
4.1 The isotopic composition of enriched $^{57}$Fe. .............................. 121
## List of Figures

1.1 Diagram of an ESI-MS ........................................... 4
1.2 Diagram of the electrospray process .......................... 5
1.3 Diagram showing the Lorentz force ............................ 9
1.4 Diagram showing how ions are detected in FT-ICR MS ...... 15
1.5 Nomenclature of fragment ions of a peptide ................. 19
1.6 Protein secondary structure .................................... 24
1.7 Deoxyhemoglobin X-ray crystal structure .................... 25
1.8 *holo-*myoglobin X-ray crystal structure .................... 27
1.9 Charge-state distributions observed in an ESI-MS .......... 29
1.10 Example of isotope pattern for ferric-cytochrome *c* ........ 34

2.1 Bruker Daltonics 9.4 T Apex III FT-ICR MS .................. 42
2.2 The brass capillary fitted to the Apex III source .......... 42
2.3 Bruker Daltonics 12 T Apex Qe FT-ICR MS ................... 44
2.4 Advion TriVersa™ NanoMate® ................................ 46
2.5 The Bruker 'Infinity' cell ..................................... 47
2.6 +ve MS of Agilent ESI tuning mix ............................ 55
2.7 -ve MS of Agilent ESI tuning mix ............................. 55
2.8 +ve MS of CsTFHA clusters .................................. 58
2.9 -ve MS of CsTFHA clusters .................................. 58
2.10 +ve MS of a peptide mix .................................... 61
2.11 +ve MS of ubiquitin ......................................... 63
4.10 FbpA X-ray crystal structure ........................................ 113
4.11 +ve MS of holo-FbpA (gel filtration) .......................... 115
4.12 +ve MS of holo-FbpA (ultrafiltration) .................... 117
4.13 Reconstitution of apo-FbpA with $^{57}$Fe .................. 120
4.14 Isotope fitting for the 12+ charge-state of $^{57}$Fe-holo-FbpA .. 121

5.1 beSOD and heSOD sequences ..................................... 125
5.2 beSOD and heSOD X-ray crystal structures .................. 127
5.3 +ve MS of monomeric apo-beSOD denaturing ............... 129
5.4 +ve MS of monomeric apo-heSOD denaturing ................ 131
5.5 +ve MS of dimeric holo-beSOD native ...................... 132
5.6 +ve MS of dimeric holo-heSOD native ..................... 134
5.7 beSOD-cisplatin X-ray crystal structure ................... 135
5.8 +ve MS of dimeric holo-beSOD after incubation with cisplatin . 138
5.9 Natural and $^{15}$N-labelled monoplatinated adduct of beSOD .... 139
5.10 Dimeric holo-heSOD after incubation with cisplatin .... 141
5.11 Monomeric apo-beSOD after incubation with cisplatin .... 144
5.12 ECD spectrum for the mono-cisplatin-beSOD adduct .... 145
5.13 Identified beSOD-cisplatin ECD fragments ............... 146
5.14 ECD fragments mapped to the beSOD sequence .............. 147
5.15 Monomeric apo-heSOD after incubation with cisplatin .... 148
5.16 ECD spectrum for the mono-cisplatin-heSOD adduct .... 150
5.17 ECD fragments mapped to the heSOD sequence ............ 151
Glossary

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>Molar extinction coefficient</td>
</tr>
<tr>
<td>$m/z$</td>
<td>Mass-to-charge ratio</td>
</tr>
<tr>
<td>A</td>
<td>Absorbance</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>APCI</td>
<td>Atmospheric pressure chemical ionisation</td>
</tr>
<tr>
<td>API</td>
<td>Atmospheric pressure ionisation</td>
</tr>
<tr>
<td>aza</td>
<td>Acetazolamide</td>
</tr>
<tr>
<td>beSOD</td>
<td>Bovine erythrocyte superoxide dismutase</td>
</tr>
<tr>
<td>CAII</td>
<td>Carbonic anhydrase (bovine isozyme II)</td>
</tr>
<tr>
<td>CID</td>
<td>Collision-induced dissociation</td>
</tr>
<tr>
<td>Cisplatin</td>
<td>cis-diamminedichloridoplatinum(II), $\text{cis-}[\text{PtCl}_2(\text{NH}_3)_2]$</td>
</tr>
<tr>
<td>CRM</td>
<td>Charge residue model</td>
</tr>
<tr>
<td>CsTFHA</td>
<td>Cesium tridecafluoroheptanoate</td>
</tr>
<tr>
<td>CTAB</td>
<td>Cetyl trimethyl ammonium bromide</td>
</tr>
<tr>
<td>DA</td>
<td>Bruker Daltonics DataAnalysis software</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
</tbody>
</table>

xix
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTNB</td>
<td>5,5’-dithiobis(2-nitrobenzoate) / Ellman’s Reagents</td>
</tr>
<tr>
<td>ECD</td>
<td>Electron-capture dissociation</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionisation</td>
</tr>
<tr>
<td>FALS</td>
<td>Familial amyotrophic lateral sclerosis</td>
</tr>
<tr>
<td>FbpA</td>
<td>Ferric-ion binding protein A</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>FT-ICR</td>
<td>Fourier transform-ion cyclotron resonance</td>
</tr>
<tr>
<td>FT-ICR MS</td>
<td>Fourier transform-ion cyclotron resonance mass spectrometry</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full width at half magnitude</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid</td>
</tr>
<tr>
<td>heSOD</td>
<td>Human erythrocyte superoxide dismutase</td>
</tr>
<tr>
<td>ICR</td>
<td>Ion cyclotron resonance</td>
</tr>
<tr>
<td>IEM</td>
<td>Ion evaporation model</td>
</tr>
<tr>
<td>IPA</td>
<td>Isopropyl alcohol, isopropanol, 2-propanol</td>
</tr>
<tr>
<td>IRMPD</td>
<td>Infrared multi-photon dissociation</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix assisted laser desorption / ionisation</td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular weight cut-off</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
</tbody>
</table>
PES  Polyethersulfone
PTM  Post-translational modification
RF   Radio frequency
SOD  Superoxide dismutase
SORI Sustained off-resonant resonance excitation
TCEP Tris(2-carboxyethyl)phosphine hydrochloride
Tris Trishydroxymethylaminomethane / 2-amino-2-hydroxymethyl-1,3-propanediol
UHV  Ultra high vacuum
Chapter 1

Introduction

1.1 Biological Mass Spectrometry

In 2002, half of the Nobel prize for chemistry was awarded to John Fenn and Koichi Tanaka for their work on electrospray ionisation (ESI)\(^1\) and matrix assisted laser desorption / ionisation (MALDI),\(^2\) respectively. These two ionisation techniques have allowed the study of large biological molecules by mass spectrometry. These techniques continue to mature, along with instrument developments, making it now possible to study molecular weights of the order of Mega-Daltons (MDa).\(^3\)\(^-\)\(^5\)

Since the initial publication of the human genome,\(^6\) proteomics has been a growing field and has been reviewed many times over recent years.\(^7\)\(^-\)\(^12\) While genes describe the code for the transcription of proteins in a cell, they do not however describe the regulation, any subsequent post-translational modifications (PTMs), the location of each protein within a cell, or how these proteins interact with each other. Proteomics is therefore the study and mapping of proteins from each cell of an organism. In traditional proteomics, the complex mixture of proteins from a cell lysate would be separated by using 2D-gel electrophoresis.\(^13\) This is usually done by first separating the proteins by charge using isoelectric point focusing in the first dimension, and then by molecular weight using sodium dodecyl sulfate electrophoresis in the second. This results in a gel that can be stained to reveal a distribution of spots. The spots can then be cut from the gel and sequenced
using automated Edman degradation. Mass spectrometry techniques however offer greater sensitivity and throughput.\textsuperscript{14} Spots are identified after digestion using a proteolytic enzyme, typically trypsin, and the resulting fragments are analysed by MS and matched against a protein database. Alternatively these fragments can be separated by liquid chromatography (LC) coupled to ESI and sequenced using tandem mass spectrometry (MS/MS). The peptide fragments are then validated and matched against a protein database.\textsuperscript{15} The term ‘Shotgun proteomics’ has been used to describe digestion of the protein mixture from a cell lysate prior to any separation, removing the need for 2D-gel electrophoresis.

Analysis of intact proteins is gaining increasing interest as fragmentation of an intact protein in the mass spectrometer can yield powerful information on the location of PTMs.\textsuperscript{16,17} This approach is known as ‘Top-down MS’ and Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR MS) plays a pivotal role in this approach.\textsuperscript{18} The fragmentation technique of electron-capture dissociation (ECD) has proven particularly powerful for sequencing the backbone and identifying the sites of protein modifications.\textsuperscript{19–21}

‘Interaction proteomics’ involves investigating in which proteins interact with each other, which is usually studied by isolating a complex. The proteins are then resolved using denaturing gel electrophoresis and identified by digestion of the separated protein bands.\textsuperscript{22,23} Another approach is to study intact complexes using mass spectrometry, which has the benefits of rapid analysis, elucidates the stoichiometry and provides dynamic real-time detection.\textsuperscript{24,25} These complexes are usually non-covalent in nature and ESI has really come into its own in this field.\textsuperscript{26} It is not just the protein-protein interactions that have been studied, but also protein-peptide, protein-ligand, protein-DNA and protein-metal ion complexes to name just a few. There have been several reviews on the study of non-covalent complexes using MS.\textsuperscript{27–30}

For maximum sensitivity using ESI, acidic conditions (pH 2-4) are employed for positive ions and alkaline conditions (pH 8-10) for negative ions. High levels
of organic solvent (usually methanol or acetonitrile) are added to aqueous solutions in order to enhance and stabilise the signal. Non-covalent complexes are typically stable in neutral conditions (pH 6-8) but are disrupted by high levels of organic solvent. These two conditions are often referred to as simply 'denaturing' or 'native', respectively. A volatile buffer is needed in order to maintain a neutral pH while being compatible with the ESI process. Two buffers have proven particularly useful: ammonium acetate and ammonium bicarbonate, but there are other examples. The addition of imidazole has also been reported to stabilise complexes in order to allow their detection.

1.2 Electrospray Ionisation

The work presented in the following Chapters has relied heavily on the use of the electrospray ionisation technique. Some examples of successes from using ESI have been stated in the previous section. ESI directly samples an analyte from solution, even large nonvolatile molecules, as long as it is sufficiently polar or charged. Peptides and proteins contain many different amino acids, many of which have side-chains that can become positively or negatively charged, depending on the pH of the solution. This has meant that ESI is particularly well suited for the study of biological samples and has been hugely successful since its invention.

The solution containing the analyte is pumped through a metal needle which is held at a high potential (positive or negative) with respect to the counter electrode. The high electric field gradient at the tip of the needle causes the surface of the emerging liquid to become charged. The Coulomb repulsion overcomes the surface tension and a fine spray of charged droplets is formed. A heated counterflow of drying gas or a heated source block is used to encourage the solvent to evaporate. As the solvent evaporates from the droplet surface, the droplets decrease in size increasing the charge density at the surface. Once the charge density overcomes the surface tension (known as the 'Rayleigh limit'), droplet fission occurs forming smaller droplets until the analyte is ionised. This process occurs
at atmospheric pressure and so a differential pumping scheme is required to reach the high vacuum required in the mass spectrometer. The ions pass through a small entrance orifice and down a transfer capillary, with the first stage of pumping, followed by a skimmer, which proceeds the second stage of pumping, behind which further pumping stages and ion guides may be present as the ions travel through the mass analyser and reach the detector. A schematic diagram of this setup is shown in Figure 1.1.

Figure 1.1: Diagram of a simple mass spectrometer implementing an ESI source with differential pumping.

The mechanism for the electrospray process is still under debate, however two mechanisms have been widely discussed. At the tip of the ESI needle, the high electric field causes repulsion of the positive ions (or negative ions in negative ionisation mode) distorting the meniscus into what is known as a Taylor cone, as shown in Figure 1.2a. At the tip of the Taylor cone, a fine filament forms. The Coulombic repulsion drives the ejection of droplets once they overcome the surface tension of the liquid. This process is described as ‘budding’. The emerging droplets then spread out due to Coulombic repulsion between them, forming an expanding plume. The droplets then reduce in size as the solvent rapidly evaporates at the surface, increasing the charge density at the surface of the
Figure 1.2: Diagram of the electrospray process. a) Formation of the Taylor cone at the tip of the ESI spray needle due to electrostatic repulsion. A fine filament is formed at the tip of the Taylor cone from which small droplets overcome the surface tension called 'budding'. The repulsive forces between droplets produce an expanding plume of droplets. b) Formation of single ions for the two mechanisms; IEM and CRM models from positively-charged droplets.
drop. Once the Coulombic repulsion overcomes the cohesive force of the surface tension, the Rayleigh limit, the droplet then splits reducing the charge and some of the mass from the parent drop. This process is known as ‘droplet fission’. Experiments by Gomez and Tang provide some insights into this process. They observed droplets formed from an electrospray of heptane using flash photography. Droplet fission began at only 70-80% of the Rayleigh limit. The offspring droplets were ejected from a filament that is drawn out of the droplet, in a similar fashion to the ‘budding’ from the Taylor cone. It is important to note that the droplet sizes were much larger and different solvents were used than those typically used for ESI-MS. It is at this point in the mechanism that the two models differ.

The charge residue model (CRM) has been attributed to early discussions by Dole et al. This model describes the droplet fission process until droplets are formed which contain a single ion. The final evaporation of solvent leads to ions entering the gas phase. The ion evaporation model (IEM), has been attributed to the ideas of Iribarne and Thomson. This model proposes that ions evaporate directly from an evaporating charged droplet when there is a sufficiently high applied electric field. This process is in competition with the Rayleigh limit of instability, so droplet fission will continue and ions can evaporate from the offspring droplets. These processes are represented diagrammatically in Figure 1.2b. There are arguments both for and against each of the mechanisms and it is likely that both models occur and the molecular weight dictates which is dominant. In the specific case of large biomolecules which carry multiple charges, the ESI process appears to be consistent with the CRM mechanism, but the debate continues.

Irrespective of the mechanism of how solvent-free analyte ions are produced, the desolvation process is crucial for sensitivity and stability of the detected ESI signal. In particular the surface tension plays an important role and so the solvent composition needs to be considered. There is a fine limit between electrospray of ions and corona discharge, especially for water solutions. The addition of volatile
organic solvent such as methanol or acetonitrile helps reduce the surface tension and aids the desolvation process. Use of a nebulizing gas, usually nitrogen, can help disperse large droplets, also aiding the desolvation process. Additives too have been introduced into the nebulizing gas flow to help prevent electric discharge during electrospray.\textsuperscript{44} In negative ionisation mode, the onset of electric discharge is even closer to electrospray voltages,\textsuperscript{45} which helps explain why this polarity has a tendency to be less sensitive. A stable spray is not only important for high sensitivity and good signal-to-noise, but can also affect the observed mass spectrum due to changes in the size of the droplet emerging from the spray.\textsuperscript{46,47} Further details on the practicalities of ESI are summarised and reviewed in the literature.\textsuperscript{48}

Miniaturisation of the electrospray technique has proven highly successful and is termed nano-electrospray ionisation (nano-ESI).\textsuperscript{49} This technique uses a tapered glass capillary that has been pulled using heat or laser irradiation to produce a very fine orifice at the tip. A metal coating or a platinum wire is inserted into the liquid to make electrical contact. An applied voltage of 500-1500 V is needed for nano-ESI. The spray is produced entirely from the electrostatic field with flow rates of around 20-50 nL/min. This results in only a few microlitres of sample solution being required for analysis, vital for when sample is limited. There is also an advantage of increased sensitivity and greater tolerance to salt and contaminants, proving particularly useful for analysis of biological samples. This has been attributed to reduction of the initial droplets formed in a nano-ESI plume.

\section*{1.3 FT-ICR MS}

The principles of FT-ICR MS have changed little over recent years. Instrument developments have included improvements in electronics and superconducting magnet technology leading to higher sensitivity and resolution. Magnetic field strengths of 7 T are now common with 9.4, 12 and 15 T increasingly more
available. Hybrid instruments using either a quadrupole or a linear ion trap have allowed for much more versatile experiments, particularly in the automation needed for proteomic experiments.\textsuperscript{50} This Section covers some of the fundamental aspects of FT-ICR MS, with focus on the implications that these aspects have on studying biomolecules. Far more detail on all aspects of the fundamentals and implementation of FT-ICR MS can be found in several extensive review articles.\textsuperscript{51–53} The milestones and a historical account of FT-ICR MS have also been well documented.\textsuperscript{54,55}

1.3.1 Principles of FT-ICR MS

The motion of ions inside an FT-ICR analyser cell are governed by the magnetic and electric fields that are present. This motion can be broken down into three independent components: cyclotron motion, trapping motion and magnetron motion.

1.3.1.1 Cyclotron Motion

An ion with mass \( m \) (kg) with charge \( q \) (C) travelling at a velocity \( v \) (m s\(^{-1}\)), in a spatially uniform magnetic field \( B \) (Tesla), will experience a force \( F \) (N) with acceleration \([dv]/[dt]\) (m s\(^{-2}\)) given by Equation 1.1 known as the Lorentz force.

\[
F = m \frac{dv}{dt} = qv \times B
\]  

(1.1)

Simply put, the Lorentz force is a vector cross product of the magnetic field and the velocity, multiplied by the charge on the particle. An example of the Lorentz force acting on a positive ion is shown in Figure 1.3. If the polarity of the ion is changed then the Lorentz force will act in the opposite direction.

For an ion that maintains a constant speed, i.e no collisions, the resulting Lorentz force will bend the ion path into an arc of radius \( r \) (m), known as cyclotron motion. If the force is strong enough the resulting path of the arc will form a complete circle. If the magnetic field lies along the \( z \)-axis, then \( B = B_0 \mathbf{k} \) and the
Figure 1.3: Diagram showing that the Lorentz force, \( F \), leads to cyclotron motion, \( \omega \), for a particle of mass, \( m \), with charge, \( q \), moving in a uniform magnetic field \( B \).

Instantaneous linear velocity in the \( xy \)-plane can be defined as \( v_{xy} \). Substituting the angular acceleration \( [d\mathbf{v}/dt] = v_{xy}^2/r \) into Equation 1.1 leads to Equation 1.2.

\[
\frac{mv_{xy}^2}{r} = qv_{xy}B_0
\]  

(1.2)

The angular frequency \( \omega \) (rad s\(^{-1}\)) about the \( z \)-axis is defined by

\[
\omega = \frac{v_{xy}}{r}
\]  

(1.3)

so that Equation 1.2 becomes

\[
m\omega^2r = qB_0\omega r
\]  

(1.4)

or simplifying

\[
\omega_c = \frac{qB_0}{m} \quad \text{(S.I. Units)}
\]  

(1.5)

Equation 1.5 is known as the 'unperturbed cyclotron' equation in which the ion cyclotron frequency is denoted as \( \omega_c \) (rad s\(^{-1}\)) in standard S.I. units. It is an important feature that the cyclotron motion is not effected by the kinetic energy of the ions. This property makes ICR especially useful for mass spectrometry as a narrow energy distribution is not essential for precise determination of \( m/q \).
In magnetic sector and time-of-flight mass analysers, for example, the resolution that can be achieved is limited by the spread in distribution of the kinetic energy of the ions.

Equation 1.5 can be expressed in terms of a cyclotron frequency, \( \nu_c \) (Hz) and the more familiar mass-to-charge ratio used in mass spectrometry, \( m/z \), where the mass \( m \) (Da) and \( z (q/e) \).

\[
\nu_c = \frac{\omega_c}{2\pi} = \left( \frac{e}{2\pi m} \right) \frac{B_0}{m/z} = \frac{1.5 \times 10^7 B_0}{m/z} \tag{1.6}
\]

As an example, an ion of mass-to-charge of 1000 \( m/z \) in a 9.4 T uniform magnetic field will have a cyclotron frequency of 144 kHz. This highlights another important point, that the detected cyclotron frequencies for typical ions used in MS lie in the range of a few kHz to a few MHz which is convenient for commercially available electronics. Rearranging Equation 1.6 leads to,

\[
m/z \propto \frac{B_0}{\nu_c} \tag{1.7}
\]

showing the fundamental relation that mass-to-charge \( m/z \) is linearly proportional to the magnetic field \( B_0 \) and inversely proportional to the cyclotron frequency, \( \nu_c \).

The cyclotron orbital radius, \( r \) (m) (Equation 1.8) can be found by rearranging Equation 1.2:

\[
r = \frac{mv_{xy}}{qB_0} \quad \text{(S.I. Units)}. \tag{1.8}
\]

The average \( xy \) translational energy of an ion in thermal equilibrium with its surroundings at a temperature \( T \) (K) is given by Equation 1.9:

\[
\frac{1}{2} m \langle v_{xy}^2 \rangle \approx kT \tag{1.9}
\]

in which \( k \) is the Boltzmann constant (J K\(^{-1}\)). Solving Equation 1.9 for \( v_{xy} \) and substituting into Equation 1.8 yields:

\[
\langle r \rangle \approx \frac{1}{qB_0} \sqrt{2mkT} \quad \text{(S.I. Units)} \tag{1.10}
\]
Converting to the usual \( m/z \) units, Equation 1.10 becomes:

\[
\langle r \rangle \approx \left( \frac{\sqrt{2}k_{\text{u}}}{e} \right) \frac{\sqrt{mT}}{zB_0} \approx 1.3 \times 10^{-6} \frac{\sqrt{mT}}{zB_0} \tag{1.11}
\]

As an example, at room temperature (298 K) using a 9.4 T magnet, a singly-charged ion with a mass of 1000 Da will have an average cyclotron radius of 0.078 mm. This is much smaller than the diameter of the FT-ICR analyser cell (usually several cm) and so the ions must be excited to a higher orbital radius to induce a detectable image current. By increasing the orbital radius, the ions gain velocity in the \( xy \)-plane.

The velocity in the \( xy \)-plane, \( v_{xy} \), can be calculated by rearranging Equation 1.8 to give:

\[
v_{xy} = \frac{qB_0r}{m} \tag{1.12}
\]

The ‘Infinity’ cell used in Bruker Daltonics FT-ICR MS instruments (see Section 2.3.4) is cylindrical in shape and has a diameter of \( \approx 6 \text{ cm} \). For a singly-charged ion of mass 1000 Da excited to a radius of 3 cm, the translational velocity will be 816 m s\(^{-1}\), with a translational kinetic energy of 408 eV:

\[
K.E. = \frac{1}{2}mv_{xy}^2 = \frac{q^2B_0^2r^2}{2m} = \left( \frac{e^2}{2u} \right) \frac{z^2B_0^2r^2}{m} = 4.8 \times 10^7z^2B_0^2r^2 \tag{eV} \tag{1.13}
\]

A 10+ charged ion of mass 10 000 Da would have a kinetic energy of 4081 eV, and for a 50+ charged ion of mass 50 000 Da the kinetic energy would be 20.4 keV. All these ions would have a \( m/z \) value of 1000 and thus the same cyclotron frequency. This illustrates one of the reasons why large highly charged proteins are difficult to detect using FT-ICR MS as the excitation needed for detection leads to very high kinetic energies. This high kinetic energy requires an ultra high vacuum to prevent collisional fragmentation. Fragmentation can be deliberately achieved by pulsing an inert collision gas into the cell and exciting the ions, see Section 1.4.2 for more details.
1.3.1.2 Trapping Motion

An ion moving parallel to the magnetic field experiences no force from the field and would therefore follow a helical motion as it continues cyclotron motion in the $xy$-plane. Ion motion needs to be constrained in the direction of the magnetic field in order to trap the ions within the region of homogeneous magnetic field for detection. Ions are constrained by applying a potential of the same polarity as the ions to two parallel plates, or cylindrical plates in the case of an open ended cell, that are perpendicular to the magnetic field. The ions undergo simple harmonic motion between the two trapping plates along the magnetic field axis. The frequency of oscillation in the $z$-axis, $\nu_z$ (Hz), along the magnetic field direction is,

$$\nu_z = \frac{1}{2\pi} \sqrt{\frac{2qV_{\text{trap}}}{ma^2}} \quad \text{(S.I. Units)} \quad (1.14)$$

$$\nu_z = \sqrt{\frac{e}{2\upi^2} \frac{zV_{\text{trap}}}{ma^2}} = 2.2 \times 10^3 \sqrt{\frac{zV_{\text{trap}}}{ma^2}} \quad (1.15)$$

where $V_{\text{trap}}$ (V) is the trapping voltage, $a$ (m) is the cell dimension from the centre to the trapping plate and $\alpha$ is a parameter dependent on the cell geometry. It is clear that this trapping motion is independent of the cyclotron motion. For an ion of $m/z = 1000$, confined by a trapping voltage $V_{\text{trap}} = 1$ V, in an ‘Infinity’ cell, ($\alpha = 2.8404$ and $a = 4.5$ cm) the ions will oscillate at a frequency of $2618$ Hz.$^{1}$

1.3.1.3 Magnetron Motion

The combination of magnetic and electric fields creates a three-dimensional ion trap. This allows ions to be stored in the ICR analyser trap for seconds, minutes or even hours, which is many orders of magnitude higher than other types of mass spectrometer. Although the cyclotron and axial oscillations are not coupled, the combination of magnetic and electric fields together produce a third fundamental motion of the ions, called magnetron motion. The magnetron motion is a side

$^{1}$Values were determined from physical measurements of the ‘Infinity’ cell and Table 1 of Marshall et al.$^{52}$
effect from the trapping potential, and like the trapping motion is related the
analyser cell design. Applying a potential to the trapping plates while holding
the excitation and detection plates at ground has the consequence of a non-zero
field at the centre of the cell. The effect of this field is to produce a force in the
$xy$-plane that is radially repulsive; the electric field acts to drive the ions away
from the centre of the analyser cell. The electric and magnetic fields combine to
produce magnetron motion, a precession of the cyclotron motion about the centre
of the cell. The frequency of the magnetron motion is given by Equation 1.16.

$$\nu_m = \frac{\alpha V_{\text{trap}}}{\pi a^2 B_0}$$

(1.16)

It is clear that the magnetron motion is independent of $m/z$ and is directly
proportional to the trapping potential while being inversely proportional to the
magnetic field. This means that magnetron motion can be reduced by using
lowering trapping potentials. Moving to a higher magnetic field also has the
benefit of reducing the effect of the magnetron motion.

As an example, for a typical applied trapping potential of 1.1 V in a magnetic
field of 9.4 T, the magnetron motion for the ‘Infinity’ cell ($a=4.5$ cm, $\alpha=2.8404$)
has a characteristic magnetron frequency of 52 Hz, which is much lower than cy-
clotron frequencies. Magnetron motion serves no analytical purpose, it perturbs
the cyclotron frequency slightly leading to shifts in the measured $m/z$. With care-
ful calibration these shifts can be corrected and do not normally pose a problem.

1.3.2 Excitation and Detection

Ion cyclotron resonance motion by itself does not generate a detectable signal.
After the ions are injected into the analyser cell, an ion may start its cyclotron
motion at any point on its orbit. When combined with many hundreds or thou-
sands of ions, their motion will be distributed over the cyclotron radius without
any coherent phase. This results in no net difference in induced charge on the
opposing parallel detection plates.
At thermal equilibrium, as shown by Equation 1.10, the cyclotron radius is very small compared to the dimensions of the analyser cell; too small to detect a signal even if the ions at each mass-to-charge have the same phase, as shown by Figure 1.4a. In order to increase the cyclotron radius to a detectable range, the ions need to be excited using an applied resonant frequency from opposing parallel excitation plates. During resonant excitation, the ions spiral out, increasing their radius, as shown in Figure 1.4b, while ions not in resonance stay in the centre of the cell. If the RF excitation voltage were to be continually applied, then the radius of motion of the ions would continue to increase until they collided with the electrodes of the analyser cell where they would be neutralised. Resonant excitation needs to be applied only for a short time, after which the ions continue cyclotron motion but at a much larger radius. This method can be used to eject ions that are not wanted in the mass spectrum. This is known as isolation and this is discussed in more detail in Section 1.4.

All ions of a particular mass-to-charge ratio are excited coherently, which means that they are grouped as tightly as they were initially. Ions of the same mass-to-charge ratio undergo cyclotron motion as a packet. The coherent ion packet will induce a signal across the detection plates, as there will be a difference between the charge induced on each electrode. As the ion packet passes each electrode, electrons will be attracted to the plate for positive ions, while negative ions will repel electrons away from the electrode, as shown in Figure 1.4c. Electrons pass through the external circuit that joins the detection plate, resulting in an alternating current, known as the image current. The signal is amplified and then digitised prior to data processing. The image current allows for non-destructive detection of the ions, unlike any other detector, such as an electron multiplier or micro-channel plate.

Sweeping through a range of excitation frequencies allows for simultaneous detection of ions at each mass-to-charge ratio. This is known as broadband excitation and detection. The most common broadband excitation is the frequency
Figure 1.4: Diagram showing how ions are detected in FT-ICR MS. a) Ions are trapped inside the cell with small initial cyclotron radius. b) Ions are excited using an RF sweep across the range of $m/z$ of interest to form a coherent ion packet at a higher cyclotron radius. c) The coherent ion packet induces an image current on the detection plates and this signal is amplified prior to Fourier transformation and calibration.
sweep, or RF chirp, in which the frequency generator sweeps rapidly through a range of frequencies.\textsuperscript{56} The result is a superposition of sinusoidal oscillations, of different frequencies and amplitudes, from all of the ions at each cyclotron frequency. Fourier transformation of the image current gives the component frequencies and abundances, leading to the same frequency spectrum that could be obtained by measuring the power absorption while sweeping infinitely slowly though the entire \( m/z \) range. The advantage of broadband excitation and detection is that a wide range of \( m/z \) values can be measured simultaneously in the same time it would take to measure just one frequency response. Once the frequency spectrum is obtained, the mass-to-charge spectrum is calculated using Equation 1.6.

The length of time over which signal is detected is a major factor in the measured resolving power and the peak shape of the isotopic distribution.\textsuperscript{57} This is another reason why a high vacuum is required for FT-ICR MS operation. The amplitude of the transient signal decays with time as collisions between the ions and neutrals destroy the coherence of the ion packet. The peak shape deviates from the ideal as the transient gets truncated. At first the peaks broaden resulting in reduced resolving power. At high pressure the peak shapes turn from a sinc function to Lorentzian, appearing sharper at half magnitude, but the base is much broader.

From the first derivative of Equation 1.5 with respect to mass,

\[
\frac{d\omega_c}{dm} = \frac{-qB_0}{m^2} = -\frac{\omega_c}{m}
\]

the useful relation,

\[
\frac{\omega_c}{d\omega_c} = -\frac{m}{dm}
\]

is obtained. The resolving power is defined as \( \omega_c/\Delta \omega_{50\%} \) or \( m/\Delta m_{50\%} \), which is the measured peak centre divided by the full width of the peak at half of the magnitude. The frequency resolving power and mass resolving power are equal in magnitude but opposite in sign. The resolution limit is where two peaks of equal
height and shape can be distinguished. Since the frequency of an FT-ICR peak is approximately \( qB_0/m \), the experimental resolving power can be expressed as

\[
\frac{m}{\Delta m_{50\%}} = -\frac{qB_0}{m\Delta\omega_{50\%}} \quad \text{(S.I. Units)}
\]  

(1.19)

showing that the resolving power is linear with magnetic field strength but inversely proportional to the mass-to-charge ratio.

In practice, the measured resolution is influenced by the transient acquisition time, the number of data points that are used to represent the transient, collisional damping, magnetron motion, the magnetic field strength and the mass of the ion being measured.

### 1.4 Fragmentation of Ions (MS\(^n\))

Measuring the precursor mass of an ion at high resolution is a source of much information. Sometimes it is useful to be able to discover some information on the structure of a particular ion. By isolating an ion of interest and subjecting it to fragmentation, the resulting fragments can be used to elucidate the structure of the precursor ion. This is known as MS/MS or MS\(^2\). The fragment ions can also be isolated and fragmented in a repeated fashion if there is sufficient ion abundance at each stage. In general this is known as MS\(^n\).

Isolation of ions from inside the ICR cell at a particular \( m/z \) can be achieved by exciting ions until they are ejected from the cell while retaining the ions of interest. This can be performed by a correlated frequency sweep, which leaves a gap in the excitation frequencies that the instrument scans through. A much more sophisticated method is to use a stored waveform inverse Fourier transform (SWIFT), which produces an excitation waveform which isolates the ions of interest with greater efficiency and flexibility than the correlated sweep method.\(^{58}\)

External isolation, outside of the ICR cell, using a hybrid design has the advantage of not inadvertently exciting the ions of interest and displacing them from the centre of the cell, which is necessary for some fragmentation techniques.\(^{59,60}\)
Isolation is used to simplify the fragmentation spectrum, as the only fragments seen will belong to the isolated parent ion.

Many fragmentation techniques have been adapted to FT-ICR MS. The following fragmentation techniques are common in FT-ICR MS and take place inside the analyser cell. Collision-induced dissociation (CID) causes fragmentation by excitation of ions in a region containing relatively high pressure of an inert gas. The excitation increases the kinetic energy of the ions which undergo collisions with the gas causing vibrational excitation leading to fragmentation of the molecule. Infrared multi-photon dissociation (IRMPD) uses an infrared CO₂ laser to excite molecular ions to higher vibrational levels. Like CID, once there is sufficient energy for bond dissociation, fragment ions are produced. Electron capture dissociation (ECD) is a technique that fires low-energy electrons at peptide ions charged by protons. The energy from this capture event is transferred into the peptide backbone which causes fragmentation.

Details of the fragmentation pathways are not discussed here as the focus is introducing terms that are used in the later Chapters of this thesis. The aim is to apply MS/MS techniques in order to expand the information gained during MS experiments. Extensive reviews are available that cover the fragmentation mechanisms and discuss these techniques in detail.

1.4.1 Nomenclature Used in Peptide Fragmentation

The nomenclature used to describe where the peptide bond breaks is shown in Figure. The peptide bond can be cleaved in one of three places. After the bond breaks, there is usually some gas-phase rearrangement to a more stable ion. During this rearrangement, the charge can reside on either the N-terminus or C-terminus of the peptide sequence. If the charge resides on the N-terminal side then the three fragments are a, b and c ions. A subscript is used to indicate at which residue along the peptide sequence the bond has been cleaved. If the charge resides on the C-terminal side then the corresponding x, y and z ions are
formed while the numbering starts from the C-terminus. It is also possible to cleave the side-chains of the amino acid and, if these remain charged, to form d, v and w ions. Internal fragments, where bond cleavage occurs at more than one place along the backbone, is also possible. The complexity of the resulting fragmentation spectra is increased when side-chain and internal fragments are also present making direct interpretation from the sequence more difficult.

1.4.2 Collision-Induced Dissociation

As indicated by Equation 1.13, exciting the ions to a higher cyclotron radius, increases the kinetic energy of the ions. An inert collision gas, such as argon, can be pulsed into the analyser cell. When the excited ions collide with the gas molecules they become vibrationally excited, leading to fragmentation.62,74 Smaller molecular ions tend to fragment more easily than larger molecular ions. Brief on-resonant excitation is all that is required for small molecules. Larger molecules may require many more collisions to cause fragmentation. If on-resonant excitation is applied for an extended time, the orbit radius of the ions would increase to beyond the analyser cell dimensions before fragmentation occurs. By applying a sustained
off-resonant resonance excitation (SORI), just a few kHz from resonance, the ions spiral out to a higher radius.\textsuperscript{75} The excitation frequency becomes out-of-phase causing the ions to collapse slowly back to the centre of the cell until the excitation is in-phase once more, increasing radius again, and the process repeats over and over. This method results in many more collisions over a longer excitation time period while preventing the ions being excited to beyond the analyser cell dimensions. Collisions with gas molecules lead to an increase in vibrational energy and after several collisional events there is sufficient vibrational energy to cause bond dissociation. In SORI-CID the pulse-gas reservoir pressure, the gas pulse duration, the offset frequency of excitation and the length of applied excitation all can be adjusted to influence the extent of fragmentation.

1.4.3 Infrared Multi-Photon Dissociation

Photodissociation is possible by heating the ions trapped inside of the ICR cell using an infrared laser (usually $\lambda = 10.6 \ \mu\text{m}$).\textsuperscript{65} The fragmentation spectra are similar to those obtained through CID experiments. The major advantage of IRMPD over SORI-CID is that introduction of a gas-pulse into the ultra-high vacuum is not needed. It takes several seconds for the gas pulse to be pumped away, a necessary step before excitation and detection of the MS/MS spectra. If the pressure is too high as the ions are excited prior to detection, they will quickly lose ion coherence or fragment further, leading to a poor signal. The ions, however, need good overlap with the IR laser beam which is just a few mm in diameter; this requires that the ions be in a region of good overlap. This can be problematic after isolating the parent ion. IRMPD can cause secondary fragmentation as all the ions that overlap with the laser beam may absorb sufficient energy to fragment. Both the irradiation time and the laser power can be adjusted easily to change the extent of fragmentation.

Dissociation through vibrational activation, as in CID and IRMPD, produces predominantly b and y ions. To a lesser extent a and x ions can also occur, along
with loss of side chains. The loss of side chains is often undesired if fragmentation is being used to discover the site of a post-translational modification, such as phosphorylation.

1.4.4 Electron Capture Dissociation

ECD fragmentation of multiply-charged polypeptides is based on the dissociative recombination of a proton with an electron.\(^7^6\) ECD therefore requires at least a 2+ charge on the ion for fragment ions to be detected. The electron source is an indirectly heated dispenser cathode.\(^7^7\) The dispenser cathode provides a greater electron current and improved overlap with the ion cloud when compared with a heated filament. A hollow dispenser cathode has also been developed to allow for on-axis IR irradiation for vibrational excitation as well as a source of electrons for ECD.\(^7^8\) An off-axis approach to combining these two techniques has also been investigated.\(^7^9\) The trapping method, electron energy, the timing and duration of the irradiation time have been shown to be key parameters in the sensitivity and efficiency of ECD.\(^8^0,8^1\) The predominant fragment ions formed are c and z ions, while side chains are usually preserved. However, other fragment ions including b ions have been seen in ECD spectra.\(^8^2,8^3\)

The benefits of ECD are that cleavages occur over the entire backbone leading to far greater sequence coverage than conventional MS/MS methods which have been used to gain structural and kinetic information on proteins.\(^6^8,8^4–8^8\) The disadvantages are the low efficiency and sensitivity of the technique. To observe the low abundant fragments for the most sequence coverage as possible, a high abundance of the parent ion is required when compared to CID techniques. A large number of scans is also required in order to increase the signal-to-noise ratio of the low abundant fragments, which makes data acquisition much slower as well. In some cases electron irradiation produce only a few fragments. The number of fragments can be increased by post vibrational activation using a collision gas or IR laser.\(^2^0\) When ECD is combined in this way with a further fragmentation step
using IRMPD or CID, the complementary information can be used to identify the site of post-translational modifications.\(^{19,79,89}\)

### 1.5 Proteins

Proteins are made up of amino acid building blocks, many of which are essential nutrients.\(^{90}\) The sequence of amino acids that make up a particular protein is known as the primary structure; the way these amino acids orient in space forms the secondary (\(\alpha\)-helices and \(\beta\)-sheets) and tertiary structure (folded structure).

#### 1.5.1 Protein Structure

The 20 most common amino acid residues, their composition, mass and \(pK\) values for the carboxylic acid and amino groups, as well as any side chains are listed in Table 1.1.\(^{90}\) In general, \(> \text{pH 3.5}\), the carboxylic acids are almost entirely in their carboxylate forms, while \(< \text{pH 8.0}\) the amino groups are almost entirely in their ammonium ion forms. Amongst the amino acids possessing side chains, aspartic and glutamic acid are the most acidic, while lysine and arginine are the most basic. Amino acids polymerise through elimination of water, forming -CO-NH-peptide bonds. A sequence of amino acids, linked together by peptide bonds, can be thought of as a polymer chain with the side-chains branching off. This is called the protein backbone. Proteins range in length from \(~40\) to over \(4000\) amino acids resulting in molecular masses from \(~4\) to over \(440\) kDa.

Protein function is strongly related to structure, that is the three-dimensional relationships between the constituent atoms. Protein structure can be described in different ways. The primary structure describes the protein in terms of the amino acids that form it. The sequence starts at the N-terminus and lists each amino acid in order of how they are connected to each other along the protein backbone, until the C-terminus is reached. Secondary structure describes the local spatial arrangement of the peptide backbone without regard to the side chains. Pleated sheets, helices, and turns are common structural features in
### Table 1.1: The 20 commonly occurring amino acids residues with their important properties.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name and composition</th>
<th>Residue structure</th>
<th>Monoisotopic Average Mass&lt;sup&gt;a&lt;/sup&gt;</th>
<th>$pK_1$ α-COOH</th>
<th>$pK_2$ α-NH&lt;sup&gt;+&lt;/sup&gt;</th>
<th>$pK_R$ Side chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala A</td>
<td>Alanine C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;NO</td>
<td><img src="image" alt="Ala structure" /></td>
<td>71.03711 71.07806</td>
<td>2.35</td>
<td>9.87</td>
<td></td>
</tr>
<tr>
<td>Arg R</td>
<td>Arginine C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;12&lt;/sub&gt;N&lt;sub&gt;4&lt;/sub&gt;O</td>
<td><img src="image" alt="Arg structure" /></td>
<td>156.10111 156.18608</td>
<td>1.82</td>
<td>8.99 12.48</td>
<td></td>
</tr>
<tr>
<td>Asn N</td>
<td>Asparagine C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;6&lt;/sub&gt;N&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;2&lt;/sub&gt;</td>
<td><img src="image" alt="Asn structure" /></td>
<td>114.04293 114.10288</td>
<td>2.14</td>
<td>8.72</td>
<td></td>
</tr>
<tr>
<td>Asp D</td>
<td>Aspartic Acid C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;NO&lt;sub&gt;3&lt;/sub&gt;</td>
<td><img src="image" alt="Asp structure" /></td>
<td>115.02694 115.08761</td>
<td>1.99</td>
<td>9.90 3.90</td>
<td></td>
</tr>
<tr>
<td>Cys C</td>
<td>Cysteine C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;NOS</td>
<td><img src="image" alt="Cys structure" /></td>
<td>103.00918 103.14414</td>
<td>1.92</td>
<td>10.70 8.37</td>
<td></td>
</tr>
<tr>
<td>Gln Q</td>
<td>Glutamine C&lt;sub&gt;5&lt;/sub&gt;H&lt;sub&gt;12&lt;/sub&gt;N&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;5&lt;/sub&gt;</td>
<td><img src="image" alt="Gln structure" /></td>
<td>128.05558 128.12950</td>
<td>2.17</td>
<td>9.13</td>
<td></td>
</tr>
<tr>
<td>Glu E</td>
<td>Glutamic Acid C&lt;sub&gt;5&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;NO&lt;sub&gt;3&lt;/sub&gt;</td>
<td><img src="image" alt="Glu structure" /></td>
<td>97.05276 97.11541</td>
<td>2.10</td>
<td>9.47 4.07</td>
<td></td>
</tr>
<tr>
<td>Gly G</td>
<td>Glycine C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;NO</td>
<td><img src="image" alt="Gly structure" /></td>
<td>57.02146 57.05144</td>
<td>2.35</td>
<td>9.78</td>
<td></td>
</tr>
<tr>
<td>His H</td>
<td>Histidine C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;N&lt;sub&gt;4&lt;/sub&gt;O</td>
<td><img src="image" alt="His structure" /></td>
<td>137.50891 137.13963</td>
<td>1.80</td>
<td>9.33 6.04</td>
<td></td>
</tr>
<tr>
<td>Ile I</td>
<td>Isoleucine C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;11&lt;/sub&gt;NO</td>
<td><img src="image" alt="Ile structure" /></td>
<td>113.08406 113.15791</td>
<td>2.32</td>
<td>9.76</td>
<td></td>
</tr>
<tr>
<td>Leu L</td>
<td>Leucine C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;11&lt;/sub&gt;NO</td>
<td><img src="image" alt="Leu structure" /></td>
<td>113.08406 113.1579</td>
<td>2.33</td>
<td>9.74</td>
<td></td>
</tr>
<tr>
<td>Lys K</td>
<td>Lysine C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;12&lt;/sub&gt;N&lt;sub&gt;2&lt;/sub&gt;O</td>
<td><img src="image" alt="Lys structure" /></td>
<td>128.09496 128.17260</td>
<td>2.16</td>
<td>9.06 10.54</td>
<td></td>
</tr>
<tr>
<td>Met M</td>
<td>Methionine C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;12&lt;/sub&gt;NOS</td>
<td><img src="image" alt="Met structure" /></td>
<td>131.04048 131.19738</td>
<td>2.13</td>
<td>9.28</td>
<td></td>
</tr>
<tr>
<td>Phe F</td>
<td>Phenylalanine C&lt;sub&gt;9&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;NO</td>
<td><img src="image" alt="Phe structure" /></td>
<td>147.06841 147.17424</td>
<td>2.20</td>
<td>9.31</td>
<td></td>
</tr>
<tr>
<td>Pro P</td>
<td>Proline C&lt;sub&gt;5&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;NO</td>
<td><img src="image" alt="Pro structure" /></td>
<td>97.05276 97.11541</td>
<td>1.95</td>
<td>10.64</td>
<td></td>
</tr>
<tr>
<td>Ser S</td>
<td>Serine C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;NO&lt;sub&gt;2&lt;/sub&gt;</td>
<td><img src="image" alt="Ser structure" /></td>
<td>87.03203 87.07746</td>
<td>2.19</td>
<td>9.21</td>
<td></td>
</tr>
<tr>
<td>Thr T</td>
<td>Threonine C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;NO&lt;sub&gt;2&lt;/sub&gt;</td>
<td><img src="image" alt="Thr structure" /></td>
<td>101.04768 101.10408</td>
<td>2.09</td>
<td>9.10</td>
<td></td>
</tr>
<tr>
<td>Trp W</td>
<td>Tryptophan C&lt;sub&gt;11&lt;/sub&gt;H&lt;sub&gt;12&lt;/sub&gt;N&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;2&lt;/sub&gt;</td>
<td><img src="image" alt="Trp structure" /></td>
<td>186.07931 186.21039</td>
<td>2.46</td>
<td>9.41</td>
<td></td>
</tr>
<tr>
<td>Tyr Y</td>
<td>Tyrosine C&lt;sub&gt;9&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;NO&lt;sub&gt;2&lt;/sub&gt;</td>
<td><img src="image" alt="Tyr structure" /></td>
<td>163.06333 163.17364</td>
<td>2.20</td>
<td>9.21 10.46</td>
<td></td>
</tr>
<tr>
<td>Val V</td>
<td>Valine C&lt;sub&gt;5&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;NO</td>
<td><img src="image" alt="Val structure" /></td>
<td>99.06841 99.13129</td>
<td>2.29</td>
<td>9.74</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Neutral mass
proteins. Figure 1.6a shows a representation of an α-helix and Figure 1.6b shows a β-sheet. The atoms of the backbone and side-chain are in stick representation showing their positions in space. As proteins are made from many hundreds or even thousands of amino acids, showing where each atom is orientated in space can soon become cluttered, making unique features of a particular protein difficult to see. The side-chain atoms can be hidden allowing the path of the backbone to be seen more clearly. Features like α-helix and β-sheets can be highlighted using a cartoon representation, shown as being slightly transparent. Figure 1.6 illustrates this representation. The pleated sheets and helices are held together by hydrogen-bonding between carbonyl oxygen and the N-H in the amide from the backbone. This helps to hold these structures together and stabilise protein folding. Tertiary structure describes the full three-dimensional structure of the protein, including the side chains, showing how the secondary structure features lie in relation to each other. These can be held together with some or all of the following; hydrogen bonds, salt and disulfide bridges. Often proteins are composed of subunits of polypeptides which can be held together by non-covalent interactions, or via disulfide bridges (from cysteine side-chains) covalently linking them together. The spatial arrangement of these subunits give rise to the protein's
Hemoglobin is a tetramer of proteins comprised of a dimer of two $\alpha\beta$-dimers. Each hemoglobin subunit contains a ferric heme group and is almost entirely $\alpha$-helical in structure.

The quaternary structure is very important for the function of hemoglobin. Each heme group can bind its target ligand, $O_2$, enabling hemoglobin to transport 4 dioxygen molecules around the body. The binding of oxygen to the four heme groups does not happen simultaneously.
As the number of oxygen molecules bound increases, the binding affinity also increases, this is known as cooperative binding. This cooperative binding has been attributed to changes in the structure of the protein. When oxygen binds to the iron in the heme, part of the protein moves causing a conformational change at another part of the protein changing the binding affinity. Salt bridges are dynamically formed and broken as this happens and 10 microstates are known for hemoglobin. This highlights that protein structure can be highly dynamic and important for understanding protein function.

1.5.2 Protein Conformation Studies by MS

Circular dichroism is sensitive to secondary structure, allowing large changes in conformation to be observed. Techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) are used to elucidate the spacial structure of a protein, which is often a difficult task. X-ray crystallography offers the highest spatial resolution providing visualisation of the tertiary and any quaternary structure, but this is a static picture which freezes conformations and interactions. The X-ray crystal structure for the protein myoglobin, which is similar in size and folding to a monomer subunit of hemoglobin, is shown in Figure 1.8a. The heme group is bound non-covalently to the protein via the axial ligands of His64 and His93 residues to the iron atom. When oxygen binds to the iron of myoglobin the proximal axial ligand of His64 helps to stabilise the binding (Figure 1.8b). Using the space-filling representation (Figure 1.8c) it becomes clear that the oxygen becomes buried just under the surface of the protein. In order for the oxygen to bind and be released from the protein there must be some flexibility in the structure that is not shown by the static snapshots of X-ray crystallography. While protein interactions and dynamics can be followed by NMR or fluorescence, these often require a special label to be added. The nature of mass spectrometry makes it is sensitive to even small changes in mass, with or without an external label. As proteins require interactions with other molecules in order
Figure 1.8: a) The X-ray crystal structure of holo-myoglobin (horse) reproduced from PDB database entry 1YMB. The His64 and His93 residues provide axial ligands to the iron of the heme group. The X-ray crystal structure (PDB entry 1MBO) of oxygen bound to the iron of the heme group of holo-myoglobin, shown using: b) cartoon representation and c) space-filling representation, illustrating that the oxygen gets buried inside the protein after binding.
to function, the dynamic properties of proteins are of interest in the field of biology. Using mass spectrometry to observe changes in charge-state distribution or hydrogen/deuterium exchange (HDX) of the backbone amides, it is possible to obtain information on protein folding, dynamics and interactions. A detailed review of the use of mass spectrometry for the study of protein conformations is in the literature.95

1.5.2.1 Charge-State Distributions

During the electrospray ionisation process, a molecule can become charged at multiple sites. Proteins contain many amino acids with side-chains that can be either protonated or deprotonated leading to a net overall charge at a given pH. While it may seem intuitive that the charge measured by ESI-MS be related to the $pK_a$ of the side-chains, in general ESI-MS of proteins leads to a variety of charged species, giving rise to a 'charge-state distribution'. The charge-state distribution has been found to be only weakly dependent on the pH unless there is a change in the protein tertiary structure.96-98 Detailed analysis of the charge-state distribution leads to the observation of distinct conformations being present in solution.47,99 A native conformation can be thought of as the natural fold of a protein, which allows the protein to perform a specific function. In terms of energy, this conformation leads to a global minimum. If there are several conformations all with a small barrier for interconversion then the structure will be dynamic and not the static minimum shown by crystallography. When a protein is denatured, non-covalent interactions are disrupted causing dissociation of these weak bonds. The protein unfolds and its native fold is lost. An extended conformation that retains some of the structure of the native fold is an intermediate state between the native and the unfolded protein. A native conformation gives rise to a narrow charge-state distribution. A tightly folded structure will protect the core of the protein from the solvent, leading to only a few charged sites at the surface. An unfolded protein will expose the core residues to the
Increasing charge

Figure 1.9: Charge-state distributions observed in an ESI mass spectrum, depending on the solution conformations present.

The observed charge-state distribution can be a combination of all these, depending on the solution conditions. A diagram summarising the typical charge-state distributions observed depending on the conformations present, is shown in Figure 1.9.

1.5.2.2 Hydrogen/Deuterium Exchange

In the early 1950s, Linderstrøm-Lang and co-workers pioneered methods of studying proteins using hydrogen/deuterium exchange (HDX). The amide, -NH-,
hydrogen of peptide bonds are in continual exchange with the solvent. Residues that are exposed to the solvent will exchange rapidly, while amides that are involved in hydrogen-bonding (such as in α-helix and β-sheets), have low solvent accessibility (such as in the core of the protein) or steric hindrance, all exchange far more slowly.\textsuperscript{102} The dynamic properties of a protein can be monitored by dissolving in D\textsubscript{2}O and monitoring the HDX over time, either by NMR\textsuperscript{102} or MS.\textsuperscript{103} NMR allows exchange rates to be calculated for individual residues, but requires a high concentration of protein. MS has the advantages of high sensitivity and the ability to distinguish between populations of protein with differing exchange characteristics, due to the shifts in mass. As the exchange rate is highly sensitive to the solvent accessibility, HDX has been used to study protein structure and dynamics,\textsuperscript{103} conformation,\textsuperscript{104} and interactions.\textsuperscript{105-111} The exchange rate will alter if there is a change in the conformation or upon binding of ligands which will protect sites from solvent exposure. Since MS is a gas-phase technique, gas-phase HDX has also been performed.\textsuperscript{112-114}

1.5.3 Metalloproteins

A variety of metals are found in biology, and often play a vital role in many biological functions.\textsuperscript{115} It has been postulated that about 30\% of enzymes contain a metal bound at the active site. These metalloenzymes facilitate a variety of reactions, which include acid-catalysed hydrolysis (carried out by the hydrolases), redox reactions (carried out by the oxidases and oxygenases), and the rearrangement of carbon-carbon bonds (by synthases and isomerases). The co-ordination to the metal ions contained in these metalloenzymes are usually highly tailored for a specific redox reaction while the folds around the active site provide high selectivity for their given function. Some proteins contain metal ions for structural reasons helping the protein obtain the correct fold.

Metal-containing molecules serve important functions as electron carriers, metal storage sites, O\textsubscript{2} transport and storage, and signal transduction. In plants
and some bacteria, the metal magnesium plays a vital role in chlorophyll, which is used to harvest the power of the sun to produce glucose from carbon dioxide and water during photosynthesis. As mentioned previously (Section 1.5.1), hemoglobin is a well known metalloprotein found in red blood cells and is responsible for transportation of oxygen around the body. Since a variety of metals are important in many vital functions, specific proteins are also used to scavenge for and transport their particular metal targets.

Mass spectrometry, in particular coupled with electrospray ionisation, has been used to study many metalloproteins.\textsuperscript{116–122} Accurate determination of the mass allows oxidation states of the metal to be deduced.\textsuperscript{123,124} The transfer of metal ions can also be followed between proteins in reactions. Mass spectrometry is also a useful tool for studying the stoichiometry and affinity of binding, ligand interaction and inhibition, DNA and drug binding to proteins.\textsuperscript{125–130}

The experimental challenges come from trying to maintain the solution characteristics of a protein into the gas-phase during the ionisation process. Proteins need to be kept in buffered solutions to maintain a neutral pH. This keeps the protein in its native-fold and keeps the metal in the binding domain of metalloproteins. Inorganic salts and non-volatile buffers present a problem in mass spectrometry of proteins as adducts form, which broaden the peaks and suppress ionisation.\textsuperscript{130,131} Volatile buffers such as ammonium acetate, ammonium formate or ammonium hydrogen carbonate with a concentration of less than 50 mM can be used to study proteins under near physiological conditions. Use of aqueous buffered solutions near pH 7.0, low flow rates, the use of low temperatures, minimising collisions in the source by using low voltages and gas pressures, and optimising the ionic strength have all been key factors for preservation of non-covalent complexes, including metalloproteins.\textsuperscript{118} Stable isotope exchange has also been shown to be a powerful technique in the study the dynamic properties of metalloproteins.\textsuperscript{132–134}
1.5.3.1 Isotope Distribution and Isotope Modelling

FT-ICR MS offers high resolving-power and mass-accuracy allowing resolution of the isotope peaks present. The measured mass needs to be defined, as each peak in the mass spectrum will be a distribution of isotope peaks rather than a single peak. Peaks measured from a low resolution instrument, correspond to the centroid of the isotope distribution of the molecular species allowing determination of the average molecular weight. Resolution of the isotopic peaks, however presents a problem, of how to report the measured mass. Senko et al. describe a method by which to determine the monoisotopic peak, which is a far more accurate way of reporting the mass. The monoisotopic mass represents the mass corresponding to all the lightest elemental isotope masses. For most proteins this represents contributions from $^{12}\text{C}$, $^{1}\text{H}$, $^{14}\text{N}$, $^{16}\text{O}$, and $^{32}\text{S}$. For large biomolecules the monoisotopic peak has such a low abundance that it will be rarely observed above the level of the noise in the spectra. Senko et al. used a $\chi^2$ statistical test to find a match between the abundance of theoretical and measured isotope distributions after the profiles had been normalised. Calculation of the molecular mass (and any calculation of the isotope distribution for that matter), are based on the elemental composition of the neutral protein molecule. The formation of any oxidised thiols, which form disulfide bridges will result in loss of 2 Da per bond ($2\times\text{H}$).

A metalloprotein presents a problem, as the metal can also carry charge. The oxidation state of the metal also presents a problem as the number of protons will need to be varied. Johnson et al. assumed that the oxidation state at the metal centre is zero, with the resulting mass shift from the calculated monoisotopic peak indicating the true oxidation state. In this work, the isotope distributions presented have been corrected to yield the best match and any adjustments needed are noted in the discussion.

The charge on the ion can easily be determined when the isotope peaks are resolved. The charge is equal to the inverse of the mass difference between two
isotope peaks as shown in Equation 1.20.

\[ z = \frac{1}{\Delta M_{\text{isotope}}} \]  

(1.20)

However the observed charge does not provide any information on where the charge is located or the species of the charge carrier, which must be inferred from the change in mass from the neutral protein.

The deconvoluted neutral mass, \( M_n \), is calculated using Equation 1.21, where \( m/z \) is the measured mass-to-charge ratio, \( z \) is the charge present on the ion and \( M_{H^+} \) is the mass of a proton.

\[ M_n = (m/z)z - zM_{H^+} \]  

(1.21)

If the mass-to-charge ratio is from the most abundant isotopic peak, this will give a deconvoluted mass similar to the neutral average mass of the protein. Metalloproteins again pose a problem to this convention, as this assumes that all the charge on the ion is provided by protons. This works well for the \textit{apo} protein, but for the \textit{holo} protein, Equation 1.21 will overestimate the number of protons and thus the deconvoluted mass will be lower than the calculated mass by the number of protons equal to the oxidation state of the metal. Where applicable in this Thesis the number and species of charge carriers have been calculated and written explicitly in the assignment of a particular peak. As an example, ferric-cytochrome \( c \), of monoisotopic mass 12352.1 Da, with charge 10+, is written as \([M + Fe^{3+} + 7H^+]^{10+}\).

When the protein sequence, along with any modifications present, is known, it is possible to calculate the theoretical isotope pattern and match it with the observed isotopic peaks. Figure 1.10 shows the observed isotope pattern for the 10+ charge-state of ferric-cytochrome \( c \) along with the calculated isotope patterns for \( Fe^{3+} + 7H^+ (\text{△}) \) and \( Fe^{2+} + 8H^+ (\text{●}) \). Further details on how this spectrum was obtained can be found in Section 2.5.3.3. The theoretical isotope pattern was calculated using the formula \( C_{566}H_{883}Fe_{1}N_{148}O_{156}S_{4} \). The empirical
Figure 1.10: The observed isotope pattern for 10+ charged ion of ferric-cytochrome c, along with the theoretical isotope pattern shown with Fe$^{3+} + 7$H$^+$ (△) and Fe$^{2+} + 8$H$^+$ (●). The theoretical isotope pattern that matches best was calculated using the empirical formula C$_{560}$H$_{883}$Fe$_1$N$_{148}$O$_{156}$S$_4$, which indicates that the charge on the iron is 3$^+$. The heme group, which is covalently bound by two cysteine thiols, each lose a hydrogen upon binding. The neutral elemental formula for the heme group in cytochrome c is C$_{34}$H$_{36}$O$_4$N$_4$Fe, assuming a 2$^+$ charge on the iron. This makes the empirical formula for the neutral protein C$_{526}$H$_{842}$N$_{144}$O$_{152}$S$_4$. The difference between the matched formula used to calculate the theoretical isotope pattern and the neutral protein is 7 hydrogen. The overall charge of 10+ ion must be provided by the charge on the iron atom along with 7 protons, meaning the iron must be in the ferric oxidation state, Fe$^{3+}$. This methodology has been used throughout this Thesis to identify and characterise the species present in each mass spectrum.
1.6 Aims

The overall aim of this Thesis was to investigate metalloproteins in the gas phase using mass spectrometry. FT-ICR MS offers the highest mass resolving power and mass accuracy available at present. Much of the work presented in the literature to date using FT-ICR MS of biomolecules, makes use of denaturing conditions, which causes unfolding and loss of any of the metals bound. The specific aims during this work were:

1. To use native-like conditions in order to preserve the native conformation and resolve the isotope patterns of metalloproteins.

2. Make use of the high-mass accuracy and mass-resolving power in order to match the measured isotope pattern to the theoretical isotope pattern.

3. To preserve and investigate non-covalent complexes using ESI-FT-ICR MS.

4. To make use of enriched stable-isotope substitution to exploit the high-resolving power and mass-accuracy of FT-ICR MS

5. To use the top-down fragmentation techniques available in FT-ICR MS, in order to locate the site of binding of a metallo-anticancer agent to a protein.
Chapter 2
Experimental: Materials and Instrumentation

This Chapter discusses the materials, techniques required for sample preparation, the instruments employed and the samples which have been used for tuning and calibration during this work. The materials section covers the solvents and samples which were purchased. The techniques section describe gel filtration and ultrafiltration needed to desalt and buffer exchange samples in order to prepare them as suitable for mass spectrometric analysis using ESI. The mass spectrometers employed are described in detail in the instrumental section along with discussion of various practical aspects of using the instruments. In the concluding section of this Chapter FT-ICR mass spectra are presented for several proteins of varying molecular weight, which were used in order to tune and calibrate the instruments.

2.1 Materials

The materials used during this study are listed in Table 2.1. Included in this list are samples which were purchased for study or calibration as well as consumables required for sample preparation, such as solvents. The name of the compound, the quoted quality and the supplier information is also provided if applicable.
Table 2.1: Materials used, together with the supplier information.

<table>
<thead>
<tr>
<th>Name</th>
<th>Quality</th>
<th>Supplier</th>
<th>Catalogue No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance P acetate salt hydrate</td>
<td>≥ 95 % (HPLC)</td>
<td>Sigma</td>
<td>S6883</td>
</tr>
<tr>
<td>Bradykinin acetate</td>
<td>≥ 98 % (HPLC)</td>
<td>Sigma</td>
<td>B3259</td>
</tr>
<tr>
<td>Angiotensin I acetate hydrate</td>
<td>≥ 90 % (HPLC)</td>
<td>Sigma</td>
<td>A9650</td>
</tr>
<tr>
<td>Angiotensin II human acetate</td>
<td>≥ 95 % (HPLC)</td>
<td>Sigma</td>
<td>A9525</td>
</tr>
<tr>
<td>Angiotensin III</td>
<td>≥ 98.0 %</td>
<td>Sigma</td>
<td>Bio-10385</td>
</tr>
<tr>
<td>Neurotensin</td>
<td>~ 95 % (HPLC)</td>
<td>Sigma</td>
<td>N6383</td>
</tr>
<tr>
<td>Bombesin acetate hydrate</td>
<td>≥ 97.0 %</td>
<td>Fluka</td>
<td>15560</td>
</tr>
<tr>
<td>Luteinizing hormone-releasing hormone (LHRH)</td>
<td>≥ 98 % (HPLC)</td>
<td>Sigma</td>
<td>L7134</td>
</tr>
<tr>
<td>Ubiquitin from bovine erythrocytes</td>
<td>≥ 90 % (SDS-PAGE)</td>
<td>Sigma</td>
<td>U6253</td>
</tr>
<tr>
<td>Trypsinogen from bovine pancreas</td>
<td></td>
<td>Sigma</td>
<td>T1143</td>
</tr>
<tr>
<td>Carbonic anhydrase isozyme II from bovine erythrocytes</td>
<td>N/A</td>
<td>Sigma</td>
<td>C2522</td>
</tr>
<tr>
<td>Superoxide dismutase from bovine erythrocytes (beSOD)</td>
<td>N/A</td>
<td>Fluka</td>
<td>Bio-86200</td>
</tr>
<tr>
<td>Superoxide dismutase from human erythrocytes (heSOD)</td>
<td>N/A</td>
<td>Sigma</td>
<td>S9636</td>
</tr>
<tr>
<td>ESI Tuning Mix</td>
<td>N/A</td>
<td>Agilent Technologies</td>
<td>G2431A</td>
</tr>
<tr>
<td>Caesium iodide</td>
<td>99.999 %</td>
<td>Aldrich</td>
<td>203033</td>
</tr>
<tr>
<td>Perfluorohexanoic acid</td>
<td>99 %</td>
<td>Aldrich</td>
<td>342041</td>
</tr>
<tr>
<td>Acetazolamide</td>
<td>≥ 99 %</td>
<td>Sigma-Aldrich</td>
<td>A6011</td>
</tr>
<tr>
<td>Iron $^{57}$Fe Oxide</td>
<td>95.3 %</td>
<td>Advanced Material Technologies</td>
<td>Fluka</td>
</tr>
<tr>
<td>Acetic acid</td>
<td>puriss. p.a., eluent additive for LC-MS</td>
<td>Riedel-de Haën</td>
<td>34965</td>
</tr>
<tr>
<td>2-Propanol CHROMASOLV®</td>
<td>≥ 99.9 %</td>
<td>Fisher Scientific</td>
<td>A/0638/17</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>LC-MS</td>
<td>Fisher Scientific</td>
<td>M/4062/17</td>
</tr>
<tr>
<td>Methanol</td>
<td>LC-MS</td>
<td>Fisher Scientific</td>
<td>W/0112/15</td>
</tr>
<tr>
<td>Water</td>
<td>LC-MS</td>
<td>Fisher Scientific</td>
<td>A/3446/50</td>
</tr>
<tr>
<td>Ammonium acetate</td>
<td>99 + % HPLC</td>
<td>Fisher Scientific</td>
<td>N/A</td>
</tr>
<tr>
<td>Microcon YM-3 Centrifugal Filter Unit</td>
<td>N/A</td>
<td>Millipore</td>
<td>42403</td>
</tr>
<tr>
<td>Microcon YM-10 Centrifugal Filter Unit</td>
<td>N/A</td>
<td>Millipore</td>
<td>42406</td>
</tr>
</tbody>
</table>
2.2 Techniques

2.2.1 Gel Filtration Chromatography

Gel filtration chromatography separates molecules based on their size using an aqueous mobile phase. The column is packed with a porous inert material (e.g., silica) forming a relatively soft gel which requires relatively low pressures so that the packing material does not get distorted, so damaging the structure of the stationary phase. As molecules pass through the column, larger molecules cannot enter small pores while small molecules can. This results in larger molecules taking a shorter path through the column. They therefore elute before smaller molecules, which are hindered by the pores and take a much longer path. The performance of the packing material is dependent on the range of pore sizes contained within stationary phase and so this affects the range of molecular sizes that can be effectively separated by that column. The technique is most effective in separating large molecules from much smaller ones. Only an isocratic solvent system is needed to elute the sample and is largely unaffected by the buffers used. This allows desalting and buffer exchange to MS-compatible buffers such as ammonium acetate. Neutral pH can also be used so that protein-ligand or protein-drug complexes can be separated from any excess, unbound ligand or drug, which might interfere with analysis. This is important in order to prevent suppression of ionisation or formation of non-specific binding during such studies. The buffer exchange is quick and 'soft'; that is the complex is maintained when eluting from the column. The major disadvantage of this technique is that the sample gets significantly diluted, requiring either injection of a highly concentrated sample or concentration of the collected fraction for effective MS detection. For this reason on-line coupling is not effective.
2.2.1.1 Gel Filtration HPLC

An Agilent 1100 series HPLC system was used for gel filtration chromatography. An aliquot of the sample (20 μL) was injected onto a Tosoh Bioscience TSKgel SuperSW 2000 gel filtration column (4.6 × 30 cm) and eluted with ammonium acetate buffer (20 mM, pH 7). The flow rate was set to 0.35 mL min⁻¹ and the UV detector monitored the absorbance at 214 nm for proteins unless otherwise stated.

2.2.1.2 PD10 Column

To remove non-volatile buffers like Tris-HCl and the high salt concentration from solution, an Amersham-Biosciences Sephadex™G-25M PD10 gel filtration column was employed. A volume of 25 ml of 10 mM ammonium acetate at pH 9 was used to equilibrate the column. The protein (2.5 ml) was then loaded onto the column and the eluent discarded. Buffer solution (3.5 ml) was then loaded and the eluent collected. Using this method, typically 95 % of the protein was recovered, with a dilution factor of 1.4, while less than 4 % of the original salt concentration remained.

2.2.2 Ultrafiltration Centrifugation

Ultrafiltration makes use of a semipermeable membrane to separate components between a certain molecular weight cut-off (MWCO). Solvent and small molecular weight species below the cut-off pass through the membrane while high molecular weight components are retained. This is a pressure-driven process and the devices used were designed for use in a centrifuge. The solvent passes through the membrane along with any buffer or inorganic salts that may be in the sample. The high molecular weight species are concentrated by the reduction of solvent, while the salt and buffer concentrations remain the same. Desalting and buffer exchange (called diafiltration) is possible simply by diluting the concentrated sample using water or a prepared buffer such as ammonium acetate. The salt
and original buffer content are diluted relative to the original sample. By repeating the concentration and dilution steps 2 or 3 times, it is possible to remove 99.9% of the original salt, while retaining > 90% of the sample. The smaller the molecular weight cut-off the longer the process takes for the solvent to filter through. This is particularly a problem when using 3 kDa MWCO filters as the concentration step can take over an hour, while for 10 kDa MWCO the process takes about 20–25 min. As most of the biomolecules analysed were in the range 12–34 kDa, the 10 kDa MWCO filters were used most often.

2.2.2.1 Ultrafiltration Using Centrifugal Units

The ultrafiltration filters can contain glycerol that can interfere with MS analysis. For this reason units were always washed twice with the exchange buffer (500 μL) before the sample was added. The sample was added and the volume made up to a total volume of 500 μL. Ultrafiltration was then performed at 13000 rpm using a Sanyo MSE Micro Centaur centrifuge until ~20–30 μL remained. The exchange buffer was then used to dilute to a total volume of 500 μL and this process was repeated 2 or 3 times. The Millipore Microcon sample reservoir was inverted into a new vial and spun at 2000 rpm for 2 min. The reservoir was removed, the vial was capped, and the sample (typically 20–30 μL in volume) was stored until required.

2.3 Instrumentation

The principal focus of this work was to take advantage of the high-resolution and high mass-accuracy of FT-ICR mass spectrometry to study several biologically important metalloproteins in order to obtain data of the highest quality. The main instrument employed was a 9.4 T Bruker Daltonics Apex III FT-ICR MS, described in detail in Section 2.3.1. However, a new 12 T instrument was purchased and installed towards the final year of study. This 12 T Bruker Daltonics Apex Qe FT-ICR MS features many advantages over the Apex III, which are
outlined in greater detail in Section 2.3.2. Unless otherwise stated, the spectra presented in this Thesis were obtained using the 9.4 T Apex III MS.

### 2.3.1 9.4 Tesla Bruker Daltonics FT-ICR MS

The Bruker Daltonics Apex III FT-ICR MS operates with a 9.4 Tesla Magnex Scientific 160 mm super-conducting magnet. Figure 2.1 is a photograph of the instrument showing the vacuum cart fitted with an Advion TriVersa™ NanoMate® (see Section 2.3.3) as the nanoESI source. The flight tube containing the cell is inside the bore of the magnet. A Bruker atmospheric pressure ionisation (API) controller, the API 1600, is fitted together with a Bruker Apollo™ ESI/APCI source. The original Apollo™ nebulizing sprayer was replaced with an Agilent Technologies electrospray nebulizer to improve stability of the spray. Nitrogen gas was used to nebulize the analyte solution as it reaches the end of the nebulizing needle. Typically, 21 psi of pressure was used for nebulizing the sample solution. The sample flow rate to the ESI nebulizer was set to 100 μL h⁻¹.

An in-house designed and fabricated directly heated brass capillary replaced the standard glass capillary to improve desolvation of the electrospray ionisation process. The modified source (shown in Figure 2.2) was heated in the range of 423 to 523 K which improved the sensitivity and stability of ionisation. Due to no electrical isolation along the metal capillary, a single voltage was applied to the whole length of the capillary unlike the glass provided by Bruker Daltonics. An electrical gradient is needed for ionisation of the sample, therefore a direct voltage is applied to the nebulizer instead of it being held at instrument ground when using the glass capillary. In positive-ion mode, 3000 to 4000 V is applied, while in negative-ion mode -2000 to -3000 V is applied to the nebulizing sprayer.

Differential pumping is required for ESI as the ions need to travel from atmospheric pressure to the ~2 × 10⁻¹¹ mBar in the analyser cell needed for detection of cyclotron motion. The pumping stages are separated into two regions, the

---

1By Dr. P. Barran and Dr. S. Shirran, who based the design on similar modifications by Prof. J. R. Eyler
Figure 2.1: The Bruker Daltonics 9.4 Tesla FT-ICR mass spectrometer.

Figure 2.2: The brass capillary fitted to the Apollo ESI source. The capillary entrance and the thermocouple used to control the temperature have been highlighted.
source and the ultra high vacuum (UHV). A beam valve can be shut to isolate the two vacuum regions, which is needed as the source needs to be vented in order to clean the capillary and skimmers. The pumping system comprises two rotary roughing pumps which provide backing stages to the UHV (70 L s\(^{-1}\) and 500 L s\(^{-1}\)) and source (250 L s\(^{-1}\) and 500 L s\(^{-1}\)) turbomolecular pumps.

The Apex III has a simple source design. After the capillary there is a set of skimmers and a hexapole for ion accumulation. A set of ion optics then guides the ions as they are transferred from the source region to the ICR cell. The simplicity of this design has proven to be tunable for a wide range of conditions from harsh desolvation, to remove undesirable buffer adducts, to gentle source conditions to detect soft non-covalent complexes. The major disadvantage is the lack of a mass-resolving quadrupole for isolation of a particular ion prior to the analyser cell which is a major benefit for fragmentation studies.

2.3.2 12 Tesla Bruker Daltonics FT-ICR MS

The Apex Qe FT-ICR MS, shown in Figure 2.3, is operated with a 12 T US/R 110 mm bore super-conducting magnet. This system has many advantages over the older 9.4 T Apex III system. Using the latest magnet advances, the 12 T US/R magnet has ultra-shielding technology to minimise the stray field outside of the magnet, along with cryorefrigeration. The cryorefrigeration comprises a ‘cold head’ and a compressor. This removes the need for a liquid nitrogen cryostat which reduces the size and weight, along with a much lower helium consumption. The power supplies and electronics in the vacuum cart have been updated, improving the stability of the ion optics, increasing the stability of the signal.

There are several advantages to the Apex Qe over the 9.4 T Apex III MS. The higher field of the 12 T magnet gives rise to increased resolution, higher charge-capacity of ions and greater mass-accuracy. The vacuum cart has a newly-designed source which has a quadrupole mass filter and an orthogonal ion source. The quadrupole is sandwiched between two hexapoles, which can be used for
Figure 2.3: The Bruker Daltonics 12 Tesla FT-ICR mass spectrometer.
accumulation or CID. The quadrupole mass filter allows for selective ion transmission, which enables isolation of a single species prior to trapping in the ICR cell. This makes it possible to fragment intact proteins much more effectively than the Apex III system which is not fitted with a mass-resolving quadrupole. Isolation of a particular ion on the Apex III must be carried out in the ICR cell using a frequency sweep to excite all ions that need to be discarded, while leaving a suitable gap in the sweep to prevent ejection of the isolated ions. This gap or 'safety belt' often needs to be wide as off-resonance excitation is also possible. This excitation moves the ions to a new radius in the cell which makes on-axis dissociation techniques such as ECD and IRMPD quite difficult or impossible due to poor overlap with the electrons or laser beam.

The ion source pressure is regulated and filled with argon gas. This can be adjusted depending on the right conditions for the collisional cooling needed to trap the ions for efficient ion accumulation. Increasing the acceleration voltage into the second hexapole causes CID prior to detection in the ICR cell. This is a powerful technique when combined with the quadrupole mass filter.

The use of ion funnels\(^\text{136}\) in the source makes it possible for an orthogonal spray design, which removes the possibility of neutral molecules entering the high vacuum regions. This is important as neutral molecules, especially solvent, increase the background pressure which leads to poor transmission of ions. An increase in pressure in the UHV region is especially bad as collisions with neutral molecules causes loss of cyclotron motion, dissociation of ions and collapse of ion coherence resulting in loss of resolution.

### 2.3.3 The Advion TriVersa\(^\text{TM}\) NanoMate\(^\text{®}\) Chip-Based NanoESI Robot

The Advion TriVersa\(^\text{TM}\) NanoMate\(^\text{®}\), pictured in Figure 2.4 is a chip-based robotic device for nano electrospray. It uses the combination of microtitre plates, conductive tips and an array of nanospray nozzles. A few microlitres of sam-
Experimental: Materials and Instrumentation

In Figure 2.4: The Advion TriVersa™ NanoMate®.

Pie are placed in the sample wells of the microtiter plate, a robotic arm picks up a new conductive sample tip and aspirates the sample into the tip and then connects the tip with the next unused spray nozzle. The electrospray voltage is applied to the tip which conducts to the solution and nozzle producing electrospray. When the spray is instructed to stop, the robot can return any remaining sample to the microtitre plate and the tip is ejected. This approach removes any cross-contamination between samples and makes it easy to load the sample. The NanoMate has been shown to be especially useful for studies of non-covalent interactions, such as protein-ligand interactions, where automated titrations can be performed due to the robust and reliable spray.\textsuperscript{125}

The TriVersa NanoMate enhances the capabilities for liquid chromatography. The robot can take the eluent flow from the LC system and split it so that the correct flow rates are sent to the nanospray chip for on-line MS detection. The remaining flow is either discarded like traditional flow splits, or can be directed to the robotic mandrel for collection into a microtitre sample plate. The collected
fractions can then be used as samples for off-line nanospray infusion as before. This is a powerful technique for collecting fractions that require longer analysis time, to improve mass resolution, or fragmentation of a protein.

2.3.4 Ion Trapping and Detection

An electric field is applied to two parallel plates perpendicular to the magnetic field to trap the ions axially in the analyser cell, in order to detect the cyclotron motion due to the high field of the super-conducting magnet. Typical trapping voltages of 0.8 to 1.5 V are used. A high applied trapping voltage results in increased magnetron motion which causes a shift in measured mass-to-charge ratio and peak broadening. The detection cell used in the Bruker Apex series instruments is the ‘Infinity’ cell. This is a cylindrical cell with segmented trapping plates which simulate the potential of an infinitely extended cylinder. Figure 2.5 shows a photograph of the ‘Infinity’ cell. The trapping plates are capacitively cou-
pled to the excitation plates so that ions experience the same trapping potential during the excitation stage prior to cyclotron detection.

The ‘Infinity’ cell has a patented ‘SIDEKICK’ trapping technique which applies a potential across two split cylindrical plates just after the cell entrance. These plates alter the $xy$-motion of the ions, and affect the ion cloud’s radial position so as to send the ions off-axis. This method of trapping does not require cooling gas to be pulsed into the cell, which would need to be pumped away prior to excitation and detection. A variety of ions can be accumulated in the cell, as the on-axis energy is converted to off-axis energy, allowing ions with a wide mass-to-charge range to be trapped at once. If the ion packet is off-axis, overlap with electrons or laser photons during ECD or IRMPD MS$^n$ experiments is more difficult but can be overcome by careful excitation.$^{138,139}$

Gated trapping confines the ions without applying the ‘SIDEKICK’ voltages and is achieved by careful timing when the voltage to the first trapping plate is applied, trapping the ions between the two plates. The disadvantage of this technique is that ions with only a narrow range of mass-to-charge can be trapped at once. If the trapping voltages remain constant, this is called static gated trapping. It is possible to change the trapping voltage during the experiment as well as pulsing in argon as a cooling gas. These techniques are known as dynamic gated trapping and gas-assisted dynamic trapping, respectively. A higher trapping voltage is used for high trapping efficiency, followed by a lower trapping voltage to reduce the magnetron motion and space-charge effects. By pulsing gas to cool the ions, multiple ion accumulations can be achieved, increasing the number of ions in the cell. The resolution may be increased compared to ‘SIDEKICK’ trapping, however the experiment time is also increased and it is difficult to achieve effective trapping.

The ‘Infinity’ cell uses dipolar excitation and detection, which are on separate independent plates. A T & C Power Conversion Inc. ULTRA series LF/HF linear power amplifier is used to amplify the signal from the RF-generator for excitation.
of the ions in the cell. A pre-amp, fitted to the back of the flight-tube, is used to amplify the tiny signal induced by the ions undergoing cyclotron motion.

2.3.5 Instrument Tuning

Although the two instruments are quite different, as outlined in Section 2.3.1 and 2.3.2, the main principles are the same. The newly designed Apollo II source of the 12 T FT-ICR MS appears to require less tuning when compared to the 9.4 T instrument, as a more stable signal is transferred to the cell. Precise control of the source energies is however more difficult, making detection of non-covalent complexes more challenging.

Tuning the instrument can be a lengthy process as there are a lot of parameters that change the many voltages throughout the instrument. Although this allows for the maximum flexibility for controlling the instrument, each sample requires some optimisation in order to obtain satisfactory spectra. The following description focuses principally on tuning aspects of the 9.4 T FT-ICR MS.

Tuning the source and the ion optics, which are outside of the magnet, is required to optimise the transmission of ions to the cell inside the magnet, and so has an effect on the sensitivity of the instrument. Once the sensitivity has been maximised, the cell parameters can then be tuned to maximise the resolution of the instrument.

Source tuning is important as it affects the energies that the ions experience in the highest pressure regions of the instrument. The most important voltages in the source region are the capillary and skimmer voltages. These allow for tuning 'soft' or 'harsh' source conditions. During normal operation, sufficiently high voltages are applied in the source in order to decluster any residual solvent molecules, to obtain a strong signal. Non-covalent complexes often require very soft conditions in order to maintain the integrity of the complex through to the ICR cell for detection. This requires tuning of the source voltages.

Using higher source voltages just after ionisation, the ions may gain sufficient
kinetic energy that collisions with gas molecules cause the ions to fragment. The
dissociation of covalent bonds requires much higher voltages than those needed
to dissociate non-covalent complexes. This dissociation technique is known as
nozzle-skimmer CID. Usually this in-source CID is undesirable, but can be useful
to extend the sequence coverage of larger intact proteins when used in conjunction
with other fragmentation techniques.\textsuperscript{140}

After transfer through the capillary and skimmers the ions are collected in
a storage hexapole. The accumulation time can be adjusted to compensate for
low or high abundance of ions from the electrospray. However, changing the con-
centration of the sample may still be required. Increasing the accumulation time
may increase the abundance of ions transferred to the cell, but the length of each
acquisition cycle increases, which is an important factor to consider when per-
forming on-line-LC analysis. Fast eluting peaks may be detected together or low
level signals may not be detected at all if there are insufficient ions accumulated.
Long accumulation times can also lead to charge reduction, resulting in a change
in the charge distribution in the mass spectrum, and fragmentation can also occur
for unstable ions.

After the ions are ejected from the hexapole, they must travel through the ion
optics to the cell before the trapping voltage is applied to the first trapping plate
closest to the source. This delay is a flight time for the ions and is related to the
mass-to-charge ($m/z$) and the kinetic energy distribution of the ions. Lower $m/z$
ions travel faster than high $m/z$ ions, and so careful timing is required to trap
the ions of the desired range.

The voltages applied to the cell storing plates play a key role in trapping the
ions and the resolution that can be achieved. High trapping voltages may increase
the trapping efficiency, which is needed for high resolution, as this is related to
the transient detection time. The magnetron motion of the ions is related to the
electric field and how far off-centre the ions are. The larger the mass of the ion
the stronger the effect of the magnetron motion. The magnetron motion causes
the observed mass to shift to high $m/z$ due to lower measured frequency.$^{51,52}$

A large abundance of ions in the cell is not always desirable as too many ions lead to space-charge effects, reducing the resolution, shifting the observed mass-to-charge of the peak.$^{141,142}$ A low abundance of ions leads to poor signal-to-noise ratio requiring a larger number of scans to be accumulated, but will reduce the space-charge effects. Therefore, in order to obtain a well resolved spectrum the number of stored ions is a critical parameter. The effect of ion-coalescence, which determines the decay of the transient, is important as more ions almost always give rise to better signal than increasing the number of scans. Along with the concentration of analyte in solution, the hexapole accumulation time can be used to control the ion abundance observed for a given charge-state in the mass spectrum. A mass-resolving quadrupole, like that of the Apex Qe instrument, can also be used to isolate a limited mass range so that number of ions can be reduced.$^{143}$

The signal-to-noise can be improved by increasing the number of transient scans which are averaged. The number of scans, $N$, increases the signal-to-noise in a statistical manner and is proportional to $\sqrt{N}$. It is therefore desirable to have a higher abundance of ions rather than increasing the number of scans, as this leads to a shorter acquisition time.

Resolution is also affected by the amplitude of the RF excitation frequency. Too little power and the ions are too far from the detection plates leading to a short transient. Too much power may cause ion loss if the radius exceeds the cell dimensions. A larger radius also increases the kinetic energy of the ions, so that collisions with background pressure neutral molecules may cause random fragmentation and loss of ions.

### 2.4 Software

The software used to acquire data on the 9.4 T Apex III FT-ICR MS system was Bruker Daltonics Xmass version 7.0.8. The instrument control software on
the 12 T Apex Qe FT-ICR MS system was Bruker Daltonics apexControl version 2.0.0.21. Acquired data from both systems were analysed using Bruker Daltonics DataAnalysis (DA) version 3.4 (build 172).

For analysing MS/MS fragments of intact proteins, a peak list of monoisotopic masses was generated using the 'SNAP’ tool of DA. The peak list was then exported, deconvoluted and corrected to be singly charged by the addition of one Dalton, [M + H+]\. The corrected peak list was then imported into the Bruker Daltonics BioTools suite version 3.0. The protein sequence was entered into BioTools and the theoretical fragments were calculated and compared to the measured masses.

Data plots were produced by exporting the calibrated measured spectra into X-Y ASCII files. These X-Y files were then plotted using Gnuplot 4.0 (http://www.gnuplot.info/). Gnuplot’s least squares fit parameter minimisation was used to optimise the fitting function, where applicable.

Protein X-ray crystal structures were obtained from the RSCB protein data bank (PDB) (http://www.rcsb.org/pdb/home/home.do). Downloaded structures were then manipulated and rendered using PyMOL (http://pymol.sf.net/).

2.5 Mass Spectra of Standards and Calibrants

For instrument tuning and calibration, samples that were readily available and cheap were used so that more interesting samples, of limited quantity, were not wasted without collecting useful data. Outlined in this section are the samples that were used for tuning the instrument. Well characterised samples were also used to calibrate the spectra, which was needed for accurate mass measurements. This was particularly necessary when isotopic modelling was used for analysis of a sample, so that the theoretical and measured isotopes matched with little ambiguity.
2.5.1 Agilent Technologies ESI Tuning Mix

The ESI tuning mix produced by Agilent Technologies, or simply known as ‘Tune Mix’ provides routine calibration over the range 500 to 3000 m/z. The calibration mix is comprised of a series of five homogeneously substituted triazatriphosphazene compounds along with betaine and hexamethoxyphosphazene added for low mass calibration. The calibration compounds in the ESI tuning mix. The calibrant can be used to calibrate a spectrum externally or can be added to a sample to calibrate the spectrum internally with little apparent problems of ion suppression, although the dilution needs to be adjusted to a suitable abundance.

In positive-ion mode, all these calibration compounds gain a proton to form singly charged ions. The low mass compounds of betaine (118 m/z) and hexamethoxyphosphazene (322 m/z) often give fairly weak peaks and so calibration masses are usually chosen from the compounds in the range 622 to 2722 m/z. The formulae and calculated monoisotopic masses of the calibration compounds are given in Table 2.3. The stock solution was diluted 10-fold in acetonitrile (ACN) with the addition of a little water (ACN : H₂O 97 : 3 v/v) and the mass

Table 2.2: The names and formulae of the calibration compounds of the Agilent ESI tuning mix.

<table>
<thead>
<tr>
<th>Compound name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betaine</td>
<td>C₅H₁₁NO₂</td>
</tr>
<tr>
<td>Hexamethoxyphosphazene</td>
<td>C₆H₁₈N₃O₆P₃</td>
</tr>
<tr>
<td>Hexakis(2,2-difluoroethoxy)</td>
<td>P₃N₃(-O-CH₂-CF₂H)₆</td>
</tr>
<tr>
<td>phosphazene</td>
<td></td>
</tr>
<tr>
<td>Hexakis(1H,1H,3H-tetrafluoropropoxy)</td>
<td>P₃N₃(-O-CH₂-(CF₂-CF₂H))₆</td>
</tr>
<tr>
<td>phosphazene</td>
<td></td>
</tr>
<tr>
<td>Hexakis(1H,1H,5H-octafluoropentoxy)</td>
<td>P₃N₃(-O-CH₂-(CF₂)₃-CF₂H)₆</td>
</tr>
<tr>
<td>phosphazene</td>
<td></td>
</tr>
<tr>
<td>Hexakis(1H,1H,7H-dodecafluorohexoxy)</td>
<td>P₃N₃(-O-CH₂-(CF₂)₅-CF₂H)₆</td>
</tr>
<tr>
<td>phosphazene</td>
<td></td>
</tr>
<tr>
<td>Hexakis(1H,1H,9H-perfluorononylxy)</td>
<td>P₃N₃(-O-CH₂-(CF₂)₇-CF₂H)₆</td>
</tr>
<tr>
<td>phosphazene</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.3: The calibration peaks observed for Agilent ESI tuning mix in positive-ion and negative-ion modes. The formula and calculated monoisotopic masses are shown along with observed masses with the mass error.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Charge</th>
<th>Theoretical mass /m/z</th>
<th>Measured Mass /m/z</th>
<th>Mass error /ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₅H₁₂O₂N</td>
<td>1+</td>
<td>118.086255</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₆H₁₉O₆N₃P₃</td>
<td>1+</td>
<td>322.048121</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₁₂H₁₉O₆N₃P₃F₁₂</td>
<td>1+</td>
<td>622.028960</td>
<td>622.028962</td>
<td>0.003</td>
</tr>
<tr>
<td>C₁₃H₁₉O₆N₃P₃F₂₄</td>
<td>1+</td>
<td>922.009798</td>
<td>922.009785</td>
<td>-0.015</td>
</tr>
<tr>
<td>C₁₉H₁₉O₆N₃P₃F₄₈</td>
<td>1+</td>
<td>1521.971475</td>
<td>1521.971525</td>
<td>0.033</td>
</tr>
<tr>
<td>C₄₂H₁₉O₆N₃P₃F₇₂</td>
<td>1+</td>
<td>2121.933152</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₅₄H₁₉O₆N₃P₃F₹₆</td>
<td>1+</td>
<td>2721.894829</td>
<td>2721.894756</td>
<td>-0.027</td>
</tr>
<tr>
<td>C₂O₂F₃</td>
<td>1—</td>
<td>112.985587</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₁₂HON₃F₂₁</td>
<td>1—</td>
<td>601.978977</td>
<td>601.978965</td>
<td>-0.019</td>
</tr>
<tr>
<td>C₁₄H₁₈N₃O₈P₃F₁₅</td>
<td>1—</td>
<td>734.007271</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₂₆H₁₈N₃O₈P₃F₂₇</td>
<td>1—</td>
<td>1033.988109</td>
<td>1033.988311</td>
<td>0.195</td>
</tr>
<tr>
<td>C₃₃H₁₈N₃O₈P₃F₵₁</td>
<td>1—</td>
<td>1633.949786</td>
<td>1633.949218</td>
<td>-0.348</td>
</tr>
<tr>
<td>C₄₄H₁₈N₃O₈P₃F₇₅</td>
<td>1—</td>
<td>2233.911463</td>
<td>2233.910484</td>
<td>-0.438</td>
</tr>
<tr>
<td>C₅₆H₁₈N₃O₈P₃F₹₉</td>
<td>1—</td>
<td>2833.873139</td>
<td>2833.875168</td>
<td>0.716</td>
</tr>
</tbody>
</table>

spectrum acquired from a single scan using 1 M (2²⁰) data set. The acquired ESI tuning mix spectrum is shown in Figure 2.6 after calibration. The observed monoisotopic mass and mass errors are also shown in Table 2.3. The mass errors are all much less than 1 ppm showing that a good calibration has been achieved.

In negative-ion mode, the substituted triazatriphosphazene compounds all ionise via the formation of an adduct with deprotonated trifluoroacetic acid (TFA—H⁺, C₂O₂F₃⁻), due to the addition of the TFA ammonium salt in the calibration mix. As with the positive-ion mode, all the ions are singly-charged. The formulae and calculated monoisotopic masses for the calibration peaks in negative mode are shown in Table 2.3. It was found experimentally that although it was possible to obtain an electrospray in negative ion mode with acetonitrile when using ESI, isopropanol (IPA) gave a far more reliable nanoESI spray when using the NanoMate in either positive- or negative-ion modes. The stock solution of ESI tuning mix was diluted 10-fold with isopropanol and 3 % water was added (ACN : IPA: H₂O 10 : 87 : 3 v/v). An acquisition of 10 scans was collected using
Figure 2.6: Mass spectrum of Agilent ESI tuning mix in positive-ion mode (ACN : H₂O 97 : 3 v/v). The observed monoisotopic masses are shown above the peaks after calibration. For peak assignments see Table 2.3.

Figure 2.7: Mass spectrum of Agilent ESI tuning mix in negative-ion mode (ACN : IPA : H₂O 10 : 87 : 3 v/v). The observed monoisotopic peaks are shown above the peaks. For peak assignments see Table 2.3.
a 512 k \( (2^{19}) \) data-set. The acquired mass spectrum of ESI tuning mix in negative mode is shown in Figure 2.7 after calibration. The observed monoisotopic masses are and the mass errors are shown in Table 2.3. The mass errors are all below 1 ppm showing that the calibration is very good. The peaks have almost all the same abundance across the mass range which is preferred for a good calibration.

2.5.2 Cesium Tridecafluoroheptanoate (CsTFHA) Clusters

It is difficult to calibrate at high \( m/z \) using peptides or proteins which carry multiple charges and the charge number increases with increasing size. A large protein in denaturing conditions (high organic solvent at low pH) will show a large number of charge-states, but typically the most abundant will be centred around 1000 \( m/z \). Use of neutral pH conditions significantly reduces the charge on the protein, and dramatically increases the measured \( m/z \) value. Typically only a few charge-states will be observed, with one being predominant. Another problem is that the number of isotope peaks increases as the mass of a protein increases. It can be difficult to be certain of the identity of a particular isotopic peak which is needed in order to accurately assign it to a calibration peak. Often the most abundant isotope peak is used for calibration, as long as the elemental formula can be determined from its amino acid sequence. This relies on the protein sequence being known and that if there any modifications, these are accurately determined. Although internal calibration is preferred, use of a cesium cluster series provides a calibration for high \( m/z \) which can then be used to calibrate native-esi spectrum of large proteins. Clusters of cesium tridecafluoroheptanoate (CsTFHA) can be used to calibrate a mass spectrometer from 600 to 10 000 \( m/z \) or beyond, in positive and negative ionisation modes.

A calibration mix was prepared using cesium iodide (100 \( \mu M \)) with tridecafluoroheptanoic acid (100 \( \mu M \)) in 1:1 (v/v) water:isopropanol. In positive-ion mode, the clusters take the form \( \text{Cs(C_7F_{13}O_2Cs)_n}^+ \) and the sodiated form
Table 2.4: Calibration table for a cesium tridecafluorohexadecanoate clusters in positive-ion and negative-ion modes. The formulae and calculated monoisotopic masses are shown, along with observed masses with the mass error.

<table>
<thead>
<tr>
<th>Species</th>
<th>Charge</th>
<th>Theoretical mass /m/z</th>
<th>Measured Mass /m/z</th>
<th>Mass error /ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na(C$<em>7$F$</em>{13}$O$_2$Cs)$_4$</td>
<td>1+</td>
<td>2006.487292</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cs(C$<em>7$F$</em>{13}$O$_2$Cs)$_4$</td>
<td>1+</td>
<td>2116.402970</td>
<td>2116.403002</td>
<td>0.015</td>
</tr>
<tr>
<td>Na(C$<em>7$F$</em>{13}$O$_2$Cs)$_5$</td>
<td>1+</td>
<td>2502.361810</td>
<td>2502.361575</td>
<td>-0.094</td>
</tr>
<tr>
<td>Cs(C$<em>7$F$</em>{13}$O$_2$Cs)$_5$</td>
<td>1+</td>
<td>2612.277488</td>
<td>2612.277584</td>
<td>0.037</td>
</tr>
<tr>
<td>Na(C$<em>7$F$</em>{13}$O$_2$Cs)$_6$</td>
<td>1+</td>
<td>2998.236328</td>
<td>2998.237072</td>
<td>0.248</td>
</tr>
<tr>
<td>Cs(C$<em>7$F$</em>{13}$O$_2$Cs)$_6$</td>
<td>1+</td>
<td>3108.152005</td>
<td>3108.151362</td>
<td>-0.207</td>
</tr>
<tr>
<td>Na(C$<em>7$F$</em>{13}$O$_2$Cs)$_7$</td>
<td>1+</td>
<td>3494.110846</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_1$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>858.843040</td>
<td>858.843014</td>
<td>-0.031</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_2$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>1354.717558</td>
<td>1354.717863</td>
<td>0.225</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_3$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>1850.592076</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_4$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>2346.466594</td>
<td>2346.464765</td>
<td>-0.779</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_5$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>2842.341111</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_6$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>3338.215629</td>
<td>3338.217876</td>
<td>0.673</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_7$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>3834.090417</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(C$<em>7$F$</em>{13}$O$_2$Cs)$_8$(C$<em>7$F$</em>{13}$O$_2$)</td>
<td>1−</td>
<td>4329.964665</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Na(C$_7$F$_{13}$O$_2$Cs)$_n^+$, where $n \geq 1$. Figure 2.8 shows an example of a CsTFHA spectrum used to calibrate the mass range from 2000 to 3000 m/z. The clusters formed have few isotopes and are all singly charged making calibration of masses easy even at high m/z. A quadratic calibration was performed using the monoisotopic masses noted in the Figure. A full list of species with the monoisotopic masses in the mass range of 2000 to 3500 m/z is given in Table 2.4, along with the measured masses and the associated mass error after calibration. The cluster ions are not the most abundant peaks in the spectrum and many peaks for other masses are present. The clusters appear to be quite 'fragile', as the species corresponding to the calibration masses appear to be fragmented in the source unless soft conditions are used. Under more 'harsh' conditions the abundances greatly reduce while unassigned peaks are predominant.

A corresponding negative-ion mode spectrum is shown in Figure 2.9 and the calibration species are listed in Table 2.4 for the mass range of 800 to 5000 m/z.
Figure 2.8: Mass spectrum of CsTFHA clusters in positive-ion mode (H₂O : IPA 1:1 v/v).

Figure 2.9: Mass spectrum of CsTFHA clusters in negative-ion mode (H₂O : IPA 1:1 v/v).
The masses correspond to clusters of \((\text{C}_7\text{F}_{13}\text{O}_2\text{Cs})_n(\text{C}_7\text{F}_{13}\text{O}_2)^-\), where \(n \geq 1\).

### 2.5.3 Peptides and Proteins

To tune the instruments samples of peptides and proteins were used in a range of masses from 756 Da to 24 kDa. The peptides and proteins used are detailed in Table 2.5, together with the calculated neutral mass, the amino acid sequence, the amino acid count and the elemental formulae used for isotopic modelling. Any features which are important for the mass measurement are also indicated. The N-terminus is indicated as follows: ‘h-’ implies that a hydrogen caps the terminus, ‘hp-E’ indicates that the first residue is pyroglutamic acid and ch\(_3\)co indicates an acetylated terminus. The C-terminus can either be the free acid, indicated by ‘-oh’ or an amide shown as ‘-nh\(_2\)’. Lower case letters have been used to distinguish the terminus from amino acids, this was important for writing a program that was able to read a sequence and output the elemental formula for that sequence. For details on this program see Appendix A.

#### 2.5.3.1 Peptide Mixture

A sample containing a mixture of peptides was made by first dissolving each crystalline stock in LC-MS grade water. Angiotensin I and II (5 mg each) were dissolved in 1 ml of water. Similarly, angiotensin III, neurotensin and bradykinin fragments 1-7 (1 mg each) were dissolved in 1 ml of water. The peptide mix was diluted a further ten times into a solution containing water, methanol and acetic acid (49:49:2 v/v), to obtain 0.005 mg / ml of each peptide. The final concentrations of each peptide were estimated to be:\(^1\) angiotensin I 3.86 \(\mu\)M, angiotensin II 4.78 \(\mu\)M, angiotensin III 5.37 \(\mu\)M, neurotensin 3.99 \(\mu\)M and bradykinin 1-7 6.61 \(\mu\)M. A 10 \(\mu\)l aliquot of the sample was loaded into a ProXeon Biosoftware’s metal-coated borosilicate nanospray emitter using a 20 \(\mu\)l Eppendorf GELloader\(^\text{®}\)

\(^1\)As the lyophilised powder may contain some water and impurities. The exact weight of the peptides and proteins have not been accurately determined unless otherwise stated using UV/VIS absorption for example.
Table 2.5: Peptides and proteins used as standards for calibration and instrument tuning.

<table>
<thead>
<tr>
<th>Name</th>
<th>Calculated average mass /Da</th>
<th>Sequence information</th>
<th>Total amino acids</th>
<th>Elemental formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heme (myoglobin)</td>
<td>616.49</td>
<td>-</td>
<td>-</td>
<td>C$<em>{34}$H$</em>{32}$O$<em>{4}$N$</em>{4}$Fe</td>
</tr>
<tr>
<td>Heme (cytochrome c)</td>
<td>684.65</td>
<td>-</td>
<td>-</td>
<td>C$<em>{34}$H$</em>{32}$O$<em>{4}$N$</em>{4}$S$_{2}$Fe</td>
</tr>
<tr>
<td>Angiotensin III</td>
<td>931.09</td>
<td>h-RVYIHPF-oh</td>
<td>7</td>
<td>C$<em>{64}$H$</em>{66}$N$<em>{12}$O$</em>{9}$</td>
</tr>
<tr>
<td>Angiotensin II</td>
<td>1046.18</td>
<td>h-DRVYIHDPF-oh</td>
<td>8</td>
<td>C$<em>{50}$H$</em>{71}$N$<em>{13}$O$</em>{12}$</td>
</tr>
<tr>
<td>Bradykinin</td>
<td>1060.21</td>
<td>h-RPPGFSPFP-oh</td>
<td>9</td>
<td>C$<em>{50}$H$</em>{73}$N$<em>{15}$O$</em>{11}$</td>
</tr>
<tr>
<td>Luteinizing hormone-</td>
<td>1182.29</td>
<td>hp-EHWSYGLRPG-nh$_{2}$</td>
<td>10</td>
<td>C$<em>{55}$H$</em>{77}$N$<em>{14}$O$</em>{13}$</td>
</tr>
<tr>
<td>releasing hormone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angiotensin I</td>
<td>1296.48</td>
<td>h-DRVYIHPFHL-oh</td>
<td>10</td>
<td>C$<em>{64}$H$</em>{69}$N$<em>{17}$O$</em>{14}$</td>
</tr>
<tr>
<td>Substance P</td>
<td>1347.63</td>
<td>h-RPKPQFFGLM-nh$_{2}$</td>
<td>10</td>
<td>C$<em>{63}$H$</em>{98}$N$<em>{18}$O$</em>{13}$S</td>
</tr>
<tr>
<td>Bombesin</td>
<td>1619.85</td>
<td>hp-EQRGNQAVGHLM-nh$_{2}$</td>
<td>14</td>
<td>C$<em>{71}$H$</em>{130}$N$<em>{24}$O$</em>{18}$S</td>
</tr>
<tr>
<td>Neurotensin</td>
<td>1671.93</td>
<td>hp-ELYENKPRRYPYI-oh</td>
<td>13</td>
<td>C$<em>{78}$H$</em>{121}$N$<em>{21}$O$</em>{20}$</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>8564.76</td>
<td>h-MQIFVTKLTVITLLEVEPSDTIENVKAKIQDKEGIPPDQQRFLPAGQLEDGR-TLSDYN1QKESTTLVRLRGG-oh</td>
<td>104</td>
<td>C$<em>{564}$H$</em>{878}$N$<em>{148}$O$</em>{156}$S$_{4}$Fe</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>12364.98</td>
<td>ch$_{3}$co-GDEVKGGKIFVQKCAQCHTVEKGGKGHTGPNLHGLFGRTQAPQFT-YTDANKNGITKWEETLMEYLELENPKYIPGTMIFAGIKKTERYELAYLKKT-ATNE-oh + heme $\text{c,d}$</td>
<td>153</td>
<td>C$<em>{769}$H$</em>{1212}$N$<em>{210}$O$</em>{218}$S$_{2}$</td>
</tr>
<tr>
<td>apo-Myoglobin</td>
<td>16951.30</td>
<td>h-GLSDGEWQVLNVGKVVEADIGHGEQVLRILFTGHPETLEKFKPHLKT-TEAEMKASEDLKKHGTVLTLGGLKKGHHAEKPLQASHATKHKPIKY-LEFISDAIIHVLHISHKHPGDFAGADAQGAMTKEALFRNDIAAKYKELGFQG-oh</td>
<td>153</td>
<td>C$<em>{803}$H$</em>{1244}$FeN$<em>{214}$O$</em>{222}$S$_{2}$</td>
</tr>
<tr>
<td>holo-Myoglobin</td>
<td>17559.72</td>
<td>h-GLSDGEWQVLNVGKVVEADIGHGEQVLRILFTGHPETLEKFKPHLKT-TEAEMKASEDLKKHGTVLTLGGLKKGHHAEKPLQASHATKHKPIKY-LEFISDAIIHVLHISHKHPGDFAGADAQGAMTKEALFRNDIAAKYKELGFQG-oh + heme$^{a}$</td>
<td>229</td>
<td>C$<em>{1039}$H$</em>{1638}$N$<em>{286}$O$</em>{338}$S$_{14}$</td>
</tr>
<tr>
<td>Trypsinogen</td>
<td>23992.81</td>
<td>h-VDDDKIVGQYTCGANTVPYYQVSLNSNGYHFCCGSGLINSQVWVSAAHCYSQ-1QVRLGEDINNVVQEGEQIISAASKSIVHPSYNSNTNNLMIKLAKASLASNVSAMLTSPCSAGTQCPCLISQWGNTKTSSGTYPDVLKCLAPKFPSSCSCKSAYFPQQT-SNMFCAAYLEGKDSQCDGDSGPVVCSCGKLQGVSWGSGCACQKNPGVYTKVCNYVSWIKQTIAISN-oh $^{f}$</td>
<td>229</td>
<td>C$<em>{1039}$H$</em>{1638}$N$<em>{286}$O$</em>{338}$S$_{14}$</td>
</tr>
</tbody>
</table>

$^{a}$Calculated using elemental formula and Bruker Xmass software.
$^{b}$UniProtKB/Swiss-Prot entry P62990
$^{c}$UniProtKB/Swiss-Prot entry P62894
$^{d}$heme binding residues marked in red.
$^{e}$UniProtKB/Swiss-Prot entry P68082; residues 2 to 154
$^{f}$UniProtKB/Swiss-Prot entry P00760; residues 15 to 243
pipette tip. A nanospray mass spectrum was produced by accumulating 32 scans. The mass spectrum for the peptide mixture sample is shown in Figure 2.10. For these peptides the major charge-states are 1+, 2+ and 3+. Neurotensin is the largest peptide in the mixture and the 3+ ion is the most abundant for this species, while bradykinin 1-7 is the shortest peptide and the most abundant ion for that species is the 1+ charge-state. This mixture of peptides with different abundances of charge-states allows for good calibration across the range of 400 to 800 m/z. There would be insufficient peaks for calibration if only one peptide were used for this mass range. The full width at half magnitude (FWHM) of the monoisotopic peak of angiotensin I at 532.77 m/z, is 0.00462 m/z corresponding to a resolving power of 113,000. The calculated monoisotopic for each species at different charge-state are shown in Table 2.6 along with the measured mass from the spectrum and the associated mass error after calibration. The estimated masses for the neutral peptides are: bradykinin 1-7 756.38 Da, an-
giotensin I 1296.66 Da, angiotensin II 1045.52 Da, angiotensin III 930.48 Da and neurotensin 1671.90 Da.

Table 2.6: Calibration table for a peptide mixture. The name, formulae and calculated monoisotopic masses are shown, along with measured masses and the mass error after calibration.

<table>
<thead>
<tr>
<th>Species</th>
<th>Formula</th>
<th>Charge</th>
<th>Theor. mass /m/z</th>
<th>Meas. Mass /m/z</th>
<th>Mass error /ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bradykinin 1-7</td>
<td>C₃₅H₄₄N₁₅O₉</td>
<td>2+</td>
<td>379.203213</td>
<td>379.203337</td>
<td>0.327</td>
</tr>
<tr>
<td>Angiotensin I</td>
<td>C₆₂H₉₂N₁₇O₁₄</td>
<td>3+</td>
<td>432.899773</td>
<td>432.899428</td>
<td>-0.799</td>
</tr>
<tr>
<td>Angiotensin III</td>
<td>C₄₆H₆₈N₁₂O₉</td>
<td>2+</td>
<td>466.261062</td>
<td>466.261025</td>
<td>-0.081</td>
</tr>
<tr>
<td>Angiotensin II</td>
<td>C₅₀F₁₇₃N₁₃O₁₂</td>
<td>2+</td>
<td>523.774534</td>
<td>523.774758</td>
<td>0.427</td>
</tr>
<tr>
<td>Neurotensin</td>
<td>C₇₆H₁₁₄N₂₁O₂₀</td>
<td>3+</td>
<td>558.310502</td>
<td>558.310705</td>
<td>0.364</td>
</tr>
<tr>
<td>Angiotensin III</td>
<td>C₄₆H₇₈N₁₂O₉</td>
<td>1+</td>
<td>931.514848</td>
<td>931.514336</td>
<td>-0.550</td>
</tr>
<tr>
<td>angiotensin II</td>
<td>C₅₀H₇₂N₁₃O₁₂</td>
<td>1+</td>
<td>1046.541792</td>
<td>1046.542096</td>
<td>0.290</td>
</tr>
</tbody>
</table>

2.5.3.2 Ubiquitin

Ubiquitin is a protein that has been well studied by mass spectrometry and its folding has been examined by using a variety of electrospray conditions. A ubiquitin (bovine, 8.5 kDa) solution was prepared by weighing out 0.0031 g of the crystalline solid into a micro test tube. This was dissolved in 1 ml of water, resulting in an estimated concentration of 362 μM. Aliquots of 50 μl of the solution were diluted 3 times with water and stored at 253 K until needed.

A fifty-times dilution to obtain a concentration of ~2.4 μM was made into a solution containing water, methanol and acetic acid (49.75 : 49.75 : 0.5 v/v). Figure 2.11 shows an electrospray mass spectrum obtained by accumulating 8 scans using a 1 M (2²⁰) data set. A broad charge-state distribution is observed in the mass spectrum as the entire amino acid chain is available for protonation. Charge-states from 7+ to 12+ are seen with a Boltzmann-type distribution. The most abundant charge-state is the 10+, for which an expansion is shown as an inset in the Figure. The isotope distribution is clearly resolved. The FWHM of the most abundant peak at 857.45 m/z, is 0.00415 corresponding to a resolving power of 206,000. The estimated mass for neutral ubiquitin is 8564.40 Da.
Figure 2.11: An electrospray ionisation mass spectrum of ubiquitin (H₂O : MeOH : AcOH 49.75 : 49.75 : 0.5 v/v). Inset is an expansion of the [M + 10H⁺]¹⁰⁺ charge-state with a resolving power > 206,000 at 857.45 m/z.

Figure 2.12: An electrospray ionisation mass spectrum of ferric cytochrome c (H₂O : MeOH : AcOH 49.75 : 49.75 : 0.5 v/v). Inset is an expansion of the [M + Fe³⁺ + 7H⁺]¹⁰⁺ charge-state with a resolving power > 176,000 at 1236.84 m/z.
2.5.3.3 Cytochrome c

Cytochrome c has been heavily studied by mass spectrometry, in terms of its folding characteristics as well as the oxidation state of the iron.\(^74,96,97,100,112,123,146-149\)

A cytochrome c (bovine, 12.4 kDa) solution was prepared from a 1 mM aqueous stock solution which had been stored at 253 K. The stock solution was diluted 160\(^\times\) to obtain a concentration of 6 \(\mu\)M in 1 ml containing water, methanol and acetic acid (49.75:49.75:0.5 v/v). An electrospray mass spectrum was obtained by accumulating one hundred scans and averaging the time-domain transient using a 1 M \((^{20})\) data set. Figure 2.12 shows the resulting mass spectrum of cytochrome c. Charge-states from 7+ to 16+ are seen with an unusual charge distribution. This could indicate only partial denaturation of the protein. The isotope distribution of the 10+ charge-state is shown as an inset. The FWHM at 1236.84 \(m/z\) is 0.00701 corresponding a resolving power of 176,000. The mass shift of 7 hydrogens indicates that there are 7 protons and a ferric ion, \(Fe^{3+}\), from the heme group. The estimated mass for cytochrome c is 12362.37 Da. Due to the strength of iron binding to the heme group and the covalent bonds of the \(c\)-type heme to Cys14 and Cys17, even at pH 2 no change in mass was observed.

2.5.3.4 Myoglobin

Myoglobin, which possesses a non-covalently bound heme group has been well studied making it now a well established system for investigating non-covalent interactions and protein conformation by MS.\(^{123,146,147,150-155}\) A stock solution of myoglobin (horse, 17.0 kDa) was prepared in water (1 mg / mL, \(~60 \mu\)M). A sample was prepared in denaturing conditions by diluting the stock 6-fold in water, isopropanol and ammonia solution (49.5 : 49.5 : 1 v/v). The sample was sprayed using the NanoMate in direct infusion mode in negative ion mode and the instrument tuned for maximum sensitivity. The resolution was then optimised and 100 scans were accumulated using a 1 M \((^{20})\) data set giving the spectrum shown in Figure 2.13. Under these conditions the \(apo\) protein is seen
Figure 2.13: An electrospray ionisation mass spectrum of myoglobin in denaturing conditions (H₂O : IPA : NH₃OH 49.5 : 49.5 : 1 v/v) in negative ion mode. Inset is an expansion of the [M - 10H⁺]¹⁰⁻ charge-state.

Figure 2.14: An electrospray ionisation mass spectrum of myoglobin in native conditions (pH 7, 20 mM NH₄AcO + 10 % IPA v/v). Inset is an expansion of the [M + Heme²⁻ + Fe³⁺ + NH₃ - 8H⁺]⁷⁻ charge-state.
with the release of the heme group. A broad range of charge-states is present, the most predominant being the $10^-$ charge-state, which is shown as an inset in the Figure, centred at 1694 $m/z$. The heme (674.21 $m/z$) appears predominantly with an adduct of $i$-propoxyl ($C_3H_7O^-$), rather than free heme (615.16) $m/z$), causing a mass shift of 59.05 Da. The free heme group has lost two protons but only has a single negative charge, indicating that under these conditions the iron is in the 3+ oxidation state.

Neutral conditions were then used to tune for soft ionisation in order to preserve the binding of the non-covalent heme group. The stock solution of cytochrome $c$ was diluted 6-fold into (20 mM, pH 7) ammonium acetate with 10% iso-propanol. An accumulation of 10 scans with a (2$^{20}$) data set gave the spectrum shown in Figure 2.14. Under these conditions just the $7^-$ charge-state is predominant. A expanded version of this peak, centred at 2511 $m/z$, is shown in the inset of the Figure. This peak is not simply due to the holo-protein but is shifted by +17 Da. This is likely to be substitution of $H_2O$ by ammonia as an axial ligand for iron. Under these conditions, the iron is in the 3+ oxidation state indicated by the loss of 8 protons to achieve the overall charge of $7^-$. The most abundant peak has been assigned as $[M + Heme^2^- + Fe^{3+} + NH_3 - 8H^+]^7^-$. This is the same oxidation state as seen in positive ion mode. The majority of previous studies of myoglobin have been conducted in positive ion mode, with the exception of that by Schmidt and Karas, which concluded that the heme complex was more stable in negative ion mode. The solution conditions they used were slightly different as 10 mM ammonium acetate was used without any organic solvent. The instruments (QIT and oTOF MS) they employed were of considerably lower resolving power than shown here. The formation of an ammonia adduct was not considered and the oxidation state of the iron was only reported in positive ion mode using an oTOF MS after collisional dissociation from the protein.
2.5.3.5 Trypsinogen

A trypsinogen (bovine, 24.0 kDa) solution was prepared by weighing out 0.0047 g of the crystalline solid into a micro test tube. This was dissolved in 1 ml of water, resulting in a concentration of 196 µM. The stock solution was diluted ten times to obtain a concentration of 20 µM in 1 ml containing water, methanol and acetic acid (49.5 : 49.5 : 1 v/v). Acidic conditions are required for the stability of trypsinogen as it can auto-activate and auto-cleave itself to form trypsin.\textsuperscript{157,158}

An electrospray mass spectrum was obtained by accumulating 32 scans using a 1 M (2\textsuperscript{20}) data set. Figure 2.15 shows the resulting mass spectrum. Charge-states from 11+ to 17+ are seen, with 12+ and 13+ having a much higher abundance than the other ions. The isotope distribution of the 13+ charge-state is shown as an inset. The FWHM of this peak at 1845.85 m/z is 0.01760 m/z corresponding to a resolving power of 104,000. The estimated mass for trypsinogen is 23982.95 Da.

![Electrospray ionisation mass spectrum of trypsinogen](image)

Figure 2.15: An electrospray ionisation mass spectrum of trypsinogen (H\textsubscript{2}O : MeOH : AcOH 49.5 : 49.5 : 1 v/v). Inset is an expansion of the [M+13H\textsuperscript{+}]\textsuperscript{13+} charge-state with a resolving power >104,000 at 1845.85 m/z. A noise spike is indicated by *.
2.6 Summary

This Chapter provided details of the solvents and samples purchased during this work, along with the quality, supplier and order codes. Gel filtration was introduced and implemented using a HPLC column or a single use column driven by gravity in order to remove salt, unsuitable buffers or other contaminants prior to MS analysis. Ultrafiltration using small centrifugal units was also described as an alternative to gel filtration for sample preparation.

The FT-ICR MS instruments used during this work were the Bruker Daltonics 9.4 T Apex III and the Bruker Daltonics 12 T Apex Qe. The Advion TriVersa NanoMate chip-based nanoESI robot that has been utilised during this work has been described. Ion trapping and detection using the ‘Infinity’ ICR cell found in the Bruker Daltonics FT-ICR MS instruments has been discussed. The software that has been used for data acquisition were Xmass for the Apex III and apexControl for the Apex Qe mass spectrometers. Data was post-processed and analysed using DA and BioTools.

Calibration for the mass range of 500–3000 m/z can be accomplished using ‘EST tuning mix’ from Agilent Technologies. This is a mixture of substituted triazatriphosphazine compounds providing singly-charge masses at 622, 922, 1522, 2122 and 2722 m/z in positive-ion mode and 602, 734, 1034, 1634, 2234 and 2734 m/z in negative-ion mode. For calibration at higher m/z values, cesium tridecafluoroheptanoate cluster series was used. CsTFHA clusters were particularly useful for calibrating spectrum of proteins under non-denaturing conditions (10-20 mM NH₄AcO, pH 7-8, ≤ 10 % MeOH or IPA) which often give rise to only a few peaks in the region of 2000–4000 m/z.

Instrument tuning was achieved using a variety of samples with a molecular weight ranging from 756 Da to 24 kDa. A peptide mixture containing angiotensin I (1296.5 Da), angiotensin II (1046.2 Da), angiotensin III (931.1 Da),
neurotensin (1671.9 Da) and bradykinin fragments 1-7 (756.4 Da) allowed for good calibration across the range from 400 to 800 m/z. Mass spectra were obtained for ubiquitin (8.6 kDa), cytochrome c (12.4 kDa) and trypsinogen (24.0 kDa) using denaturing conditions (1:1 H₂O : MeOH + 0.5-1 % AcO₂H v/v) each showing a broad range of charge-states. Myoglobin (17.0 kDa) was studied under denaturing (H₂O : IPA : NH₃OH 49.5 : 49.5 : 1 v/v) and native-like conditions (pH 7, 20 mM NH₃AcO + 10 % IPA v/v) using negative-ion mode. Using denaturing conditions a broad range of charge-states were observed for myoglobin and the heme group was not retained on the protein. Using native-like conditions, along with ‘gentle’ source conditions it was possible to retain the heme group on the protein (17.6 kDa). A mass shift of +17 Da was attributed to the substitution of H₂O by NH₃ as the axial ligand to the iron in the heme group. The oxidation state of the iron in the heme groups of cytochrome c and myoglobin were determined to be Fe³⁺ using isotopic peak fitting.
In the work described in this Chapter bovine carbonic anhydrase isozyme II (CAII) has been used as a model system for studying a medium sized protein (29 kDa). Carbonic anhydrase catalyses the reversible hydration of carbon dioxide: \[ \text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+. \] A catalytic zinc is bound at the centre of the protein via 3 three histidine residues. Sample preparation has a strong influence on the analysis and type of experiments that can be performed. Using high levels of organic solvent (MeOH ~50 % v/v) in acidic conditions will lead to high sensitivity. However, this will denature the protein leading to protein unfolding, releasing any bound metals. Buffered solutions (NH₄AcO, pH 7–8) will preserve protein folding and retain any metals allowing for ligand binding studies to be performed. Sulfonamides are a potent family of CAII inhibitors. By adding an non-covalent inhibitor to a solution of CAII, a complex of the protein and inhibitor should form binding to the \( \text{Zn}^{2+} \) site. Successful transfer of the complex into the gas phase will result in a mass shift corresponding to the inhibitor binding to the protein. Systematic excitation of the complex can lead to dissociation providing information on the strength of binding between the protein and ligand. Careful titration of the ligand against the protein can provide absolute binding affinities by detection of complex and free protein peaks in the mass spectrum. Fragmentation of a protein in the gas phase can provide useful sequence informa-
tion and sites of modified residues. During the course of the study ECD spectra have also been recorded for CAII, using the 12 T FT-ICR MS, to illustrate the extent of fragmentation and information that can be obtained for medium sized protein.

### 3.1 Carbonic Anhydrase II

In order to tune the FT-ICR MS instrument for larger proteins, CAII was chosen due to its availability and its size 29 kDa. The amino acid sequence is shown in Figure 3.1. The N-terminus is acetylated at the serine. The three histidine residues that bind to the zinc are His93, His95 and His118. The final ligand in the catalytic tetrahedral co-ordination site is usually a water / hydroxide molecule. The isoelectric point, pI, was calculated from the sequence to be 5.83. The net charge was calculated to be $-2.88$ at pH 7, while in acidic conditions (pH 2) the net charge is $+37.22$. The crystal structure of CAII is shown in Figure 3.2. The three histidine residues and the zinc ion are shown.

![Figure 3.1: The 259 amino acid sequence of bovine carbonic anhydrase II. The zinc binding residues are highlighted in red and the N-terminal acetylserine is highlighted in blue.](image)

### 3.2 Mass Spectra of CAII

In the following two sections mass spectrometric analyses of carbonic anhydrase under denaturing conditions and native-like conditions are compared and con-
Figure 3.2: The X-ray crystal structure of bovine carbonic anhydrase II reproduced using 1V9E entry from the RCSB protein data bank. The zinc ion is shown in the centre along with the three histidine ligands, His93, His95 and His118.


Figure 3.3: A broadband electrospray ionisation mass spectrum of carbonic anhydrase II using denaturing conditions (H$_2$O : MeOH : AcOH 49.5 : 49.5 : 1 % v/v). Inset is an expansion of the [M + 29H$^+$]$^{29+}$ charge-state at 1001.86 m/z, a resolving power of 140,000 along with the theoretical isotope pattern (△).

3.2.1 Denaturing Conditions

A 10 μM carbonic anhydrase sample was prepared in 1:1 H$_2$O : MeOH containing 1% acetic acid (v/v). The sample was infused using nano-electrospray with the Advion NanoMate robot into the 12 T Apex Qe mass spectrometer. The mass spectrum, shown in Figure 3.3, was obtained from an accumulation of 100 scans using a 1 M (2$^{20}$) data set. The spectrum exhibits a broad number of charge-states from 17+ to 41+, showing that the protein had been denatured and that a large variety of sites are available for protonation. The large number of charge-states is consistent with the protein being denatured and losing its native fold. The zinc ion has not been retained on the protein. An expansion of the
peak corresponding to the 29+ charge-state is shown as an inset in the Figure. The spectrum was calibrated internally using the most abundant isotope of each charge-state, calculated from the amino acid sequence. The theoretical isotope distribution (▲) is overlaid in the Figure, and was calculated from the elemental formula C$_{1312}$H$_{2025}$N$_{358}$O$_{384}$S$_3$. After calibration the measured mass for the most abundant isotopic peak was 0.2 ppm different from the theoretical value. The isotopic peaks shown are clearly baseline resolved, with a mass resolution of 140,000 and signal-to-noise ratio of 267 for the most abundant isotopic peak at 1001.86 m/z. The deconvoluted mass of apo-CAII was calculated from the most abundant isotope as 29024.71 Da (calculated average molecular mass 29024.32 Da).

### 3.2.2 Native-Like Electrospray Conditions

A carbonic anhydrase II solution was prepared by dissolving 0.0010 g of the crystalline solid in 1 ml of water containing 10 mM ammonium acetate pH 7, resulting in a concentration of 34 μM. The stock solution was diluted three times with the buffer to obtain a concentration of ~11 μM CAII in 10 mM ammonium acetate at pH 7 with the addition of 10 % methanol (v/v). An electrospray mass spectrum was obtained by accumulating 100 scans with a 1 M (2^20) data set using the 9.4 T Apex III mass spectrometer. The capillary voltage was 269 V. Figure 3.4 shows the resulting mass spectrum. There are two charge-states which are predominant in the spectra, 10+ which is the most abundant and the 11+ charge-state. The 12+ and 9+ charge-states can be identified but are only just above the baseline noise level. The zinc ion in the protein has been retained and the few charge-states observed suggest that most of the native fold has been conserved. Shown as an inset in Figure 3.4, is an expanded view of the 10+ charge-state for the protein, [CAII + Zn$^{2+}$ + 8H$^{+}$]$^{10+}$, showing the isotopically resolved peaks. For the most abundant isotopic peak at 2909.67 m/z a resolving power of > 200,000 has been achieved with the signal-to-noise ratio of 840. In order to achieve this mass resolution on the 9.4 T instrument gated dynamic trapping was used as
Figure 3.4: An electrospray ionisation mass spectrum of carbonic anhydrase II using native-like conditions (NH₄AcO 10 mM, pH 7: MeOH 90 : 10 v/v). Shown in the inset is an expansion of the [CAII + Zn²⁺ + 8H⁺]¹⁰⁺ charge-state at 2909.67 m/z, a resolving power of > 200,000 along with the theoretical isotope distribution (▲).

described in Section 2.3.4. The mass spectrum was calibrated externally using CsTFHA clusters as described in Section 2.5.2. The theoretical isotope distribution (▲) for the 10+ charge-state was calculated from the elemental formula \( \text{C}_{1312}\text{H}_{2004}\text{N}_{358}\text{O}_{384}\text{S}_{3}\text{Zn} \). The mass accuracy between the theoretical and measured value for the most abundant isotopic peak is 6 ppm, which is good for this mass-to-charge range using external calibration. The deconvoluted mass of holo-CAII was calculated from the most abundant isotope as 29088.65 Da. The mass difference between the holo and apo protein is 63.94 Da, close to the mass for the most abundant isotope of zinc, 63.93 Da.
3.3 Carbonic Anhydrase II Inhibitor Studies

Unsubstituted sulfonamides are an important family of compounds which are specific and potent inhibitors of carbonic anhydrase.\textsuperscript{159,161-164} These bind to the metal as anions \( R-\text{SO}_2\text{NH}^- \) via the nitrogen atom of the sulfonamide group.\textsuperscript{159} Acetazolamide (aza) is a particular example of one such sulfonamide and has been well studied against the human\textsuperscript{165,166} and bovine\textsuperscript{167-169} forms of carbonic anhydrase II. The structure of acetazolamide is shown in Figure 3.5 and has an elemental formula of \( C_4H_6N_4O_3S_2 \). When acetazolamide binds to the protein, little change in the secondary and tertiary structure is observed.\textsuperscript{169}

![Figure 3.5: The structure of the carbonic anhydrase II inhibitor; acetazolamide.](image)

The dissociation constant for bovine CAII-acetazolamide complex is reported to be 8 nM, which is similar to the values found for human CAII.\textsuperscript{166,167,170} This relatively strong binding makes it a useful system for studying non-covalent complexes in the gas phase.

CAII inhibitor complexes have been studied using FT-ICR MS previously.\textsuperscript{171,172} These earlier studies used a library of inhibitors in competitive binding experiments. The complexes were isolated and fragmented to reveal which ligands were binding and the intensities of these fragment peaks were correlated to the solution binding constants obtained from fluorescence studies. The relative binding affinities measured in the gas phase matched well with the solution constants, but absolute binding constants were not determined. SORI-CID was used to probe the stability of adducts with 8 selected inhibitors, but only for the 10+ charge-state of the complex.\textsuperscript{173}
This section describes the measurement of CAII-acetazolamide complexes in the gas phase. In-source nozzle-skimmer dissociation and IRMPD techniques were used in order to probe the strength of the interaction. This approach allows for simultaneous measurement of the stability of all the charge-states of the complex as there is no isolation of the parent ion.

### 3.3.1 In-Source Nozzle-Skimmer Dissociation

Bovine carbonic anhydrase II (5 mg) was dissolved in 1 mL of water and aliquots were stored at 253 K until required. The protein was desalted and buffer exchanged into 10 mM ammonium acetate using ultrafiltration. An aliquot (4 μL) of acetazolamide solution (2.1 mg) dissolved in 1 mL of (DMSO) was diluted with 996 μL of ammonium acetate solution (10 mM, pH 7). In order to determine more accurately the solution concentrations for the ligand acetazolamide as well as that for the protein CAII their UV absorbance spectra were recorded. The resulting values were 47 ± 1 μM for acetazolamide ($A_{270} = 0.37016, ε_{270} = 7900$ mol$^{-1}$ dm$^3$ cm$^{-1}$) and 180 ± 15 μM for CAII ($ε_{280} = 55300$ mol$^{-1}$ dm$^3$ cm$^{-1}$ using $A_{280}^{100\%} = 1.9$, molecular weight of 29088.81 Da). The solutions were combined and diluted with ammonium acetate buffer and methanol to give a resulting sample of 11 ± 1 μM CAII, 12.2 ± 0.3 μM acetazolamide in 10 mM ammonium acetate pH 7 containing 10 % MeOH and 0.04 % DMSO (v/v). Acetic acid was used to adjust the pH of a 10 mM ammonium acetate solution to a value of 2. A sample was similarly prepared to produce apo-CAII using the pH 2 buffer with acetazolamide.

Mass spectra for the CAII-aza complex at three different capillary voltages are shown in Figure 3.6 (acquired using similar tuning parameters as for Figure 3.4). At a capillary voltage of 269 V, see Figure 3.6a, there was no apparent mass increase on any peak to indicate formation of the protein-ligand complex. The charge-states observed had reduced from $9^+ \rightarrow 11^+$ to $7^+ \rightarrow 9^+$, with the $9^+$ ion being the most predominant. When the capillary voltage was lowered to
Figure 3.6: Electrospray ionisation mass spectra of a solution containing carbonic anhydrase II and the inhibitor acetazolamide (aza). The 7+, 8+ and 9+ charge-states are labelled. The capillary voltage was set to a) 269 V, b) 175 V and c) 100 V.
100 V, the complex remained intact, see Figure 3.6c. The charge-state distribution shifted once again, resulting in the 8+ ion being the most abundant. Further mass spectra were recorded using the same sample while varying the capillary voltage over the range 100 → 269 V in order investigate the dependence of complex stability on this source voltage. With the capillary voltage set to 175 V the 9+ charge-state of the protein has almost no acetazolamide bound, see Figure 3.6b. For the 8+ charge-state the abundance of the complex bound is ca twice that of the protein, and for the 7+ charge-state almost all the ions have acetazolamide bound. In the complexes the ligand is labelled as az$^-a$ denoting a negative charge on the acetazolamide ligand. This is in accordance with the literature and with the observed shift in the charge-states. The deconvoluted mass of the holo-CAII-aza complex was calculated from the most abundant isotopic peak for the 8+ charge-state as 29310.84 Da. The mass difference between the holo-protein and the holo-protein-inhibitor complex is 222.19 Da. The calculated mass of the most abundant isotope of acetazolamide is 221.99 Da. The overall mass difference is very close to the neutral mass of the ligand and so the formal charge on the ligand must balance out a with proton binding elsewhere on the protein. The observation that the most abundant charge-state for the protein ligand complex is 8+, compared to 10+ for the holo protein alone, would suggest that another proton has been displaced potentially from the binding domain. The binding of acetazolamide and the inhibition of CAII has been shown to be pH dependent and the involvement of proton transfer from side-chain amino acids was suggested to be involved in the mechanism for the catalytic activity of CAII. Further studies involving a pH titration may highlight a proton mass shift and the pH dependence on ligand binding.

The abundance of the ions with acetazolamide bound [RL] relative to that of the free protein [R] was calculated from the integrated intensity of each peak at each charge-state including salt adducts (Figure 3.7). The relative abundances for the different charge-states show that ligand binding is weaker the higher the
Figure 3.7: The ratio of carbonic anhydrase II with bound acetazolamide [RL] to unbound protein [R] was calculated from the integrated peak intensity for each charge-state. An expansion of the 8+ charge-state at capillary voltage of 175 V, showing the integral boundaries for [R] and [RL].

charge. Figure 3.8 shows the normalised relative abundances \([\text{RL}] / [\text{R}]\) as a function of capillary voltage. For the following charge-states, 7+, 8+ and 9+ at pH 7 and the 9+ charge-state at pH 2. A sigmoidal curve was used to fit the data to the line of best fit as shown using Equation 3.1.

\[
f(x) = \frac{a}{1 + \exp((x-b)/c)} + d
\]  

(3.1)

The four parameters, \(a\), \(b\), \(c\) and \(d\) were used to optimise the fitting curve to the data. The least squares fit minimised parameters are given in Table 3.1. As the data has been normalised, parameter \(a\) should be equal to 1 and \(d\) should be 0. The parameter \(b\) correlates to the capillary voltage, and in turn to the collisional energy, required to dissociate half of the protein-inhibitor complex. As the charge-state increases, the parameter \(b\) decreases. It seems reasonable that as the charge-state increases, the presence of a greater number of protons destabilises
the protein and induces unfolding as the electric charge repels different parts of
the structure. The parameter $c$ is related to the gradient of the sigmoidal function.
There is less of a correlation with the charge-state and this parameter. Instead
there is a clear difference between the respective values for the holo ($c=8.31$) and
apo protein ligand complexes ($c=16.30$). An increase in the $c$ parameter means
that dissociation occurs over a broader range of energies. This might be expected
for non-specific binding to the number of conformations that would be present
for the partially unfolded apo protein. However, the relative strength should not
be considered in isolation, as abundance of the apo protein-ligand complex was
very low, so the abundance of this complex was much smaller than that of the
holo protein. Removal of the zinc ion by using a chelating agent, such as EDTA,
instead of using acidic conditions, might provide a more meaningful comparison

![Figure 3.8: A plot of the normalised ratio of carbonic anhydrase II with bound acetazolamide [RL] to unbound protein [R] versus the capillary voltage for the 7+ (■), 8+ (▲) and 9+ (●) charge-states of the holo protein in pH 7 buffer. For the apo protein in pH 2 buffer, only data for the 9+ charge-state (●) is shown. The sigmoidal curves were calculated by using a minimised fit to Equation 3.1.](image)
between apo- and holo-CAII-ligand binding.

Table 3.1: The parameters for the sigmoidal fit to Equation 3.1 for each charge-state for the nozzle-skimmer dissociation of the carbonic anhydrase-acetazolamide complex. The capillary voltage was altered to probe the strength of binding.

<table>
<thead>
<tr>
<th>Species</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>7+ at pH 7</td>
<td>0.99</td>
<td>192.07</td>
<td>10.79</td>
<td>-0.02</td>
</tr>
<tr>
<td>8+ at pH 7</td>
<td>1.00</td>
<td>169.22</td>
<td>7.73</td>
<td>-0.01</td>
</tr>
<tr>
<td>9+ at pH 7</td>
<td>1.00</td>
<td>148.74</td>
<td>8.31</td>
<td>0.00</td>
</tr>
<tr>
<td>9+ at pH 2</td>
<td>1.03</td>
<td>159.89</td>
<td>16.30</td>
<td>0.00</td>
</tr>
</tbody>
</table>

### 3.3.2 IRMPD

Another method of dissociating the complex is by using infrared photons. In this case dissociation occurs inside the analyser cell, rather than in the source region where the ions are formed. Infrared Multi-Photon Dissociation (IRMPD) has previously been used to fragment non-covalent complexes and studies have shown that the technique is not capable of providing absolute activation energies.\(^{175,176}\) However, when applied to similar systems under similar conditions, IRMPD can provide relative activation energies which can be compared between samples.\(^{177}\) Large biomolecules will most likely have many vibrational modes which could absorb energies at IR irradiation frequencies, and the energy transferred to the molecule is difficult to determine. For these reasons the study that follows is not comprehensive enough to provide an activation energy for the gas-phase complex, but explores the feasibility of investigating such a system using IRMPD.

A sample was prepared from stock solutions of ligand and protein as described previously, to give a resulting solution of 11 ± 1 µM CAII, 12.2 ± 0.3 µM acetazolamide in 10 mM ammonium acetate pH 8 with 10 % MeOH (v/v) and 0.04 % DMSO. As can be seen from the spectrum shown in Figure 3.6c, at a capillary voltage of 100 V virtually no signal due to the unbound protein can be detected. Therefore, in the IRMPD experiments, the ions generated using this capillary voltage were transferred and trapped in the analyser cell where they
were irradiated with infra-red photons. Successive spectra were recorded from an accumulation of 10 scans using a 1 M (20) data set. For each scan the trapped ions were irradiated by the CO₂ laser for a period of 0.4 s. Spectra were recorded with the power of the 25 W laser attenuated to 10, 12.5, 14, 14.5, 15 and 15.5 % of the maximum power output. IRMPD mass spectra of the holo-CAII-aza complex after irradiation are shown Figure 3.9. The spectra clearly show that the complex can be preferentially dissociated using low power IRMPD. Higher laser power lead to new peaks due to protein fragmentation. The 8+ and 9+ charge-state peaks in the spectra were integrated in a similar way as for the earlier study using nozzle-skimmer dissociation to fragment the complex. The 7+ charge-state had too low abundance to be included in this study. The relative abundance [RL]/[R] of the bound complex [RL] to the unbound protein [R] was calculated at each charge-state. This ratio was normalised and plotted as a function of laser power and is shown in Figure 3.10. Equation 3.1 was again used to fit the data to a sigmoidal curve and the optimised fitting parameters are given in Table 3.2. As before the parameter a should be 1 and d should be 0 due to normalisation. However, the quality of this data set is worse than for the nozzle-skimmer dissociation study. This is in part due to fewer data points but also due to the coarse control over the laser power for which the smallest increment was 0.5 %. An improvement on the method would be to use a fixed low laser power and to adjust the irradiation time, over which there is a much finer control, as well repeating the measurements several times in order to better establish reproducibility. Low power has been shown to follow more closely black body heating, which provides a much more reliable

Table 3.2: Parameters for the sigmoidal fit, to Equation 3.1 for each charge-state for the dissociation of the carbonic anhydrase-acetazolamide complex. Infrared laser radiation was used to probe the strength of binding.

<table>
<thead>
<tr>
<th>Species</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>8+</td>
<td>1.06</td>
<td>3.00</td>
<td>0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>9+</td>
<td>1.12</td>
<td>3.05</td>
<td>0.21</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Figure 3.9: IRMPD of holo-CAII-aza complex. The 8+ and 9+ charge-states are labelled. IR laser power was set to a) 3.6 W, and b) 3.1 W. Noise spikes are indicated by *
Chapter 3 — Mass Spectrometric Studies of Carbonic Anhydrase II

Figure 3.10: A plot of the normalised ratio \([RL]/[R]\) of carbonic anhydrase II with bound acetazolamide \([RL]\) to unbound protein \([R]\) versus infrared laser power, for the 8+ (\(\triangle\)), and 9+ (\(\bullet\)) charge-states of the *holo* protein in pH 8 buffer. The sigmoidal curves were calculated by using a minimised fit to Equation 3.1.

estimate of the activation energy.\(^{178}\) A short laser pulse of high power may not yield the same result as a long pulse of low power. The nature of the cyclotron motion means that overlap of the trapped ions with the laser beam may not be uniform for the entire irradiation time as ions could move in and out of the beam during this time. This experiment highlights the importance of the irradiation time and the laser power.

3.3.3 Titration of Acetazolamide With CAII

Once the source conditions of the instrument had been optimised for soft transfer of the CAII-acetazolamide complex, an attempt was made to measure the binding constant for this complex using ESI-MS. Direct measurement of dissociation constants is possible using electrospray ionisation by titration of the ligand against a constant concentration of the target molecule.\(^{125,130,131,179-183}\)
Samples were prepared from the CAII stock solution (180 ± 30 μM) described previously, to give a resulting solution of 11 ± 2 μM CAII in 10 mM ammonium acetate pH 8 with 10 % MeOH (v/v) and ≤0.04 % DMSO (v/v). Acetazolamide was added at concentrations of 12.2 ± 0.6, 9.3 ± 0.5, 6.2 ± 0.3 and 3 ± 0.2 μM. Spectra were obtained from an accumulation of 100 scans using a 1 M (2^10) data set. The capillary voltage was kept constant at 100 V. The ESI spectra obtained are shown in Figure 3.11. The measurements on the 6.2 and 12.2 μM solutions were repeated using the NanoMate chip-based nano electrospray robot, in this case but only 10 scans were accumulated.

The binding energy is dependent on the charge-state as shown in Section 3.3.1. For this reason the 10+, 9+, 8+ and 7+ charge-states were all considered for the calculation of the binding constant. The area of peaks corresponding to the bound complex [RL] and unbound protein [R] were calculated from the integral of each peak, using the same mass interval for each spectrum. The relative abundance of bound complex to unbound protein [RL]/[R] was determined from the sum over all charge-states.

A protein R, and a ligand L in equilibrium forming a complex RL can be written as

\[ R + L \rightleftharpoons RL \] \hspace{1cm} (3.2)

The dissociation constant \( K_D \) can be calculated from the concentrations of each of the species at equilibrium.

\[ K_D = \frac{[R][L]}{[RL]} \] \hspace{1cm} (3.3)

The initial concentration of protein and ligand added before equilibrium can be defined as [R]i and [L]i respectively. At equilibrium the total amount of protein and ligand must be conserved such that

\[ [L]i = [L] + [RL] \] \hspace{1cm} (3.4)

\[ [R]i = [R] + [RL] \] \hspace{1cm} (3.5)
Figure 3.11: Titration of acetazolamide with holo-CAII at pH 8. ESI spectra were obtained for a) 3 $\mu$M, b) 6.2 $\mu$M, c) 9.3 $\mu$M and d) 12.2 $\mu$M acetazolamide concentrations with a constant 11 $\mu$M concentration for holo-CAII. Approximate stoichiometric ratios of [holo-CAII : aza] are shown.
Equation 3.3 can be rewritten by substituting Equation 3.4 and rearranging to give
\[
\frac{[RL]}{[R]} = \frac{([Li] - [RL])}{K_D}
\] (3.6)
A plot of $[RL]/[R]$ versus $([Li]-[RL])$ should be a straight line with a gradient of $1/K_D$. This approach was used successfully by Zhang et al. for the determination of μM dissociation constants. Figure 3.12 shows a plot of $[RL]/[R]$ versus $([Li]-[RL])$ with a line of best fit to the data points. The calculated value for the binding constant is $K_D = 0.9 \pm 0.2 \, \mu\text{M}$. There is a relatively large error as the expected $K_D$ value is in the nM range. The concentration for the free ligand is calculated from the values of $[Li]$ and $[RL]$ both of which are in the μM range, these need to be very precisely known in order to determine $[L]$ with any accuracy. It is for this reason that the error bars are so large. The uncertainty in the so-

Figure 3.12: A plot of the ratio of carbonic anhydrase II with bound acetazo-

amide to unbound protein against the calculated concentration of free ligand at
equilibrium. Points shown in green (▲) are from pneumatic-driven electrospray
and red points (■) are from chip-based nano-electrospray. The line of best fit
was produced from Equation 3.6.
olution concentrations were determined from several UV measurements of protein and ligand solutions of varying dilutions. The reported literature values for the molar extinction coefficients used were only reported to 2 significant figures, and any error in these values would also be propagated in these calculations. The stock concentrations were determined from the mean of the values several measurements and the error bars represent one standard deviation (±1σ). While it is not uncommon for the solution and gas phase determined dissociation constants to differ, in this case the difference is a thousand fold.

An alternative approach is to plot the ratio [RL]/[R] versus the initial concentration of the ligand, which should be more accurately known than the equilibrium concentrations calculated from the ion abundance within the mass spectra. Figure 3.13 shows the plot of [RL]/[R] versus [Li]. A linear fit should also give a gradient of $1/K_D$ while the intercept should be zero as there can be no complex formed when there is no ligand present. This approach was used by Sannes-Lowery et al. The linear fit results in a calculated value for the binding constant of $K_D = 3.2 ± 0.4 \text{ μM}$. The result is the same order of magnitude as that derived above, however, the relative error is much lower.

Daniel et al. have used a different fit to the same plot shown in Figure 3.13. The fit takes the form shown in Equation 3.7

$$\frac{[RL]}{[R]} = \left( -1 - K_a \times [Ri] + K_a \times [Li] + \sqrt{4 \times K_a \times [Li] + (K_a \times [Li] - K_a \times [Ri] - 1)^2} \right) \quad (3.7)$$

While this equation is quite unwieldy, the only variable is the value of $K_a$, as the initial concentration of the protein, [Ri], has been kept constant throughout the experiment. The association constant, $K_a$, can be related to the dissociation constant by

$$K_D = \frac{1}{K_a} \quad (3.8)$$

The dashed line in Figure 3.13 shows the fit to Equation 3.7. This non-linear fit does correlate to the data quite well. The calculated value of the binding constant
Figure 3.13: A plot of the ratio of carbonic anhydrase II with bound acetazolamide to unbound protein against the concentration of ligand added. Points shown in green (▲) are from pneumatic-driven electrospray and red points (■) are from chip-based nano-electrospray. Two lines of best fit have been overlaid onto the data points. The solid line calculated using Equation 3.6 and the dashed lines from Equation 3.7.

is $K_D = 1.4 \pm 0.4 \mu M$. This value is again in the low $\mu M$ range, far higher than the 8 nM reported for solution.$^{166,167,170}$

Mass spectrometry of large biomolecules has a typical sensitivity in the low $\mu M$ range, although this is instrument- and system-specific. The sensitivity of fluorescence measurements is much higher allowing nanomolar concentrations to be employed; while titration calorimetry also requires the use of micromolar concentrations.$^{181}$ Good agreement has been reported for $K_D$ values in the micromolar range measured by mass spectrometry and solution-based techniques.$^{184}$ However, discrepancies can occur when $K_D$ values are nM or less.$^{126}$ This makes it necessary to know the magnitude of the strength of interaction prior to the design the experiment, which needs to be within the sensitivity of the equipment. Although
the use of mass spectrometry for determining absolute binding constants appears not to be very reliable, as sensitivity and reproducibility are problems especially for weak interactions, studies of competitive interactions, whereby the ligands are placed into a relative order of binding are far more successful.\textsuperscript{171,172,183,185} This approach seems to be quite robust even for hydrophobic interactions which are normally weakened in the gas phase.\textsuperscript{182}

Even when the ion source optics and other instrument parameters have been tuned for the softest possible transfer, some dissociation of a non-covalent complex cannot be ruled out. For this reason it can be difficult to determine binding constants using mass spectrometry. The response factor is a measure of how effective ionisation and ion transfer is from solution to the detector. It is a difficult parameter to determine directly. It is also possible for a compound, its ligand and the complex to all have different response factors, thus skewing the relative abundances observed in the mass spectra. Methods have been developed which try to take into account the response factor, using a reference complex or internal standard.\textsuperscript{130,181,183} The response factor for the carbonic anhydrase II-acetazolamide complex could also be an issue in determining the binding constant directly. Usually a large protein and a small ligand would be thought to have a similar response factor and thus not be affected by this problem.\textsuperscript{186} However, there is a change in the charge-state distribution with complexation which could have an effect on the ionisation efficiency.

While its use was kept to a minimum, DMSO was needed to dissolve the ligand initially. DMSO is a highly polar and non-volatile solvent, its presence will have an effect on the solvent evaporation, an important step in the ESI process, and thus lead to reduction in the ionisation efficiency.

The electrospray ionisation process may effect the chemical equilibria as ions go from solution to the gas phase. The droplets formed by electrospray shrink as solvent evaporates. It cannot be ruled out that droplet shrinkage could induce complex formation as the molecules become closer, although it is thought that
the shrinking of the droplets is too rapid to influence chemical equilibria in most cases.\textsuperscript{186,187}

While there have been successes in direct determination of binding constants using mass spectrometry, this is not always the case. The CAII system studied here is such an example. The errors might be reduced by repeating experiments and acquiring more data points, but this is unlikely to improve the sensitivity required for this nM binding model complex. This is one example why absolute determination of binding constants by MS has its limitations, while competitive and relative binding determination seems to be far more informative. Further discussions on this issue have been published by Peschke \textit{et al.}\textsuperscript{186}

\subsection*{3.4 Top-Down Analysis of CAII}
Carbonic anhydrase was also used as a model protein for top-down fragmentation. Top-down fragmentation is a two step process whereby the intact protein is measured in a first stage MS and is then selected for fragmentation for the second stage MS/MS. This approach is usually performed on an FT-ICR MS instrument using either CID, IRMPD or ECD. The high resolution and mass accuracy of FT-ICR MS is needed to enable resolution and identification of the many multiply-charged fragments that are formed. Ion/ion reactions in quadrupole ion traps have also been used for top-down analysis\textsuperscript{188,189}. However, this has been implemented only very recently in a commercial instrument.\textsuperscript{190} Top-down fragmentation of a protein has the advantage over traditional ‘bottom-up’ approaches of providing an accurate intact mass as well as the ability to select and sequence any modifications that may be present.\textsuperscript{191}

The 29+ charge-state of CAII, as shown in Figure 3.3, sprayed from denaturing conditions, was isolated in the mass-resolving quadrupole on the 12 T Apex Qe instrument. The mass isolation on the quadrupole was set to 1000 \textit{m/z} with a mass resolution of 20 \textit{m/z}. This isolation window was narrow enough to select a single charge-state whilst wide enough to maximise the transmission of the ions
to the ICR cell for fragmentation. The ECD lens voltage was set to 20 V and the heater current to 1.8 A. The cathode was set to 1.2 V and electrons were emitted from the cathode surface for 8 ms. The resulting ECD spectrum, shown in Figure 3.14, was acquired from an accumulation of 1000 scans using a 512 K (2¹⁹) data set.

The mass spectrum was processed using Bruker Daltonics DataAnalysis program (version 3.4 Build 172). The SNAP algorithm was used to produce a list of the monoisotopic peak and the charge for each peak identified in the mass spectrum. The peak list was then opened into Bruker BioTools (version 3.0 build 2.9) the MS/MS view was used to assign the peaks to fragments. The carbonic anhydrase II sequence was entered into the software and the theoretical fragments were calculated. Measured peaks were then matched with theoretical peaks for identification. Manual validation and peak assignment was performed in order to label the peaks as shown in Figure 3.14. A total of 364 peaks were identified using the SNAP algorithm. From this 310 peaks (85%) were assigned to ECD fragments (171 c' and 139 z'), 4 peaks were assigned as charge-reduced species ([M + nH]ⁿ⁻¹⁺) and the remaining 50 peaks could not be assigned to any fragment with confidence. The fragments were mapped onto the sequence to show where the cleavages occurred, as shown in Figure 3.15.

The sequence coverage is very good for the N- and C- terminal regions while there are far fewer fragments from the middle of the protein. A total of 140 backbone bonds have been cleaved leading to 54% coverage of the protein sequence. For a single experiment, this represents good coverage of the protein sequence, which can be attributed to the high charge-capacity of the 12 T magnetic field.

Carbonic anhydrase II and larger proteins have been fragmented before. In order to identify unambiguously a modified residue, it is necessary to obtain as much sequence coverage as possible. Using multiple spectra from different experiments, by altering source conditions to cause nozzle-skimmer dissociation, and thermal decomposition to complement CID and/or ECD data, it is possible
Figure 3.14: An electron capture dissociation spectrum of the isolated \([\text{M} + 29\text{H}]^{29+}\) charge-state of CAII at 1000 \(m/z\). 364 fragment peaks have been assigned. The major peaks are labelled. However due to space and overlapping species, not all fragments are labelled in the spectrum.
to obtain far greater sequence coverage." For carbonic anhydrase II, using 49 experiments Sze et al. have shown that it is possible to obtain 250 of the 258 possible peptide bond cleavages. This approach with the extension of additives to the electrospray solution has been used to analyse a 144 kDa protein using top down mass spectrometry. In this work, Han et al. showed that, of the 1314 residues only 287 peptide bond cleavages were possible, which were mostly identified from within the first 200 residues of the termini. This highlights the problem that the protein needs to be completely unfolded to sequence the centre portions of the protein effectively. The presence of disulfide bonds which hold the protein structure together exacerabates the problem. A study of a 300 kDa protein has also been reported. In this work, Karanam et al. used limited proteolysis with cyanogen bromide (CNBr) to produce large protein fragments. This approach is now being called ‘middle-down’ which uses methodology from both top-down and bottom-up techniques. It is a good compromise between the two and is useful for studying very large proteins and will no doubt gain popularity in the field of proteomics.
Chapter 3 — Mass Spectrometric Studies of Carbonic Anhydrase II

3.5 Summary

Bovine carbonic anhydrase II (29 kDa) was studied under denaturing and native-like conditions. Using denaturing conditions (H₂O : MeOH : AcOH 49.5 :49.5 : 1 % v/v) a broad charge distribution was observed from 17+ to 41+ indicating that the protein had unfolded. Ammonium acetate (10 mM, pH 7 + 10 % MeOH) was used in order to keep CAII in a native-like conformation. Far fewer charge-state were seen, 12+ to 9+, with the 10+ being the most abundant. Good resolving power (> 200,000) and signal-to-noise (840) was achieved at 2909.67 m/z. Comparing the masses of the apo and holo protein indicated the mass shift consistent with retention of the zinc ion.

Acetazolamide (222 Da) is an example of a sulfonamide which acts as an inhibitor of CAII. A complex of CAII and acetazolamide was studied at pH 8 (NH₄AcO 10 mM + 10 % MeOH). A mass shift was observed indicating that the inhibitor complex was being maintained through ionisation and into the gas phase. The strength of interaction between CAII and acetazolamide was investigated using in-source CID, IRMPD and a titration method. The capillary voltage was a critical parameter in tuning the source conditions for maintaining the complex into the gas phase. The capillary voltage was varied between the range 100–269 V in order investigate the stability of the complex. The relative stability of each charge-state could easily be distinguished, with the lower charge-state requiring a higher capillary voltage for dissociation, indicating stronger binding. IRMPD was used to dissociate the complex inside the ICR cell, in the gas-phase rather than during the ionisation process. Although only a few data points were acquired, the IRMPD data also suggested that the lower charge-state was more strongly bound. In order to obtain an absolute binding affinity rather than the relative binding affinities provided by CID or IRMPD, a titration method was used. The protein concentration was fixed at ~10 μM while the ratio of acetazolamide was altered. Accurately determining the concentrations of each species present at
equilibrium was one of the largest sources of error. Different methods for analysing
the data were also used in order to calculate an absolute binding constant. The
analysis gave a $K_D$ value ranging from 0.9 to 3.2 $\mu$M which is 125-400 hundred-
fold larger than values obtained using fluorescence studies in solution$^{166,167,170}$

Under denaturing conditions (H$_2$O : MeOH : AcOH 49.5 :49.5 : 1 % v/v) the
29+ charge-state was isolated in the mass resolving quadrupole on the Apex Qe
MS and subjected to ECD fragmentation. From the total of 364 peaks which
were observed, 310 peaks (85%) were assigned to ECD fragments, 4 peaks were
assigned as charge reduced species and the remaining 50 peaks could not be
assigned. These fragments showed that 54% of the protein sequence were cleaved
in a single experiment, showing how useful ECD can be for obtaining sequence
information.
Chapter 4
Mass Spectra of Recombinant Metalloproteins

In this Chapter FT-ICR mass spectra for the zinc binding protein, YdaE, and the ferric-ion binding protein, FbpA, are presented. The stoichiometry of the metal binding can be determined directly from the measured mass of the protein and the shift from the apo-protein to the metal-bound protein (holo) can be used to identify the metal present. Metalloproteins are sensitive to the pH, as protons will readily compete with the metal at the binding site. An appropriate pH is needed in order to maintain the metal binding and keep the protein stable. In-source collisions can often break non-specific and undesirable buffer adducts from the protein, which improves the signal-to-noise ratio and therefore improves the overall spectral quality. Gentle ionisation is required in order to preserve as much of the solution binding as possible into the gas-phase. However gentle source conditions mean that buffer adducts are more likely to be transferred into the mass spectrometer reducing the ionisation efficiency. In order to use such gentle conditions it is imperative that the protein be pure and free from non-volatile buffers used to stabilise the protein for storage. Mass spectrometry can also be used to study metal exchange reactions or proteins with isotopically-enriched metals which can be followed by a the shift in mass.
4.1 YdaE: A Zinc Finger Protein

In earlier work Blindauer et al. showed that the bacterial metallothionein SmtA from *Synechococcus* PCC 7942, contained 4 zinc ions per protein. While 3 of those zinc ions were readily exchanged like other metallothionein clusters, one of the zinc ions was found to be completely inert, and had an associated zinc-finger-like fold. A sequence similarity search was performed in order to find other bacterial zinc-finger-like metallothioneins. Four proteins were discovered and studied in further detail including the YdaE protein for *Escherichia coli*, which was a zinc-finger protein which binds 1 zinc ion. While the role of YdaE remains unknown, the lack of a zinc cluster suggests that it is unlikely to be involved in zinc regulation. An involvement in macromolecular interactions has been suggested for YdaE. The T4 endonuclease VII protein which has been shown to form an S-shaped dimer and resolve Holliday junctions has been used for comparative modelling of YdaE. YdaE could therefore interact with DNA or RNA as part of its function like many other zinc-finger proteins.

The 56 amino acid sequence of YdaE is shown in Figure 4.1. When recombinantly expressed, the N-terminal methionine is observed to be cleaved in some but not all of the protein. The predominant species is 55 amino acids long. This is common in the over expression of recombinant proteins.

Initial studies by Blindauer et al., using a titration of *p*-(hydroxymercuri)phenylsulfonate (PMPS) in the presence of 4-(2-pyridylazo)resorcinol (PAR) and monitoring the absorbance at 492 nm show that zinc is held in place by four cysteines, Cys7, Cys12, Cys42 and Cys46 which are shown in Figure 4.2. The structure shown is a homology model using an alignment of SmtA and T4 endo-
Figure 4.2: A suggested structure for YdaE using cartoon representation. The zinc and the four bound cysteines are shown at the top centre of the structure.

donuclease VII, subjected to energy minimisation after the addition of a zinc ion. The antiparallel $\beta$-sheets and $\alpha$-helix form a characteristic treble-clef fold.\(^{201,202}\)

Recent studies on YdaE have provided contradictory data from different analytical methods (N. L. Fowlis and C. A. Blindauer unpublished). During gel-filtration purification, YdaE elutes much earlier than would be expected from a relatively small 56 amino acid protein, suggesting that the molecular size is much larger than a single $\sim$6.5 kDa protein. There may also be a possibility that more than 1 metal ion can bind to the protein. A potential second binding site is labelled in Figure 4.2, which includes His27, Asp50 and His54. This raises the question as to whether YdaE may form multimers in solution and if it does is this important for its function?
Chapter 4 — Mass Spectra of Recombinant Metalloproteins

4.1.1 Sample Preparation of YdaE

The zinc finger protein YdaE from *Escherichia coli* K-12 was expressed in *Escherichia coli* BL21(DE3) cell lines using the YdaEpET29a plasmid. The protein was extracted and purified using an Amersham-Biosciences Q-Sepharose anion exchange column followed by a gel filtration column in 25 mM pH 9 Tris-HCl buffer. A salt gradient (0 to 500 mM) NaCl was used to elute the protein from the anion exchange column, while a constant 150 mM concentration of NaCl was used for gel filtration. This was a slightly modified method than reported in the literature. The protein was concentrated using a Vivascience Vivaspin 3000 MWCO PES concentrator.

4.1.2 Studying YdaE by Mass Spectrometry

The protein was first desalted and buffer-exchanged using a PD10 column (see Section 2.2.1.2) and then concentrated to 200 μM using ultrafiltration. The protein was then diluted twenty times into 0.5 ml in ammonium acetate buffer (10 mM), at either pH 2 (for *apo* protein) or pH 9 (for *holo* protein) and methanol (90:10 v/v). Spectra were calibrated externally using ESI tune mix as described in Section 2.5.1. Attempts to calibrate internally using ESI tune mix resulted in ion suppression of tune mix by the protein resulting in poor calibration due to very low abundance of the calibration peaks.

4.1.2.1 Investigation of *apo*-YdaE

The protein was dissolved in ammonium acetate buffer at pH 2 in order to study the *apo* form. Figure 4.3 shows the mass spectrum obtained from acquisition of 100 scans with a 1 M (2²⁰) data set. The protein shows a broad number of charge-states from 4+ to 9+ with the most abundant being the 6+. The 6+ charge-state is labelled as [YdaE + 6H⁺]₆⁺ and the theoretical isotope distribution (▲) calculated from C₂₇₈H₄₅₇N₇₉O₈₅S₇ and overlaid onto the spectrum.

---

†Expression and purification of YdaE was carried out by Dr N. L. Fowlis
Figure 4.3: The mass spectrum for apo-YdaE sprayed from ammonium acetate at pH 2. The 6+ charge-state has been expanded and the theoretical isotope pattern (▲) calculated for C\textsubscript{278}H\textsubscript{457}N\textsubscript{79}O\textsubscript{83}S\textsubscript{7}. The peaks labelled with * indicate a mass shift for the additional N-terminal methionine.
There is a good match between the measured and the calculated isotope pattern, however there are some discrepancies with the isotope abundances. The masses of isotopes lighter than the most abundant appear to have lower abundance than the calculated values while heavier isotopes appear to have higher abundance than the calculated values. The monoisotopic peak is 2 mass units lighter than expected. The reason for this is discussed in Section 4.1.3. The mass accuracy was calculated to be 0.46 ppm at 1291.44 \( m/z \). The mass resolution for the most abundant isotope of the 6+ charge-state is 60,000.

Each charge-state appears as two peaks with a mass difference of 131.05 Da. This is very close to the mass of the methionine which has been cleaved off of the N-terminus. The peaks for proteins that contain methionine are labelled with *, but these peaks are much less abundant than those without methionine. Unless otherwise stated when referring to YdaE this implies the predominant species without the methionine.

The deconvoluted mass for the most abundant isotope of apo-YdaE was calculated to be 6452.16 Da.

### 4.1.2.2 Investigation of holo-YdaE

In order to study holo form, the protein was dissolved in ammonium acetate at pH 9. Figure 4.4 shows the mass spectrum of holo-YdaE obtained from an acquisition of 100 scans using a 1 M (2^{20}) data set to record the transient. The most abundant charge-state is the 5+, while 4+ and 6+ ions have much lower abundance. The 3+ state is visible, however the abundance is very low. This narrow charge-state distribution is typical for a protein sprayed from native-like conditions. The 5+ charge-state is labelled as \([\text{YdaE} + \text{Zn}^{2+} + 3\text{H}^+]^{5+}\) and the theoretical isotope distribution (\(\bigtriangleup\)) calculated from \(\text{C}_{278}\text{H}_{454}\text{N}_{79}\text{O}_{83}\text{S}_{7}\text{Zn}\) is overlaid onto the spectrum. There is a good agreement between the measured and the calculated isotope pattern. The mass accuracy is calculated to be 0.4 ppm at 1304.02 \( m/z \). The mass resolution for the most abundant isotope of the
Figure 4.4: The mass spectrum for holo-YdaE sprayed from ammonium acetate at pH 9. The 5+ charge-state has been expanded and the theoretical isotope pattern (▲) calculated for C_{278}H_{454}N_{79}O_{83}S_{7}Zn. The zinc ion has been retained by the protein during ionisation. The peaks labelled with * indicate a mass shift for the additional N-terminal methionine.
5+ charge-state is 140,000.

The deconvoluted mass for the most abundant isotope of \textit{holo}-YdaE was calculated to be 6517.09 Da. The mass difference between the \textit{holo} and \textit{apo} forms is 64.92 Da. This is close to the atomic mass of zinc (65.40 Da). Although there is a small discrepancy in mass difference, which comes from comparing the most abundant isotope peaks of each protein, the spectrum does show that the zinc has been retained at pH 9.

In order to investigate whether additional zinc could bind to the protein, a sample with a 5-fold excess of zinc added was incubated for 1 h or overnight prior to removal of any excess metal ion by ultrafiltration. Acquisition of the protein mass spectrum after incubation and clean-up showed no additional zinc binding to the protein.

4.1.2.3 Formation of YdaE-Dimer

At a protein concentration of 10 \textmu{}M there is evidence for formation of a dimer. Both \textit{apo}- and \textit{holo}-YdaE form a dimer, as shown in Figure 4.5, for the respective 7+ charge-states. This charge-state was the only one identified for the dimers. Even charge-states would of course underlay the more intense monomer peaks. Like the monomer, the dimer can be formed with and without the N-terminal methionine. The ratio of the most abundant peak of the 5+ charge-state of the monomer to the 7+ charge-state of the dimer is of the order 40 : 1. The deconvoluted masses of the \textit{apo} and \textit{holo} dimer were calculated to be 12905.43 Da and 13035.24 Da which correspond to exactly twice the mass of the monomers. The mass difference between the \textit{apo} and \textit{holo} dimer is 129.81 Da, which indicates that each monomer still carries one zinc ion in the dimeric protein.

At 5 \textmu{}M concentrations of protein no dimer species were detected. This would indicate that there is an equilibrium between monomer and dimer states which is concentration dependent. Unfortunately higher concentrations could not be investigated due to limited availability of the protein. Optimisation for the transfer
of the dimer into the gas-phase would also need to be further investigated, including solution conditions such as buffer concentration and the methanol content as well as tuning the source to limit in-source CID.

### 4.1.3 The pH Dependence of Zinc Binding

In order to investigate the pH at which YdaE releases its zinc ion, a pH titration was conducted. Samples of 10 \( \mu \text{M} \) YdaE were prepared as described previously, and the protein was diluted with 10 mM ammonium acetate at pH values in the range of 2 to 9. A mass spectrum was acquired for each pH value using the same instrument parameters. The ratio of the *holo* and *apo* protein was calculated from the abundance of each species at each charge-state. The ratio of *holo*- to *apo*-YdaE was normalised so that one zinc bound to the protein would represent 100% when no *apo*-YdaE was detected. This procedure was then repeated for the acquired spectra at each pH value and the data are shown in Figure 4.6. The
Figure 4.6: The pH dependence of zinc binding to YdaE. The ratio of holo- to apo-YdaE was calculated from the abundance of each species from each mass spectrum which was then normalised to 100% for one zinc ion.
displacement of zinc by protons occurs rapidly below a pH value of 6 and zinc is completely displaced by pH 4.

Looking more closely at apo-YdaE below pH 6 shows some unexpected behaviour. After loss of the zinc ion there is also a mass shift which is pH dependent, as shown in the spectra presented in Figure 4.7. At pH 2 the theoretical isotope pattern (▲) that fits closest is that calculated by the addition of 6 hydrogens, i.e. C$_{278}$H$_{457}$N$_{79}$O$_{83}$S$_{7}$. This is what would be expected from the protonation of 6 sites to produce the 6+ charge-state of the ion. At pH 3 the best match was the isotope pattern calculated (●) for C$_{278}$H$_{458}$N$_{79}$O$_{83}$S$_{7}$. This corresponds to an increase in mass of 1 Da and is seen for all charge-states. For the spectrum shown in Figure 4.7b the charge-state is still 6+ so the charge must still come from 6 protons while there is an additional hydrogen which has no charge. At pH 4 the best match was from the isotope pattern calculated (■) for C$_{278}$H$_{455}$N$_{79}$O$_{83}$S$_{7}$ as shown in Figure 4.7c. This corresponds to a decrease of 2 hydrogens, which is surprising. There are 4 cysteines which are in close proximity which could form a disulfide bond after the zinc has been displaced which would result in loss of 2 hydrogens. The formation of disulfide bonds from free thiols occurs readily in air, is pH dependent and favoured at higher pH values, as long as the oxidation reaction is not prevented by zinc coordination. The process of buffer exchange and washing by ultrafiltration can take several hours, long enough for the formation of disulfide bonds. The rate of formation at lower pH is reduced, as shown by the mass difference between the samples at pH 3 and 4. The spectrum shown at pH 2 was prepared previously and the time for disulfide formation is not directly comparable to those at pH 3 and 4, but should be the slowest.

The broad isotope distribution seen for the apo protein below pH 4 would indicate that there is an equilibrium of protonation states as the mono-isotopic peak in all cases matches with that of the protein at pH 4. The holo protein does not show any mass shift with varying pH.

There are chemical tests that can be conducted to detect the presence of free
Figure 4.7: The pH dependent mass shift of apo-YdaE. The mass of YdaE as measured at a) pH 2, theoretical (△) C_{278}H_{457}N_{79}O_{83}S_{7} as expected for the 6+ b) pH 3, theoretical (○) C_{278}H_{458}N_{79}O_{83}S_{7} an extra H for the 6+ and c) pH 4, theoretical (■) C_{278}H_{456}N_{79}O_{83}S_{7} which is 2H less than expected for the 6+ ion.
cysteine and therefore infer the number of disulfide bonds. One such test uses Ellman's Reagents (5,5'-dithiobis(2-nitrobenzoate) (DTNB)) which reacts with free thiol and can be followed using a UV/VIS spectrophotometer at 412 nm.²⁰³

4.1.4 Cadmium Exchange Reactions

Metal exchange with cadmium was also used to probe the stability of zinc binding to YdaE. A 5-fold excess of cadmium acetate was added to mono-Zn-holo-YdaE and incubated for 1 hour at room temperature in pH 7 ammonium acetate buffer, before salt extraction using a gel filtration PD10 column, to remove free zinc and cadmium from solution.

After the metal exchange with cadmium, the mass spectrum for the 5+ charge-state shown in Figure 4.8b was observed. The shift in mass corresponds to displacement of zinc by cadmium. The theoretical isotope pattern calculated for C_{278}H_{454}N_{79}O_{83}S_{7}Cd differed from the experimentally observed spectrum by

Figure 4.8: A comparison of the 5+ charge-state of YdaE a) as purified and b) 1 hour after incubation with a 5× excess of cadmium acetate.
only 3 ppm. The calculated deconvoluted mass from the most abundant isotope peaks for cadmium exchanged holo-YdaE was 6563.05 Da. The mass difference (111.88 Da) from apo-YdaE is close to the average mass of cadmium 112.41 Da. Cadmium has 8 naturally abundant isotopes ranging in mass from 105.9 to 115.9 Da and 6 of these have a natural abundance of > 20 %, which means it has a very broad isotope distribution with the most abundant isotope being 113.90 Da.

Spectra recorded for the dimer also exhibited full exchange of zinc for cadmium, resulting in a neutral mass of 13129.07 Da. The observed mass difference the apo-YdaE dimer and the cadmium bound holo-YdaE dimer is 223.63 Da which shows conclusively that the dimer also exchanges both zinc ions for cadmium.

The stoichiometry of cadmium binding is the same as for zinc with one metal ion per protein, even with incubation of 5× molar excess for 24 h. The exchange is very rapid, so the method used of incubation followed by gel filtration to remove free metal ions from the solution provides little information on the rate of exchange of zinc by cadmium. Using a titration with cadmium without removal of free metal ions from solution, it would be possible to determine the relative amounts of bound zinc and cadmium. The relative strength and the kinetics of the binding could then be investigated using time-resolved ESI-MS.\(^\text{154,204}\)

### 4.2 Ferric-ion Binding Protein A

Iron is an essential nutrient for all organisms. Vertebrates have developed complex iron regulation systems in order to retain their own access to the metal while trying to withhold it from pathogenic organisms. The efficiency of the iron-uptake system in these organisms is related to their ability to cause disease. The iron transporter, ferric-ion binding protein A (FbpA) from Neisseria gonorrhoeae, is up-regulated in response to a low iron environment. FbpA forms the first step in a chain to binding partners FbpB and FbpC, which form an Fe\(^{3+}\) ABC transport system.\(^\text{205–207}\)
The 309 amino acid sequence for FbpA is shown in Figure 4.9. The X-ray crystal structure of recombinant holo-FbpA, which is shown in Figure 4.10, contains a single Fe$^{3+}$ ion bound by His9, Glu57, Tyr195 and Tyr196. The binding residues, highlighted in red in the sequence and labelled on the structure, form the interdomain binding cleft. The counter-ion shown is a phosphate ion, PO$_4^{3-}$, which has also been labelled. The counter-ion can be easily displaced with other anions.

### 4.2.1 Expression and Purification of the FbpA Protein

The iron transport protein, ferric-ion binding protein A (FbpA) from *Neisseria gonorrhoeae* was expressed in *Escherichia coli* TOP10 cell lines using the pTrc99A/FbpA/Ng plasmid. After expression of the protein, the cells were spun down using a centrifuge. The resulting red pellet containing the holo-FbpA was re-suspended in 50 mM HEPES buffer at pH 8 with 1% cetyl trimethyl ammonium bromide (CTAB) for 3 h at 310 K or overnight at room temperature. After centrifugation the supernatant was a strong red colour, containing FbpA.

---

**Figure 4.9:** The 309 amino acid sequence for ferric binding protein A from *Neisseria Gonorrhoeae*. The iron binding residues His9, Glu57, Tyr195 and Tyr196 are highlighted in red.

---

Expression and purification of FbpA was carried out by Dr P. R. Bilton
Figure 4.10: The X-ray crystal structure of FbpA using cartoon representation. The iron, phosphate and binding residues; His9, Glu57, Tyr195 and Tyr196 are labelled in the centre of the protein. This structure was reproduced from the 1D9Y entry from the RCSB protein data bank.\textsuperscript{208}
while the pellet was white. The supernatant was loaded onto a SP-Sepharose cation exchange column. The protein was eluted from the column using a NaCl salt gradient in 10 mM HEPES buffer at pH 8. The purified protein eluted at a salt concentration of 200 mM NaCl.

4.2.2 MS Study of Wild-Type holo-FbpA

4.2.2.1 MS of holo-FbpA after Sample Clean-up using Gel Filtration

The purified protein was buffer exchanged to remove the HEPES buffer and to reduce the salt content using 10 mM ammonium acetate at pH 8 with a gel filtration (Sephadex™G-25M PD10) column, resulting in an approximate protein concentration of 80 μM. This solution was diluted four times in water and methanol (90:10 v/v) containing 10 mM ammonium acetate at pH 8.

Mass spectra of ferric-ion binding protein A (FbpA) were obtained using the same acquisition parameters as for carbonic anhydrase II. Spectra were acquired from 32 scans with a capillary voltage of 150, 220 and 269 V and externally calibrated using CsTFHA clusters (Section 2.5.2).

The formation of buffer adducts to FbpA can severely affect the ionisation efficiency and thus the spectral quality of the acquired mass spectrum. Use of a single gel filtration step to buffer exchange from 10 mM HEPES and 200 mM NaCl to 10 mM ammonium acetate at pH 8, proved to be inadequate for removing all of the HEPES buffer, as can be seen from the spectra presented in Figure 4.11.

As with the formation of the carbonic anhydrase II-acetazolamide complex, the capillary voltage was a key parameter for breaking non-covalent interactions. In this case buffer adducts are undesirable, but can be reduced by increasing the capillary voltage in the source. At a capillary voltage of 269 V (see Figure 4.11a) there was sufficient collisional dissociation in the source to completely remove all buffer molecules which bind non-covalently and non-specifically to the protein. On reducing the capillary voltage to 220 V (see Figure 4.11b) it is possible to see that there are HEPES buffer (238.10 Da, C₆H₁₈N₂O₄S) molecules which bind to
Figure 4.11: An expansion of the 12+ charge-state from the spectra acquired for holo-FbpA after gel filtration was used to exchange the buffer. Spectra were recorded at capillary voltages of a) 269 V, b) 220 V and c) 150 V. The HEPES buffer forms adducts with the protein at lower capillary voltages reducing the ionisation efficiency. However, phosphate binding is preserved. Inset shows the full mass spectrum that was obtained in each case.
the protein during ionisation, as indicated by 238 Da mass shifts with as many as 3 adducts for the 12+ charge-state. It is also apparent that the lower charge-states suffer from more adducts than higher charge-states which must be a result of charge neutralisation of a more highly charged protein by the buffer. While the formation of buffer adducts is undesirable, by reducing the capillary voltage it has been possible to observe the phosphate (94.95 Da, PO$_4^{3-}$) counter-ion, that was in the crystal structure, bound to the protein. In the spectrum shown in Figure 4.11c the observed mass shift is 98 Da. Since the overall charge remains the same, an additional 3 protons bind to the protein along with the phosphate anion in order to maintain the same charge. At the capillary voltage of 150 V, although the number of buffer adducts increases, all the observed protein peaks contain the phosphate counter ion.

4.2.2.2 MS of holo-FbpA after Sample Clean-up using Ultrafiltration

Clearly an improved sample clean-up step was needed to remove the HEPES and NaCl salt from the protein prior to mass spectrometric analysis. Ultrafiltration was chosen as an alternative clean-up step to the gel filtration column. A 10K MWCO ultrafiltration centrifugal filter was used to buffer exchange and desalt the protein into 10 mM ammonium acetate at pH 8. The protein was diluted 10 fold with ammonium acetate buffer and 10 % methanol was added prior to ESI analysis.

Figure 4.12 shows the mass spectrum obtained for holo-FbpA after ultrafiltration clean-up. An acquisition of 100 scans with the capillary voltage set to 150 V, together with a data set of 1 M (2$^{20}$) was used to record the transient. External calibration was performed using CsTFHA clusters. The spectrum shows that ultrafiltration was far more effective than gel filtration for exchanging the HEPES with ammonium acetate. At the capillary voltage of 150 V, no HEPES buffer adducts are seen and almost all of the protein has the phosphate ion bound to holo-FbpA. The 13+ and 12+ are the most abundant charge-states using 10 mM
ammonium acetate at pH 8. An expansion of the spectrum for the 12+ charge-state is shown in the inset together with the theoretical isotope pattern (▲) calculated for the elemental formula C$_{1508}$H$_{2427}$N$_{423}$O$_{450}$SFeP, which gave a very good match. The charge on the protein was provided by 12 additional protons indicating that the charge on the phosphate ion and iron neutralise each other, and that the overall charge is provided by the protons. Analyses of spectra with and without the phosphate adduct are consistent with the iron having a 3+ charge, Fe$^{3+}$. For the most abundant isotopic peak at 2816.75 m/z a mass accuracy of 4 ppm was obtained with a resolving power of 210,000. This is a result of the significant improvement in the spectrum once the protein had been sufficiently cleaned-up.

The calculated deconvoluted mass of apo-FbpA was 33639.02 Da. The calculated deconvoluted mass of holo-FbpA was 33693.90 Da without the phosphate adduct.
ion, and 33788.97 Da with phosphate bound. The mass difference from the \textit{holo} to the \textit{apo} protein is 54.88 Da which is very close to the average atomic mass of iron (55.85 Da). The mass difference between the \textit{apo-} and \textit{holo} protein with iron and the phosphate bound is 149.95 Da, which is close to the calculated average mass for iron and phosphate (150.82 Da).

\subsection*{4.2.3 Reconstitution and Analysis of $^{57}$Fe-FbpA}

It has been shown in the literature that crystals of FbpA can bind multi-metal clusters in the binding domain and that phosphate can be exchanged for other anions.\cite{207,209-212} Mass spectrometry was used to investigate whether such metal clusters can be detected from the solution phase. An enriched stable isotope of iron (95.3 \% $^{57}$Fe) was chosen in order to observe the incorporation of the metal into the protein and to determine whether sufficient mass accuracy could be obtained to detect the mass shift of the enriched iron compared to the natural iron abundance of \textit{holo}-FbpA.

Repeated ultrafiltration was employed to prepare \textit{apo}-FbpA (colourless) from \textit{holo}-FbpA using sodium citrate (250 mM, pH 4.5) followed by exchange with 50 mM HEPES buffer (pH 8). The labelled iron oxide ($^{57}$Fe$_2$O$_3$, 1 mg, 95.3 \%) was dissolved using conc. HCl acid. Sodium citrate was added and the pH raised using NaOH (1 M) until the citrate dissolved forming a solution of iron citrate. A 3-fold molar excess of iron citrate was added to the \textit{apo}-FbpA (50 mM HEPES buffer, pH 8). The colourless protein solution turned red showing incorporation of the iron into the protein. Desalting and buffer exchange to ammonium acetate (10 mM, pH 8) was performed using a 10 kDa MWCO ultrafiltration centrifugal filter (Section 2.2.2.1).

The reconstituted $^{57}$Fe-FbpA was diluted 10-fold with ammonium acetate buffer and 10 \% methanol was added for ESI analysis. Spectra were acquired from 100 scans with a 1 M (2\textsuperscript{20}) data set at capillary voltages of 269 and 150 V using the same acquisition parameters as for carbonic anhydrase II. The spectra
were externally calibrated using CsTFHA clusters (Section 2.5.2). Figure 4.13 shows the spectra acquired at the two capillary voltages, 269 and 150 V. There is a relatively low abundance of the holo-FbpA. The apo protein remains the dominant species. At 269 V, $^{57}$Fe-holo-FbpA protein is seen at about half of the abundance of the apo protein peak. At this capillary voltage no counter-ion is bound to $^{57}$Fe-holo-FbpA. However, a low abundant 2-iron species along with a citrate anion (189.10 Da, $C_6H_5O_7^{3-}$) as the counter ion is observed. This must be particularly stable, as all other non-covalent complexes are dissociated at this voltage. Reducing the capillary voltage to 150 V, reduces the collisional energy allowing more non-covalent complexes to be transferred into the mass spectrometer. The holo-FbpA peak almost entirely disappears at this lower capillary voltage, and many higher mass peaks containing iron are observed. The citrate counter ion is seen to bind to both the apo protein and the holo protein. Up to two iron and two citrate ions can be seen to bind to the protein. However the abundance of each species is low when compared to the apo protein. This sample was extensively buffer-exchanged using ultrafiltration to remove free ions and HEPES buffer which should have removed all non-specific binding. The iron-clusters observed in the crystal structure contain Fe$_3$O units. However, no such oxo-metal clusters were observed in these mass spectra.

In order to conclusively confirm the uptake of $^{57}$Fe theoretical isotope patterns were calculated for spectrum Figure 4.13a, assigned to the 12+ charge-state of the $^{57}$Fe-holo-FbpA at 2808.66 m/z. Figure 4.14 shows an expansion of this peak along with the simulated isotope patterns. The theoretical isotope patterns (▲) were first calculated for $C_{1508}H_{2424}N_{423}O_{446}SFe$ and the natural abundance iron. A new element, containing the isotope abundances of enriched $^{57}$Fe (Table 4.1), was added to the periodic table used for calculating the isotope distribution. The isotope pattern was then re-calculated (●) using this user defined element with the formula $C_{1508}H_{2424}N_{423}O_{446}S^{57}$Fe. The measured spectrum clearly shows the mass shift expected compared to the natural abundance of iron and the simulated
Figure 4.13: An expansion of the 12+ charge-state from the spectra acquired after reconstitution of apo-FbpA with $^{57}$Fe. Capillary voltage a) 269 V and b) 150 V. The assigned peaks are labelled with reference to the 12+ charge-state of the apo-FbpA peak. Inset shows the full mass spectrum that was obtained in each case.
Figure 4.14: The 12+ charge-state of holo-FbpA after reconstitution with $^{57}$Fe citrate. The theoretical isotope patterns are shown both for the natural abundance of iron (▲) calculated from $C_{1508}H_{2424}N_{423}O_{446}SFe$, and for enriched $^{57}$-Fe (●).

Table 4.1: The isotopic composition of enriched $^{57}$Fe.

<table>
<thead>
<tr>
<th>Isotopic mass /Da</th>
<th>Abundance /%</th>
</tr>
</thead>
<tbody>
<tr>
<td>53.9396148</td>
<td>0.01</td>
</tr>
<tr>
<td>55.9349421</td>
<td>2.70</td>
</tr>
<tr>
<td>56.9353987</td>
<td>95.30</td>
</tr>
<tr>
<td>57.9332805</td>
<td>1.99</td>
</tr>
</tbody>
</table>
pattern using $^{57}$Fe matches very closely. The mass accuracy at 2808.66 m/z was 2 ppm.

### 4.3 Summary

Ammonium acetate buffer was used successfully as a suitable medium to study the metalloproteins YdaE and FbpA, providing a near-neutral pH while still being sufficiently volatile to be compatible for mass spectrometry.

The zinc finger protein, YdaE (6.5 kDa), from *Escherichia coli*, binds one metal atom (oxidation state 2+) even after incubation with excess zinc or cadmium overnight at 293 K. *Apo-* and the *holo*-YdaE both show the formation of a dimer. However, this species, was present in very low abundance at the 10 μM protein concentration used in this study. The pH dependence on zinc binding was determined by measurement of the mass spectrum of YdaE over the pH range 2-8. The zinc atom is released from YdaE below a pH value of 6 and no zinc binding was observed by pH 4. *Apo*-YdaE showed unusual behaviour with the measured mass shifting at different pH. At a pH of 4, the protein is 2 Da lighter than expected which might suggest the formation of a disulfide bridge between the cysteines, which are close in space once the metal has been released. Exchange with cadmium was too rapid to determine kinetic information on the rate of metal exchange.

Ferric-ion binding protein A, FbpA (34 kDa), from *Neisseria gonorrhoeae*, binds only one iron atom, (oxidation state Fe$^{3+}$) per protein when naturally expressed and purified from *Escherichia coli*. The crystal structure shows a phosphate ion, PO$_4^{3-}$, ligated to the iron at the centre of the protein. In order to preserve this non-covalent interaction the *holo* protein with phosphate ligand bound, gentle source conditions were required. Gel filtration to buffer-exchange the protein from HEPES to ammonium acetate proved to be insufficient as many non-specific HEPES adducts were observed in the mass spectrum. However, using repeated washing with ultrafiltration, it was possible to observe the phosphate
adduct of *holo*-FbpA without undesirable HEPES adducts. Reconstitution of *apo*-FbpA (50 mM HEPES, pH 8) with a 3-fold excess of stable isotope enriched iron citrate ($^{57}$Fe, 95.3 %) at 293 K, showed that the uptake of $^{57}$Fe was not complete. The binding of citrate anions ($C_6H_5O_7^{2−}$) to the *apo* and *holo* protein was observed. Although present at very low abundance, a species with two iron atoms bound to a single FbpA protein was observed. High resolution and mass accuracy was achieved for FbpA with the spectra for the reconstituted $^{57}$Fe-*holo*-FbpA showing a clear mass shift over the natural abundance iron, which matched the expected theoretical isotope pattern for a protein containing $^{57}$Fe.
Chapter 5

Superoxide Dismutase (SOD)

Molecular oxygen, $O_2$, is a paramagnetic gas that constitutes about one fifth of the Earth’s atmosphere. The superoxide radical, $O_2^-$, is a frequent intermediate of dioxygen reduction.\(^{213}\) While the superoxide radical can act as both an oxidant and a reductant, neutralisation with other molecules leads to a much more reactive oxidant. Neutralisation with a proton leads to $HO_2^*$ which is a much stronger oxidant than the superoxide radical. However, complexation of the superoxide radical with metal cations is far more damaging, as this initiates a free radical chain reaction, which dramatically amplifies the effect of $O_2^-$.  

Reactions 5.1–5.3 show the spontaneous dismutation of the superoxide radical, which proceeds most rapidly at pH 4.7; the reaction rates fall as the pH increases.

$$HO_2^- + HO_2 \rightarrow O_2 + H_2O_2 \quad (5.1)$$
$$HO_2^- + O_2^- + H^+ \rightarrow O_2 + H_2O_2 \quad (5.2)$$
$$O_2^- + O_2^- + 2H^+ \rightarrow O_2 + H_2O_2 \quad (5.3)$$

Superoxide dismutase (SOD) may be the single most important enzyme which protects biological systems from the toxicity of molecular oxygen.\(^{214}\) This enzyme catalyses the dismutation of superoxide as shown in Reaction 5.3. The focus of the work in this Chapter has involved Zn,CuSOD which is found primarily in the cytosol of eukaryotic cells, and also in chloroplasts and in some bacteria.\(^{213}\) There are two other forms, MnSOD and FeSOD, which have a very high homology
and are believed to have a similar origin. The catalytic activity of Zn,CuSOD is through successive reactions with superoxide causing the reduction of Cu(II) and reoxidation of the Cu(I) intermediate. The formation of hydrogen peroxide is then dealt with by other enzymes, with conversion to O$_2$ + H$_2$O by catalases, and to H$_2$O by peroxidases.

The human ($M_w$ (avg) 15846.44 Da) and bovine ($M_w$ (avg) 15593.23 Da) forms of SOD have remarkably similar sequences, as shown in Figure 5.1. The bovine form has 151 amino acids while the human enzyme has 153. Insertion of a gap of two residues at position 24, results in 82% sequence identity. Using this alignment, all the features of the enzyme are conserved. In the human enzyme, the zinc binding residues are His71, His80, Asp83 with a bridging histidine; His63, links to the copper which binds to His46, His48 and His120.

![Amino acid sequences of bovine and human superoxide dismutase](image)

Figure 5.1: The amino acid sequences of bovine and human superoxide dismutase. Sequence homology is indicated with '*' and sequence gaps with '-'. The N-terminal alanine is acetylated and the two cysteines linked by a disulfide bridge are highlighted in orange. The metal binding residues are highlighted in red.
these residues correspond to Zn binding to His69, His78, Asp81 with His61 bridging to Cu along with His44, His46 and His118. There is one internal disulfide bridge linking Cys57 to Cys146 in the human form and Cys55 to Cys144 for the bovine form. Again, once aligned, this disulfide bridge is also conserved. The N-terminal alanine is also acetylated in both species.

All known SOD enzymes form multimers in their active form and Zn,CuSOD is no exception, forming a homodimer.\textsuperscript{213} The X-ray crystal structure of beSOD, reproduced using the PDB entry 2SOD, is shown in Figure 5.2a.\textsuperscript{218} Similarly the crystal structure for heSOD is shown in Figure 5.2b, reproduced from PDB entry 2C9U.\textsuperscript{219} The structures show that Zn,CuSOD monomers are held together by non-covalent interactions and studies have shown this to be a particularly strong association.\textsuperscript{220} The structure of SOD is dominated by predominant $\beta$-sheets which form a $\beta$-barrel. The zinc ion is hidden, while the copper ion in the active-site is found at the bottom of a deep channel.\textsuperscript{221} The non-metal-binding solvent-accessible histidine residues and the sulfur-containing residues of methionine and cysteine are labelled and will be referred to in the following sections.

Superoxide dismutase is being actively researched upon discovery that 20-25\% of familial amyotrophic lateral sclerosis (FALS), a progressive, lethal motor neuron degenerative disease, have mutations in SOD.\textsuperscript{222} The work described in this Chapter was, however, motivated upon publication of a crystal structure showing cisplatin binding to beSOD, with unusual and controversial chlorido ligands still attached to the platinum (i.e. the ammonia ligands had been lost).\textsuperscript{223}

In this work, FT-ICR mass spectrometry has been employed to study both bovine and human forms of Zn,CuSOD. The anticancer drug, cisplatin (cis-diamminedichlorodiplatinum(II), cis-$[\text{PtCl}_2(\text{NH}_3)_2]$, $M_w$ (avg) 300.04 Da), was incubated with beSOD and heSOD and MS was used to elucidate the stoichiometry of the cisplatin adducts. The coordination sphere of the platinum was probed using $^{15}$N-labelled cisplatin (cis-$[\text{PtCl}_2(^{15}\text{NH}_3)_2]$, $M_w$ (avg) 301.99 Da). Top-down fragmentation of the intact protein was used in order to localise the site of cis-
Figure 5.2: The X-ray crystal structures of dimeric a) bovine and b) human erythrocyte superoxide dismutase. Common features are labelled in each protein including the metal ions, Cu$^{2+}$ and Zn$^{2+}$, non-metal-binding solvent-accessible histidine residues and the sulfur containing residues of methionine and cysteine. Structures produced from entries in the PDB database 2SOD and 2C9U for boSOD and heSOD, respectively.\textsuperscript{218,219}
platin modification to the protein. The challenges that arose during the course of this work were firstly to preserve the metal binding as well as the homodimer of the native, active enzyme into the gas phase. It was also necessary to remove unbound cisplatin which binds non-specifically and reduces the ionisation efficiency during the ESI. Finally it was necessary to obtain sufficient sequence coverage in order to identify the site of platinum modification.

5.1 Study of SOD using Mass Spectrometry

Prior to conducting more complicated and potentially more time-consuming experiments, it was important to establish the identity of the sample. As superoxide dismutase forms a dimer and contains metal ions, denaturing conditions consisting of a high content of organic solvent and low pH, should ionise the protein readily and thus make the protein easily detected. While the literature sequence for the bovine enzyme matched the UniProtKB/Swiss-Prot database entry P00442, there are several conflicting entries for the human enzyme in the literature. The UniProtKB/Swiss-Prot database entry P00441 does, however, match the sequence from the genome. The sequences reproduced in Figure 5.1 were taken from the UniProtKB/Swiss-Prot database. By comparing the theoretical isotopic pattern, calculated from the reported sequence, with the measured isotopic pattern any discrepancies will be apparent.

Using neutral pH along with ammonium acetate buffer, presented the best possibility of keeping the protein in as native-like a configuration as possible for MS analysis. The zinc and copper metal ions along with the non-covalent dimer formation would therefore be preserved and observable in the mass spectrum.

5.1.1 Studying SOD using Denaturing Conditions

Lyophilised bovine (beSOD, 320 μM) and human (heSOD, 320 μM) erythrocyte superoxide dismutase were dissolved in water and aliquots were stored at 253 K. Ultrafiltration was performed on each enzyme in order to desalt and remove any
buffers that may be present from the purification process prior to mass spectrometric analysis. Samples were then diluted 20-fold into 1 : 1 water : methanol containing 1 % acetic acid. The samples were infused using the NanoMate chip-based nanospray robot interfaced to the 12 T Apex Qe mass spectrometer. An acquisition of 100 scans in the mass range of 500 to 3000 m/z was recorded using a 512 k (2^19) data set. External calibration was performed using Agilent ESI tuning mix, as described in Section 2.5.1.

5.1.1.1 MS of beSOD

The mass spectrum acquired from the beSOD sample using the 12 T Apex Qe MS is shown in Figure 5.3. Charge-states are observed from 7+ to 18+, with what appears to be a bimodal distribution. One set consists of charge-states from 10+ to 18+, with the 14+ being the most abundant; this can be attributed to
an unfolded conformation. A second distinct set includes those below 10+, with the 8+ charge-state being the most abundant, which can be attributed to a more tightly folded conformation. The inset in Figure 5.3 shows an expanded region of the 14+ charge-state along with isotopic fitting (▲) calculated from the amino acid sequence. The empirical formula for beSOD calculated from the 151 amino acid sequence shown in Figure 5.1 is C_{672}H_{1076}N_{198}O_{221}S_{4}. The match shown in Figure 5.3 was calculated from C_{672}H_{1083}N_{198}O_{221}S_{4}, which is an increase of only 12 hydrogen atoms. This is consistent with the loss of 2 hydrogens to form a disulfide bond, as reported in the crystal structure and the UniProtKB/Swiss-Prot database, and the gain of 14 protons to provide the observed charge. The mass error between the most abundant peak measured isotope and the calculated value was 1.5 ppm.

5.1.1.2 MS of heSOD

The mass spectrum acquired from the heSOD sample using the 12 T Apex Qe MS is shown in Figure 5.4, and is very similar to that of beSOD. However higher charge-states are observed and the bimodal charge-state distribution is much less apparent. The charge-states from 12+ to 19+ can be attributed to an unfolded conformation, with the 16+ being the most abundant. The low abundance of 9+ and 10+ states is likely due to a more compact conformation than the unfolded protein. The inset in Figure 5.4 shows an expansion of the 16+ charge-state together with the simulated isotope distribution (▲) overlaid. The empirical formula for heSOD, C_{681}H_{1087}N_{203}O_{225}S_{4}, was calculated from the 153 amino acid sequence from UniProtKB/Swiss-Prot database shown in Figure 5.1. The simulated isotope pattern shown was calculated from the formula C_{681}H_{1101}N_{203}O_{225}S_{4}, which represents an increase of 14 hydrogen atoms. As with beSOD there is a loss of 2 hydrogens for the disulfide bond, and a gain of 16 protons to provide the net charge. Phosphate buffer adducts (seen as the addition of H₃PO₄) are also seen in the spectrum. The mass error between the most abundant measured
isotopic peak and the calculated value was 0.3 ppm.

There is an excellent match for both beSOD and heSOD between the measured and the simulated isotopic patterns, which confirms that the sequences shown in Figure 5.1 are correct. The acetylation of the N-terminal alanine and the formation of an internal disulfide bond are also confirmed from the match to the measured mass.

![Mass Spectrum of heSOD in Denaturing Conditions](image)

Figure 5.4: The mass spectrum of heSOD in denaturing conditions (H$_2$O : MeOH : AcOH 49.5 : 49.5 : 1 v/v). The inset shows the 16+ charge-state with the theoretical isotope (△) overlaid, calculated from C$_{681}$H$_{1101}$N$_{203}$O$_{225}$S$_4$. This again is consistent with the loss of 2 hydrogen atoms to form an internal disulfide bond. Peaks due to phosphate buffer adducts are marked with *.

5.1.2 Native-Like Conditions

Once the sequences were identified from the samples sprayed in denaturing conditions, samples were then prepared under native-like conditions. If the native-like conformation is preserved through the ionisation process, the metal ions and the dimer formation can be observed in the gas phase. An aliquot (20 μL) of beSOD
(320 μM) was injected on an HPLC gel filtration column and the sample was eluted using 20 mM ammonium acetate (pH 7) as described in Section 2.2.1.1. A single fraction was collected between the elution time of 9 to 10 min. The sample was then transferred into a syringe and ionised using ESI at a flow rate of 100 μL h⁻¹ using a pneumatic pump, and an ESI voltage of 3000 V. The capillary voltage was set to 50 V to minimise in-source nozzle-skimmer dissociation of non-covalent interactions, as described in Section 3.3.1. Gated dynamic trapping was used to reduce space-charge effects in the ICR cell. All other instrument parameters were similar to those used to record the carbonic anhydrase II spectra, shown in Section 3.2.2.

5.1.2.1 MS of beSOD Homodimer

An acquisition from 2000 to 3500 m/z consisting of 500 scans employing with 1 M (2²⁰) data set was used to accumulate the mass spectrum of beSOD, which is shown in Figure 5.5. An external calibration was performed using CsTFHA clus-

![Figure 5.5: The mass spectrum for the homodimer of beSOD sprayed from 20 mM ammonium acetate (pH 7). The zinc and copper metal ions in the dimer are preserved in the gas phase. The inset shows an expansion of the 11+ charge-state and the simulated isotope pattern (▲), calculated from C_{1344}H_{2151}N_{396}O_{442}S_{8}Cu_{2}Zn_{2}.](image-url)
ters, as described in Section 2.5.2. There are just 3 charge-states observed under these conditions, 10+, 11+, 12+, which is typical of a highly folded native-like conformation. The inset shows an expansion of the 11+ charge-state, which is the most abundant, together with the simulated isotope pattern (▲) overlaid, calculated from $C_{1344}H_{2151}N_{396}O_{442}S_8Cu_2Zn_2$. The mass error between the measured and calculated value for the most abundant isotope peak was 0.5 ppm. This very close match was obtained by doubling the monomer elemental formula, including the loss of 4 hydrogens for the 2 disulfide bonds, the addition of two Cu$^{2+}$, two Zn$^{2+}$ and three protons to make up the charge.

5.1.2.2 MS of heSOD Homodimer

The gel filtration HPLC desalting and buffer exchange procedures employed in some of the other work presented were however unreliable and often not reproducible. The amount of human protein available was much less than that of the bovine sample and so ultrafiltration for heSOD was used as it was far more reliable, and the protein could be concentrated into a small volume rather than diluted as with gel filtration. The sample was washed and buffer exchanged into 10 mM ammonium acetate (pH 7). The sample was diluted 3 in 20 (~10 μM dimer) with 10 mM ammonium acetate containing 10 % methanol to aid desolvation. The same parameters employed for beSOD were used in order to acquire the mass spectrum of heSOD, which is shown in Figure 5.6. Even with the addition of 10 % methanol to heSOD to aid ionisation, the signal and the resulting acquired spectrum was not as good as that for beSOD. This is apparent from the increase in noise and thus a much broader baseline for heSOD. The same charge-states were observed as for beSOD. The inset shows the 11+ state with the simulated isotope pattern overlaid (▲) calculated from $C_{1362}H_{2173}N_{406}O_{450}S_8Cu_2Zn_2$. The mass error between the measured and calculated values for the most abundant isotopic peak was 5 ppm. Again, the measured mass matches twice the empirical formula for the heSOD monomer, minus 4 hydrogens for the two disulfide bridges,
the addition of two Cu$^{2+}$, two Zn$^{2+}$ and three protons to make up the charge.

These results demonstrate that for beSOD and heSOD, it was possible to study their native-like conformations in the gas-phase, shown by the preservation of the homodimer and the retention of the metal ions.

### 5.2 Incubation of SOD with Cisplatin

Cisplatin, cis-[PtCl$_2$(NH$_3$)$_2$], is a widely used anticancer drug. There is interest in understanding not only the mechanism of its cytotoxicity, but also its side-effects and cellular resistance mechanisms. Although the major target site is thought to be DNA,$^{226}$ interactions with proteins are likely to play important biological roles.$^{227-233}$

In aqueous media, cisplatin is usually activated by aquation, the substitution of its chlorido ligands by water. Aqua adducts are more reactive than cisplatin
Chapter 5 — Superoxide Dismutase (SOD)

Figure 5.7: The X-ray crystal structure of beSOD after incubation with cisplatin. The platinum was shown to bind to His19, however, the NH₃ ligands have been lost while the chlorido ligands have been retained. This figure was reproduced from entry 2AEO from the PDB database.

itself. Platinum(II) has a high affinity for sulfur ligands, and strong binding sites on proteins are usually at the sulfur atoms of methionine (thioether) and cysteine (thiol). The nitrogen atoms in the imidazole side-chain of histidine are also potential binding sites. Although methionine sulfur is normally the kinetically-preferred binding site, transfer to more thermodynamically-stable Pt-N coordination has been observed.

In 2006, Calderone et al. reported some remarkable and unexpected results. Their X-ray crystal structure for bovine superoxide dismutase (beSOD) after reaction with cisplatin showed platination of His19, but with both chlorido ligands still bound to Pt and both ammonia ligands absent. Their crystal structure is shown in Figure 5.7. Apart from Ne of His19, the other ligand(s) could not be defined in the electron density map. Such an observation is intriguing and raises
the possibility that reactions of cisplatin with this enzyme can follow an unusual pathway in which ammonia release from platinum is readily promoted, perhaps induced by the high trans effect of protein sulfur-ligands.

In the work presented in the remainder of this Chapter the interaction of cisplatin with both the human and the bovine SOD enzyme from solution has been investigated in order to establish the stoichiometry, to define the coordination sphere of the Pt upon binding and to determine whether MS techniques can be used to locate the site of binding.

5.2.1 Studies of Intact SOD-Cisplatin Adducts

Native-like conditions for MS analysis were used to study the stoichiometry of binding for the SOD-cisplatin adducts and to retain as much of the solution conformation intact in the gas phase. As the chlorido ligands bound to platinum can be exchanged for water in aqueous conditions, it was important to accurately track the co-ordination sphere of the platinum from cisplatin. Two experiments were conducted in parallel; beSOD was incubated with cisplatin and $^{15}$N-labelled cisplatin. Accurate mass measurements of the prospective complexes would indicate a mass shift of 1 Da for each ligand that is retained during binding. If the ammonia ligand has been exchanged for an hydroxyl group, which has a very similar mass (-OH 17.00 Da, -NH$_3$ 17.03 Da), there would be no shift between the two mass spectra. Due to limited amount of sample available, heSOD was incubated only with the unlabelled cisplatin.

5.2.1.1 Sample Preparation

Reactions with cisplatin and $^{15}$N-labelled cisplatin (> 98 % enrichment) were performed with beSOD with a 10-fold molar excess of drug for 16 hours at 310 K. After the reaction, the pH was in the region of 5 to 6. Cisplatin drug in 10-fold excess was allowed to react with heSOD for 18.5 hours at 310 K. The lyophilised heSOD contains phosphate buffered saline salt which is likely to control the pH.
However, the final pH for this sample is unknown due to the small volumes used, which prevented measurement using a pH meter. The samples were stored at 253 K prior to MS analysis. Unbound cisplatin was removed using gel filtration for beSOD and eluted with 20 mM ammonium acetate (pH 7) as described in Section 2.2.1.1. The beSOD fraction collected from the gel filtration column was loaded into a glass syringe and infused without any further preparation. Ultrafiltration was used for heSOD which was washed repeatedly with 10 mM ammonium acetate. The sample of heSOD was diluted 3 in 20 with 10 mM ammonium acetate containing 10 % methanol to aid desolvation. Each sample was introduced into the MS by using pneumatically assisted ESI at a flow rate of 100 μL h⁻¹.

5.2.1.2 beSOD-Cisplatin Adducts under Native-like Conditions

The same instrument parameters and calibration were used to acquire data for the cisplatin treated beSOD samples that had been used for the unmodified samples, as described in Section 5.1.2. Figure 5.8 shows the mass spectrum obtained for the homodimer of beSOD after incubation with the unlabelled cisplatin. The mass spectrum shows that up to 4 adducts can form with beSOD. The monoplatinated 11+ charge-state is the most abundant, while the fourth in the series is of a very low abundance, only just above the noise level. This would indicate that up to 2 cisplatin adducts could be formed with each monomer. The 10+ charge-state is very weak and its position is labelled in the Figure. For each cisplatin adduct there is a loss of one hydrogen, assumed to be from the enzyme, and 1 chloride, indicating that only one of the chlorido ligands has been been displaced and that the two ammonia ligands have been retained, see Figure 5.9a. Comparison with the ¹⁵N-labelled sample showed that there was a 2-Dalton increase in mass for each of the cisplatin adducts. The monoplatinated 11+ charge-state is shown as an example in Figure 5.9b. Simulated isotopic patterns are shown for both the naturally occurring cisplatin adduct (△), calculated from.
C_{1344}H_{2157}Cl_{1}N_{396}O_{442}Pt_{1}S_{8}Cu_{2}Zn_{2}, and the $^{15}$N-labelled cisplatin adduct (*), calculated from $C_{1344}H_{2157}Cl_{1}N_{396}^{15}N_{2}O_{442}Pt_{1}S_{8}Cu_{2}Zn_{2}$. The mass accuracy between the measured and calculated values for the most abundant isotopic peak for the monoplatinated dimer was 6.8 ppm for the sample containing unlabelled cisplatin and 7.6 ppm for that with $^{15}$N-labelled cisplatin.

As stated previously, potential sites for Pt binding are the sulfur atoms of methionine (thioether) and cysteine (thiol) and the nitrogen atoms in the imidazole side-chain of histidine. The X-ray crystal structure shown in either Figure 5.2a or 5.7 indicates that Cys55 and Cys144 form a disulfide bond, Cys6 and Met115 are buried in the core of the protein making them unlikely sites for binding. The most likely binding sites are therefore the imidazole rings of His19 and His41, which are situated at the surface of the protein and are not involved in binding.

Figure 5.8: The mass spectrum of the homodimer of beSOD after reaction with cisplatin for 16 h at 310 K. The adducts are indicated as a modification to the natural beSOD $[M]^{11+}$, where $M=\text{beSOD}_2 + 2\text{Cu}^{2+} + 2\text{Zn}^{2+} + 3\text{H}^+$. Noise spikes are indicated by *.
Figure 5.9: Comparison of the expansions of the monoplatinated 11+ adduct of (beSOD$_2$-PtCl(NH$_3$)$_2$) obtained from reactions of a) natural abundance cisplatin and b) $^{15}$N-labelled cisplatin. The 2-Dalton increase in the isotopic pattern is consistent with the retention of the $^{15}$NH$_3$ ligands after incubation. Simulated isotopic patterns are shown for the naturally occurring cisplatin adduct (▲), calculated from C$_{1344}$H$_{2157}$Cl$_1$N$_{396}$O$_{442}$Pt$_1$S$_8$Cu$_2$Zn$_2$ and the $^{15}$N labelled cisplatin adduct (●), calculated from C$_{1344}$H$_{2157}$Cl$_1$N$_{396}^{15}$N$_2$O$_{442}$Pt$_1$S$_8$Cu$_2$Zn$_2$. 

a)

b)
to the copper or zinc. The X-ray crystal structure shows only a single binding site to His19, suggesting this is the most likely site of modification. The mass spectrum obtained for the homodimer of beSOD does however suggest that there is at least one other strong binding site. Repeating the reaction of beSOD with cisplatin and taking aliquots at incubation times of 16, 48 and 72 h did not show a significant change in the binding ratio, suggesting that an equilibrium is established in a shorter time period than 16 h. However, the relative abundances for the adduct series did not exactly reproduce those shown in Figure 5.8, which could have been due to differing experimental conditions.

5.2.1.3 heSOD-Cisplatin Adducts under Native-Like Conditions

The same instrument parameters were used for the acquisition of spectra from the heSOD enzyme incubated with cisplatin as for the beSOD experiments. Figure 5.10 shows an expansion of the 11+ charge-state showing multiple adducts for up to 4 platinum species. Unlike beSOD, after heSOD is incubated with cisplatin the loss of the second chlorido ligand is apparent showing the aquated species as well as platinum chelation to a second site on the protein, greatly increasing the complexity and heterogeneity of the spectrum. With the presence of further platinum adducts, this complexity is compounded. When the second or further cisplatin adducts are considered, it is possible that the loss of HCl ($\Delta = 35.98$ Da) is due to platinum chelating to a second site on the protein. The mass difference corresponding to the loss of both remaining chlorido ligands and addition of two water ligands to the platinum is similar ($-2\text{Cl} + 2\text{OH} \Delta = 35.93$ Da) and these two species will be indistinguishable at the measured resolution. The increased heterogeneity of the sample can help explain why the signal quality is far reduced for heSOD when compared to beSOD after incubation with cisplatin. These results show that cisplatin does bind to the human form of the enzyme and not just the bovine form which was studied previously. The X-ray structure of heSOD (see Figure 5.2b) shows that as for beSOD, the sulfur-containing residues
Figure 5.10: An expansion of the 11+ region of the mass spectrum obtained for the homodimer of heSOD after reaction with cisplatin for 18.5 h at 310 K. The adducts are indicated as a modification to the homodimer of heSOD [M]^{11+}, where M=[heSOD$_2$ + 2Cu$^{2+}$ + 2Zn$^{2+}$ + 3H$^+$]. Loss of the second chlorido ligand is apparent giving an aquated species, as well as platinum chelation to a second site on the protein, greatly increasing the complexity and heterogeneity of the spectrum. Several peaks due to phosphate buffer adducts are marked with *.
are buried in the core of the protein. The most likely sites of binding are again histidine residues, such as His43 and His110 which are on the surface of the protein and not involved in Zn or Cu binding, are near to the positions of His41 and His19, respectively, of beSOD. If the site of binding is similar to that of be-SOD then the His110 residue, which is located on a loop near the dimer binding interface will be the most likely site of modification. The mass spectrum once again indicates that is it the chlorido ligands that are lost and that the ammonia ligands are retained upon binding to the protein.

5.2.2 Identifying the Sites of SOD-Cisplatin Adducts

Electron capture dissociation (ECD) has been used successfully in order to localise the sites of post-translational modifications on proteins.\textsuperscript{19,21} While platination of a protein is not a naturally occurring process in biology, it will occur in patients receiving platinum-based anticancer treatment. ECD may therefore be used in order to localise the site of modification of a platinum-bound protein, such as the cisplatin-adducts of superoxide dismutase. In order to maximise the ECD fragments, denaturing conditions were used to increase the charge, as the capture process is charge-dependent, as well as to unfold the protein to minimise structural interactions that may prevent fragments from being separated and detected.\textsuperscript{68}

5.2.2.1 Sample Preparation

After ultrafiltration to remove any unbound drug, the incubated cisplatin-SOD was diluted 1 in 100 into denaturing conditions just prior to analysis (1 : 1 water : methanol 0.1 % formic acid). The samples were infused using the NanoMate chip-based nanoESI robot into the 12 T Apex Qe mass spectrometer. Spectra were acquired by accumulation of 100 scans with a 512 k (2\textsuperscript{7}) data set. The most abundant charge-state was isolated using the quadrupole, prior to irradiation with electrons inside the ICR cell. Spectra of the resulting ECD fragments were acquired by accumulation of 500 scans for the unmodified and platinated enzyme.
in order to probe the site of platination.

**5.2.2.2 Fragmentation of beSOD-Cisplatin Adducts using ECD**

The mass spectrum acquired for beSOD-cisplatin under denaturing conditions is shown in Figure 5.11, with the major peaks identified and labelled. This expansion of the most abundant 13+ charge-state, shows that while most of the protein has not been modified, there are up to 2 platinum adducts bound the protein. The second adduct is, however, of very low abundance and can only be identified clearly when that region is further expanded. Under denaturing conditions the loss of a chlorido ligand is apparent. This occurs along with the additional loss of an hydrogen suggesting that platinum is chelated to the protein backbone at an adjacent site to the first ligand substitution. For the di-platinum adduct, there is only loss of one chlorido ligand suggesting that only one binding site is suitable for chelation to platinum.

Isolation of the cisplatin-adduct at around 1220 \( m/z \), and subjecting this species to electron-bombardment for 0.015 s led to the ECD fragmentation spectrum shown in Figure 5.12. The same procedure as described in Section 3.4 was used for the peak assignments. Almost all of the identified fragment peaks are from the N-terminus, forming a series of c-type ions from Val15 up to Gln53 with every peptide bond being cleaved with the exception of the N-terminal bond of Pro13, due to the cyclic nature of this residue. The spectrum is almost identical to that from the unmodified protein. Fragmentation stops just before the site of the disulfide bridge between Cys55 and Cys144 and unfortunately no fragments can be identified from within this region of the protein. The many charge reduced species seen in the spectrum can be attributed to cleavage of the disulfide bridge rather than a backbone residue. From the 150 possible peptide bonds, 49 were cleaved and only 67 peaks from a total of 298 were directly assigned to the sequence.

The cisplatin adduct \((+\text{Pt(NH}_3)_2\text{Cl})\) was identified on two new fragment
Figure 5.11: An expansion of the 13+ charge-state of beSOD ([M]^{13+} = [beSOD + 13H^+ - 2H]^{13+}) after incubation with cisplatin for 72 h. Insets show the full mass spectrum (left) as well as the di-platinated adduct (right).
Figure 5.12: The ECD fragmentation spectrum for the mono-cisplatin adduct formed with beSOD isolated at 1220 m/z. The positions of new peaks identified containing cisplatin adducts (see Figure 5.13 for expansions) have been highlighted in green. Peaks labelled with * indicate a charge-reduced species of the form \([M + nH^+ + mH_2O]^{(n-m)+}\).
Figure 5.13: Expansion of the ECD spectrum from a) unmodified beSOD and b) monoplatinated beSOD. New peaks in the monoplatinated beSOD are clearly seen when compared to the same region in the unmodified beSOD. The simulated isotope patterns for the platinated species are shown \[c^{3+}_{19} + \text{Pt(NH}_3\text{)}_2\text{Cl}\] (○) and \[c^{5+}_{46} + \text{Pt(NH}_3\text{)}_2\text{Cl}\] (●).

peaks, the \(c_{19}\) and \(c_{46}\) ions. An expansion and comparison of the same region for the unmodified beSOD ECD spectrum and the monoplatinated beSOD ECD spectrum are shown in Figure 5.13. The simulated isotope patterns were also calculated in order to confirm the identity of the assignments. The pattern for the cisplatin adduct \[c^{3+}_{19} + \text{Pt(NH}_3\text{)}_2\text{Cl}\] (○) calculated from \(C_{84}H_{151}ClN_{27}O_{25}PtS\), and the cisplatin adduct \[c^{5+}_{46} + \text{Pt(NH}_3\text{)}_2\text{Cl}\] (●) calculated from \(C_{206}H_{339}ClN_{61}O_{65}PtS\). The discrepancy between the measured and simulated isotope abundance for each species can be attributed to the low signal-
Chapter 5 — Superoxide Dismutase (SOD)

Figure 5.14: The protein sequence of beSOD with the ECD fragments mapped to indicate bond cleavages.

to-noise for each peak, as the ion abundance is very low especially for the platini-
nated species. The \([c^{3+} + Pt(NH_3)_2Cl]\) fragment overlaps with \(c^2\) which is one of
the most abundant fragments in the spectrum. Since the \(c_{46}\) fragment contains
multiple histidine residues, the cisplatin adduct could be bound to any of the
His19, His41, His44 or His46. The \(c_{19}\) fragment is at the first histidirie which
is the most likely site for the platinum to be bound and agrees with the crys-
tal structure shown in Figure 5.7. There was insufficient signal for the isolated
diplatinum species to yield any fragments and so the second binding site remains
unknown. The fragments have been mapped onto the protein sequence as shown
in Figure 5.14, to indicate the sites of bond cleavages more readily.

5.2.2.3 Fragmentation of heSOD-Cisplatin Adducts Using ECD

The same techniques were used to analyse the human enzyme as for the bovine.
The spectrum acquired of heSOD after incubation with cisplatin for 18.5 h sprayed
from denaturing conditions, is shown in Figure 5.15. The most abundant charge-
state for the heSOD sample was 14+ which has been expanded with the full
spectrum shown in the inset. As with the spectra obtained under native-like
conditions, this spectrum for heSOD is much more heterogeneous when compared
to that of the boSOD sample. The additional peaks have been attributed to
loss of chlorido ligands to chelate to the protein, as well as substitution of the
chlorido ligand by water. As was observed in the case of the bovine enzyme,
there are at least 2 cisplatin adducts of the protein, and possibly a third under
these spray conditions. Unlike beSOD, for the human enzyme the diplinated
Figure 5.15: An expansion of the 14+ charge-state of heSOD ([M]$^{14+}$ = [heSOD + 14H$^+$ - 2H]$^{14+}$) after incubation with cisplatin for 18.5 h. Inset shows the full mass spectrum. Peaks attributed to substitution of a chlorido ligand by water are indicated with *.
adduct appears to lose all of the chlorido ligands, again suggesting chelation to the protein.

Additional complexity due to phosphate (seen as the addition of $\text{H}_3\text{PO}_4$) and oxidation (addition of $\text{O}_2$) are also present. Although speculative, the oxidation of a cysteine thiol, (Cys-SH) to sulfinic acid (Cys-SO$_2$H) is the most likely candidate for this species. Cisplatin is known to produce superoxide radical species \textit{in vivo},$^{239}$ and as long as the enzyme is still active under these conditions, the superoxide dismutase will form hydrogen peroxide as shown in Reactions 5.1–5.3. With the absence of any enzymes to remove the hydrogen peroxide formed, this can then oxidise any free cysteines present. This oxidation adduct is only seen for the human enzyme. While both the human and bovine SOD have a free cysteine at Cys6, only the human form has a free cysteine located at Cys111 and this has been shown to be oxidised by hydrogen peroxide.$^{240}$

The monoplatinated species of the 14+ charge-state was isolated and irradiated with electrons for 0.01 s giving the ECD spectrum shown in Figure 5.16. A total of 85 peaks was found for the heSOD-cisplatin adduct, far fewer peaks than for the bovine equivalent, and, of these, 42 were assigned from the sequence. Only 33 of the 152 possible peptide bonds were cleaved. The fragments once again did not sequence past the disulfide bond between Cys57 and Cys147. The most likely site for platinum to bind is His110, which has a similar spatial position to His19 of boSOD. However, this is within the sequence containing the disulfide bridge and therefore no sequence information is available around that residue. No cisplatin adducts could be identified for any of the fragment ions and no new peaks were present when compared to the fragments formed from the unmodified heSOD. This suggests either that platinum does not bind in the N-terminal region, unlike boSOD, or that there were simply insufficient ions to produce detectable fragments of interest.

The identified fragments were mapped onto the protein sequence for heSOD, as shown in Figure 5.17.
Figure 5.16: The ECD fragmentation spectrum for the mono-cisplatin adduct formed with heSOD isolated at 1149 m/z. Peaks labelled with * indicate a charge-reduced species of the form \([M + nH^+ + nH^-]^{(n-m)+}\).
5.3 Summary

The bovine (15.6 kDa) and human (15.8 kDa) forms of superoxide dismutase have been studied under denaturing and native-like conditions. Using denaturing conditions, a 2-Dalton shift was easily detected for both boSOD and heSOD, which was attributed to loss of 2 hydrogens from the formation of an internal disulfide bridge. Ammonium acetate was used successfully to allow the spraying of boSOD and heSOD in native-like conditions. The binding of the zinc and copper metal ions to the protein, as well as the formation of a homodimer (~32 kDa) found in solution were preserved into the mass spectrometer using electrospray ionisation.

The publication of an unusual and controversial X-ray crystal structure showing a cisplatin adduct on boSOD led to mass spectrometric investigations between cisplatin and boSOD. The human form of the enzyme was also investigated. A 10-fold excess of cisplatin for 16 h at 310 K was incubated with boSOD. Gel filtration chromatography was used with elution of 20 mM ammonium acetate (pH 7) in order to remove unbound cisplatin from solution prior to ESI-FT-ICR MS analysis. A number of species were observed including unmodified boSOD along with 4 cisplatin adducts, with the mono-platinated species being the most prominent. Use of $^{15}$N-labelled cisplatin under identical conditions to the unlabelled cisplatin showed unambiguously that the ammonia ligands were retained after binding to the protein, in direct contradiction to the published X-ray crystal structure, but supporting established cisplatin solution behaviour. Studies with heSOD, which was incubated with cisplatin for 18.5 h showed a similar binding
pattern; however, the adduct species observed were far more heterogeneous. The increased complexity was attributed to phosphate buffer adducts and loss or substitution of the remaining chlorido ligand. Although at a low level, oxidation of the protein was also seen. This has been attributed to the formation of sulfenic acid at Cys111 on one of the monomers, due to formation of superoxide by cisplatin, which SOD then dismuted into hydrogen peroxide.

Denaturing conditions were combined with ECD in order to determine whether the site of cisplatin-modification could be localised. The ECD fragmentation spectra of the isolated mono-platinum adduct of beSOD at 1220 m/z showed two new peaks that were not present in the unmodified spectrum. It was confirmed that these peaks were from the cisplatin adduct, with His19 being the most likely site for modification, consistent with the detection of a peak due to [c19 + Pt(NH3)2Cl]. This matches the site found in the X-ray crystal structure. Fragmentation of the heSOD-cisplatin adduct did not localise the site of modification, which is mostly likely to be in the region containing the disulfide bridge. A reducing agent such as tris(2-carboxyethyl)phosphine hydrochloride (TCEP) could be used to reduce the disulfide bond in order to increase the sequence coverage. Using this reagent would hopefully not affect the bound cisplatin, as other reducing agents typically use a thiol group that may cause displacement of platinum from the protein by preferentially binding to the reducing agent. Mass spectrometry could be performed on crystals which have a solved structure which would help identify the species being analysed. The dissolution of the crystals and required sample clean-up may change the system being studied by electrospray. Using a suitable matrix, MALDI-MS might be possible on crystals directly.

This is not the only example where MS has been used to identify the platination site of a protein. Khalaila et al. used a bottom-up approach in order to investigate the binding of cisplatin to transferrin. While the authors successfully identified a platinated peptide from the trypsin digest, they do not show the MS/MS data that they claimed localised the modification to Thr457. Use of a
bottom-up approach could lead to the migration of cisplatin to a more favourable site once the protein has been digested, making the core sequence of the protein accessible when it may not have been under native conditions. A bottom-up approach, using trypsin, was attempted in the work reported here for both beSOD and heSOD along with the cisplatin-incubated samples, but resulted in very few fragments, most likely due to the presence of the disulfide bridge. The metal ions were an additional problem, as EDTA was needed to remove them prior to digestion, since they also stabilise protein structure. While only ECD results are presented, CID fragmentation was also investigated. No peaks in the CID spectra could be identified as containing, and therefore localising, the cisplatin adduct. This is the first time that a top-down approach using ECD has been used to identify the site of modification of a protein by cisplatin.
References


47. Nemes, P.; Goyal, S.; Vertes, A., Conformational and noncovalent complex-
ation changes in proteins during electrospray ionization. *Anal Chem* 2008,
80, 387–95. (Cited on pages 7 and 28.)

48. Cech, N. B.; Enke, C. G., Practical implications of some recent studies in
(Cited on page 7.)

49. Karas, M.; Bahr, U.; Dülcks, T., Nano-electrospray ionization mass spec-
trometry: Addressing analytical problems beyond routine. *Fresenius J Anal

50. Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne,
M. T.; Burke, P. V.; Kwast, K. E.; Kelleher, N. L., Top-down proteomics on
a chromatographic time scale using linear ion trap Fourier transform hybrid

31, 1325–1337. (Cited on pages 8 and 51.)

52. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion
cyclotron resonance mass spectrometry: A primer. *Mass Spectrom Rev*
1998, 17, 1–35. (Cited on pages 8, 12 and 51.)

53. Marshall, A. G.; Hendrickson, C. L., Fourier transform ion cyclotron res-
one detection: Principles and experimental configurations. *Int J Mass
Spectrum* 2002, 215, 59–75. (Cited on page 8.)

transform ion cyclotron resonance (FT-ICR) spectroscopy. *J Mass Spectrom*

55. Marshall, A. G., Milestones in Fourier transform ion cyclotron resonance


89. Håkansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L., Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield comple-


97. Konermann, L.; Douglas, D. J., Acid-induced unfolding of cytochrome c at different methanol concentrations: Electrospray ionization mass spectrom-


112. Wood, T. D.; Chorush, R. A.; Wampler, F. M.; Little, D. P.; O’Connor, P. B.; McLafferty, F. W., Gas-phase folding and unfolding of cytochrome c
cations. Proc Natl Acad Sci U S A 1995, 92, 2451–2454. (Cited on pages 30 and 64.)


120. Taylor, P. K.; Kurtz, D. M.; Amster, I. J., Electrospray ionization Fourier


170


171


174


173. Wu, Q. Y.; Gao, J. M.; JosephMcCarthy, D.; Sigal, G. B.; Bruce, J. E.; Whitesides, G. M.; Smith, R. D., Carbonic anhydrase-inhibitor binding:


178


192. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W., Top down characterization of larger proteins (45


208. Mcree, D. E.; Bruns, C. M.; Williams, P. A.; Mietzner, T. A.; Nunn, R., Structural basis of iron uptake in the pathogen *Neisseria gonorrhoeae*. *To be Published 1999*. (Cited on pages 112 and 113.)


231. Mandal, R.; Kalke, R.; Li, X. F., Interaction of oxaliplatin, cisplatin, and


238. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W., Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other

185


Appendix A

Source Code

Included in this appendix is the source code for a program written in order to calculate the elemental formula of proteins and oligonucleotides (work not presented in this Thesis). The code (shown in Listing A.1) is written in the C programming language and has been compiled under GNU/Linux using GCC and Microsoft Windows using MinGW. The code is presented as-is with no in-line comments with the intention that the code is self-explanatory.

A brief explanation of the code design is as follows; an array is initialised which contains the numbers of carbon, hydrogen, nitrogen, oxygen and sulfur atoms for proteins while the last element in the array is phosphor, in place of sulfur, is used for oligonucleotides. The 20 amino acids are defined in a 20 by 5 array while the nucleotide bases are defined in a 4 by 5 array. The calculated elemental formula is held in array called ‘protein’, named for legacy reasons before expanding the code to also calculate elemental formulae for sequences of oligonucleotides. The sequence is written in a file held on disk which is given as an argument to the program, ‘-p’ for proteins and ‘-n’ for oligonucleotides. To make the code simple, the sequence is read into a array up to 1024 characters long and the filename can be up to 60 characters long. The N- and C- terminus (or 5’- and 3’- for oligonucleotides) are separated by ‘-’ and these are written using the elements explicitly with lower case letters in order to distinguish them from the sequence units which are all uppercase with which the elements have been defined in an array at the
beginning. The terminus needs to be defined as these are common modifications to the sequence. The protein and oligonucleotide formula calculations are computed separately due to the different outputs that are printed to the terminal. The output was formatted specifically for insertion into Xmass for calculation of the theoretical isotope distribution.

The following contains usage and output examples for both protein and oligonucleotide sequences.

Usage for a protein sequence:

```
$ ./pem -p carbanh.seq
```

Protein output:

```
Protein sequence = ch3co-SHHWGYKHKNGPEHWHKDFPIANGERQSPV
DIDTKAVVQDPALKPLALVGEATSRMVNNNGSHFNVEYDDSDQDKAVLKDGPLTG
TYRLVQFHFWGSSDQGSEHTVDRKKYAAELHLVHWNTKYGDFGTAAQQPDGLA
VVGVFLKVDANPQLKVDLALDSIKTKGSTDFPFPDGSLPLLPNVLYWTYPGS
LTTPPLLESVTWIVLKEPISVSSQMLKFRTLNFAEGEPELLMLANWRPAQPLK
NRQVRGFPK-oh
Length of protein sequence is 259 amino acids.
Elemental composition = C 1312 H 1996 N 358 O 384 S 3
```

Usage for an oligonucleotide sequence:

```
$ ./pem -n 47TR.seq
```

Oligonucleotide output:

```
Oligonucleotide sequence = oh-GCCTAACCACGTGGTGCGTACGAG
CTCAGGCGCATGCGTAGGCGGG-oh
Length of oligonucleotide sequence is 47 nucleotides.
Elemental composition = C 455 H 572 N 184 O 279 P 46
```
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

int main(int argc, char *argv[]) {
    char acids[] = "ARNDCQEGHILKMFPSTWYV";
    int adef[20][5] = {{3, 5, 1, 1, 0},
                       {6, 12, 4, 1, 0},
                       {4, 6, 2, 2, 0},
                       {4, 5, 1, 3, 0},
                       {3, 5, 1, 1, 1},
                       {5, 8, 2, 2, 0},
                       {5, 7, 1, 3, 0},
                       {2, 3, 1, 1, 0},
                       {6, 7, 3, 1, 0},
                       {6, 11, 1, 1, 0},
                       {6, 11, 1, 1, 0},
                       {6, 12, 2, 1, 0},
                       {5, 9, 1, 1, 1},
                       {9, 9, 1, 1, 0},
                       {5, 7, 1, 1, 0},
                       {3, 5, 1, 2, 0},
                       {4, 7, 1, 2, 0},
                       {11, 10, 2, 1, 0},
                       {9, 9, 1, 2, 0},
                       {5, 9, 1, 1, 0}};

    int protein[5] = {0, 0, 0, 0, 0};
    char sequence[1024], nterm[10], output[60], outsulfur[5], file[256];

    char nucleo[] = "AGTC";
    int ondef[4][5] = {{10, 12, 5, 5, 1},
                       {10, 12, 5, 6, 1},
                       {10, 13, 2, 7, 1},
                       {9, 12, 3, 6, 1}};

    int length = 0, i, j, begin = 0, end = 0;
    FILE *INPUT;

    if (argc > 1) {
        if ((strcmp(argv[1], "-p") == 0) || (strcmp(argv[1], "-n") == 0)) {
            sprintf(file, "%s", argv[2]);
        } else {
            printf("Please provide a filename for the protein sequence.\n"");
            exit(EXIT_SUCCESS);
        }
    }

    INPUT = fopen(file, "r");

    if (INPUT == (FILE *) 0) {
        printf("Failed to open file\n");
        exit(EXIT_FAILURE);
    }

    189
fscanf (INPUT, "%s", sequence);

length=strlen(sequence);

for (i=0; i<length; i++)
{
    for (j=0; j<20; j++)
    {
        if (sequence[i]==', - ')
        {
            if (begin==0)
            {
                begin=i;
            }
            else if (begin>0)
            {
                end=i+1;
            }
        }
    }
}

if ((begin==0)&&(end==0))
{
    printf("Protein u is u incomplete\n");
    exit(EXIT_SUCCESS);
}

for (i=0; i<begin; i++)
ncterm[i]=sequence[i];

ncterm[i]='0';

if (strcmp(ncterm, "h")==0)
{
    protein[0]+=0;
    protein[1]+=1;
    protein[2]+=0;
    protein[3]+=0;
    protein[4]+=0;
}
else if (strcmp(ncterm, "hco")==0)
{
    protein[0]+=1;
    protein[1]+=1;
    protein[2]+=0;
    protein[3]+=1;
    protein[4]+=0;
}
else if (strcmp(ncterm, "h3")==0)
{
    protein[0]+=0;
    protein[1]+=1;
    protein[2]+=0;
    protein[3]+=1;
    protein[4]+=0;
}
else if (strcmp(ncterm, "ch3co")==0)
{
    protein[0]+=2;
    protein[1]+=3;
    protein[2]+=0;
    protein[3]+=1;
protein[4] += 0;
}
else if (strcmp(ncterm, "oh") == 0)
{
    protein[0] += 0;
    protein[1] += 1;
    protein[2] += 0;
    protein[3] += 1;
    protein[4] += 0;
}
else
{
    printf("N-terminus not recognised\n");
    exit(EXIT_SUCCESS);
}

for (i = 0; i < (length - end); i++)
{
    nterm[i] = sequence[i + end];
}

nc term[i] = 'O';

if (strcmp(nc term, "oh") == 0)
{
    protein[0] += 0;
    protein[1] += 1;
    protein[2] += 0;
    protein[3] += 1;
    protein[4] += 0;
}
else if (strcmp(nc term, "nh2") == 0)
{
    protein[0] += 0;
    protein[1] += 2;
    protein[2] += 1;
    protein[3] += 0;
    protein[4] += 0;
}
else
{
    printf("C-terminus not recognised\n");
    exit(EXIT_SUCCESS);
}

printf("%s", argv[1]);

if (strcmp(argv[1], "-p") == 0)
{
    for (i = begin; i < end; i++)
    {
        for (j = 0; j < 20; j++)
        {
            if (sequence[i] == acids[j])
            {
                protein[0] += aadef[j][0];
                protein[1] += aadef[j][1];
                protein[2] += aadef[j][2];
                protein[3] += aadef[j][3];
                protein[4] += aadef[j][4];
            }
        }
    }
}
else

191
printf("\n");

printf("Protein\ sequence=\%s\n",sequence);
printf("Length\ of\ protein\ sequence\ is\ %d\ amino\ acids.\n",(end-
begin-2));
	sprintf(output,"Elemental\ composition=\%C_0\%H_0\%N_0\%O_0\%d", 
protein[0],protein[1],protein[2],protein[3]);

if (protein[4]!=0)
{
    sprintf(outsulfur,"\%S\%d",protein[4]);
    strcat(output,outsulfur);
}
printf("%s\n",output);

if (strcmp(argv[1],"-n")==0)
{
    protein[1]=1;
    protein[3]=4;
    protein[4]=1;
    for (i=begin; i<end; i++)
    {
        for (j=0;j<4;j++)
        {
            if (sequence[i]==nucleo[j])
            {
                protein[0]+=ondef[j][0];
                protein[1]+=ondef[j][1];
                protein[2]+=ondef[j][2];
                protein[3]+=ondef[j][3];
                protein[4]+=ondef[j][4];
            }
        }
    }
}
printf("\n");

printf("Oligonucleotide\ sequence=\%s\n",sequence);
printf("Length\ of\ oligonucleotide\ sequence\ is\ %d\ nucleotides.\n
",(end-begin-2));
	sprintf(output,"Elemental\ composition=\%C_0\%H_0\%N_0\%d\%P_0\%d", 
protein[0],protein[1],protein[2],protein[3],protein[4]);

printf("%s\n",output);
}

fclose(INPUT);
exit(EXIT_SUCCESS);
Appendix B

Publication Reprints
Platination of superoxide dismutase with cisplatin: tracking the ammonia ligands using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS)

Stefan K. Weidt, C. Logan Mackay, Pat R. R. Langridge-Smith and Peter J. Sadler*

Received (in Cambridge, UK) 7th February 2007, Accepted 23rd March 2007
First published as an Advance Article on the web 5th April 2007
DOI: 10.1039/b701903c

The high mass accuracy of FT-ICR MS combined with 15N-labelling shows that mono- and di-platinated products from the reaction of erythrocyte superoxide dismutase with the anticancer drug cisplatin in solution retain their ammine ligands, in contrast to a recent X-ray crystallographic study.

Cisplatin, cis-[PtCl₂(NH₃)₂], is a widely used anticancer drug. There is interest in understanding not only the mechanism of its cytotoxicity, but also its side-effects and cellular resistance mechanisms. Although the major target site is thought to be DNA, interactions with proteins are likely to play important biological roles.

In aqueous media, cisplatin is usually activated by aquation, the substitution of its chlorido ligands by water. Aqua adducts are more reactive than cisplatin itself. Platinum(II) has a high affinity for sulfur ligands, and strong binding sites on proteins are usually at the sulfur atoms of methionine (thioether) and cysteine (thiol). The nitrogen atoms in the imidazole side-chain of histidine are also potential binding sites. Although methionine S is normally the kinetically-preferred binding site, transfer to the more thermodynamically-stable Pt–N coordination has been observed.

A recent paper by Calderone et al. reported some remarkable and unexpected results. Their X-ray crystal structure of bovine superoxide dismutase (beSOD) after reaction with cisplatin showed platination of His19 but with both chlorido ligands still bound to Pt and both ammine ligands absent (Fig. 1). Apart from the Nε of His19, the other ligand(s) could not be defined in the electron density map. Such an observation is intriguing and raises the possibility that reactions of cisplatin with this enzyme can follow an unusual pathway in which ammonia release from platinum is readily promoted, perhaps induced by the high trans effect of protein S-ligands. We report here solution studies of the platination of bovine erythrocyte superoxide dismutase by cisplatin using FT-ICR MS. The high resolving power has allowed us to define the coordination sphere of bound Pt and to compare our findings with the reported X-ray crystal structure.

First we recorded the FT-ICR MS mass spectrum of beSOD (Mₚ 31432 Da). The instrument was tuned so that the source region could transfer the ions as gently as possible into the high vacuum system. The gentle conditions from both the buffered solution and the desolvation enabled the beSOD non-covalent dimer to be observed and the bound copper and zinc ions to be retained. Fig. S1a shows the isotopic modelling of the beSOD 11+ ion, which has a very close match over all the isotoponiers giving peaks above the noise level. The mass error from the isotopic fit is
0.7 ppm, indicative of a good fit to the expected elemental composition. Fig. S1b shows that only three charge-states are seen for the dimer, the predominant charges being 11+ along with 10+, while the 12+ ion is of a very low abundance.

Copper-zinc superoxide dismutase (133 μM) was then incubated with a 10-fold molar excess of cisplatin or 15N-labelled cisplatin (>98% enrichment) in aqueous solution at 310 K for 16 h. The samples were then stored for up to 32 days at 253 K until they were ready to be thawed, injected onto the gel filtration column as previously described, and analysed.

After reaction of beSOD with cisplatin,§ the predominant ion observed by FTMS was a monoplatinated 11+ ion (Fig. 2). Peaks for four platinated adducts can be seen, with the fourth being of very low abundance, just above the noise. The 10+ ion is very weak, but does appear to show the same binding pattern with the monoplatinated species being the most abundant. The mass spectrum shows that up to two \( \text{[PtCl(NH}_3\text{)]}^+ \) cisplatin fragments can bind per monomer. The observed mass is consistent with one of the chlorido ligands on each platinum being displaced upon binding, while the two ammine ligands are still present (Fig. 3a).

Increasing the capillary voltage allowed collisionally-induced dissociation within the source region of the homo-dimer, while maintaining the bound platinum complex. This increased the signal-to-noise ratio and reduced the charge on the molecular species making it easier to resolve the isotope pattern. The dissociation of the dimer led to the appearance of peaks from the unmodified beSOD monomer along with peaks from the mono- and di-platinum adducts (Fig. S2b). The increased internal energy also appears to promote dissociation of chloride from platinum. Fig. S2a shows the isotopic modelling of a beSOD monomer with an increased mass of PtN2H4Cl, which is consistent with the binding of a \( \text{[PtCl(NH}_3\text{)]}_2 \) fragment to the protein.

Use of 15N-labelled cisplatin in the reaction led to a clear mass shift of up to two atomic mass units per platinum bound to the protein, further supporting the conclusion that the ammine ligands have not been displaced during the binding of cisplatin to the protein (Fig. 3b).

A summary of the observed masses for the most abundant isotopomers using native ESI of beSOD along with the dimer and monomer peaks arising from reaction with cisplatin is shown in Table 1. The proposed elemental formulae are also listed along with the calculated most abundant mass at the same charge. It can be seen that the mass errors between calculated and observed values (based on external calibration) are all <10 ppm. The peak resolving powers are over 100,000, which meant that isotopic resolution was easily achieved, and this facilitated isotopic modelling for comparison with the proposed elemental formulae. Therefore our MS data suggest that in solution one or two \( \text{[PtCl(NH}_3\text{)]}^+ \) units readily bind to beSOD via loss of one chlorido ligand from cisplatin, but with retention of the two ammine ligands.

Consideration of space-filling models of beSOD (PDB 1Q0E or 2AEO) suggest that the likely binding sites are His19 and His41. The sulfurs of Cys6 (the only free thiol in the protein) and Met115 are buried in the core of the protein and are likely to be inaccessible to cisplatin. Cys55 and Cys144 form a disulfide bond, usually a very weak binding site. Hence the course of the reaction between cisplatin and beSOD in solution over a period of reaction (16 h) at 310 K (body temperature) studied here is more in line with the expected chemical reactivity of cisplatin,13 based on an initial aquation and targeting of imidazole nitrogen in the absence of access to sulfur ligands. The crystals used in the reported study11 of beSOD were obtained from a sample of protein which had reacted with 10 mol equiv. of cisplatin for 2 weeks (under unspecified conditions). Although a single binding site (His19) was detected in the X-ray structure, in solution there is clearly at least

Table 1. The masses of the observed most abundant isotopomer of each major ion and the proposed formula and the calculated masses. The mass error and resolving power of the most abundant mass are also listed.

<table>
<thead>
<tr>
<th>beSOD</th>
<th>Charge</th>
<th>Observed ( m/z^a )</th>
<th>Formula</th>
<th>Calculated ( m/z^a )</th>
<th>Error/ppm</th>
<th>Resolving power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimer, Fig. S1</td>
<td>11+</td>
<td>2858.487</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2858.485</td>
<td>0.7</td>
<td>150,000</td>
</tr>
<tr>
<td>Monomer</td>
<td>6+</td>
<td>2620.281</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2620.277</td>
<td>1.5</td>
<td>220,000</td>
</tr>
<tr>
<td>Dimer + Pt, Fig. 2b</td>
<td>11+</td>
<td>2882.464</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2882.483</td>
<td>6.8</td>
<td>209,000</td>
</tr>
<tr>
<td>Dimer + 2 Pt, Fig. 2c</td>
<td>11+</td>
<td>2906.565</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2906.573</td>
<td>2.7</td>
<td>156,000</td>
</tr>
<tr>
<td>Dimer + 3 Pt, Fig. 2d</td>
<td>11+</td>
<td>2930.563</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2930.572</td>
<td>3.0</td>
<td>114,000</td>
</tr>
<tr>
<td>Dimer + 4 Pt, Fig. 2e</td>
<td>11+</td>
<td>2954.666</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2954.661</td>
<td>1.5</td>
<td>N/A</td>
</tr>
<tr>
<td>Monomer + Pt-Cl</td>
<td>6+</td>
<td>2658.262</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2658.280</td>
<td>6.5</td>
<td>160,000</td>
</tr>
<tr>
<td>Monomer + 1 Pt, Fig. S2</td>
<td>6+</td>
<td>2664.427</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2664.443</td>
<td>5.9</td>
<td>151,000</td>
</tr>
<tr>
<td>Monomer + 2 Pt-Cl</td>
<td>6+</td>
<td>2702.266</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2702.276</td>
<td>3.6</td>
<td>136,000</td>
</tr>
<tr>
<td>Monomer + 2 Pt</td>
<td>6+</td>
<td>2708.257</td>
<td>( \text{C}<em>{1344}\text{H}</em>{2157}\text{N}<em>{396}\text{O}</em>{342}\text{Pt}<em>{5}\text{S}</em>{8}\text{Cu}<em>{2}\text{Zn}</em>{2} )</td>
<td>2708.272</td>
<td>5.7</td>
<td>160,000</td>
</tr>
</tbody>
</table>

a Most abundant isotopomer.
one other strong Pt binding site, possibly His41. It is possible that monoplatinated beSOD crystallized preferentially.

Our work demonstrates the potential of high resolution FT-ICR MS for defining the coordination spheres of protein adducts of metallodrugs in solution. Such adducts can play important roles in mechanisms of action and, in the case of cisplatin, especially in toxic side-effects.

We thank RCUK (Interdisciplinary Research Collaboration in Proteomic Technologies, RASOR), BBSRC (studentship for SW) and SHEFC for support, members of the EC COST Action D39 for stimulating discussions, and Dr Vivienne Munk (Edinburgh), for her gift of natural and $^{15}$N-labelled cisplatin.

Notes and references

‡ To prepare the samples for mass spectrometric analysis and to remove unbound cisplatin and thus prevent formation of further adducts during the electrospray desolvation process, a gel filtration column was used with 20 mM ammonium acetate pH 7 as mobile phase. An aliquot of the sample (20 pL) was injected onto a Tosoh Bioscience TSKgel SuperSW 2000 gel filtration column (4.6 x 30 cm) and eluted with the ammonium acetate buffer. An Agilent 1100 series l-LPLC was used at a flow rate of 350 pL min$^{-1}$ and the UV detector monitored the absorbance at 214 nm.

The major fraction with an elution time of ca. 8 to 12 min with a volume of ca. 1.5 mL was collected (ca. 75 x dilution of the sample).

The gel filtration fraction was analysed using electrospray ionisation at a flow rate of 100 pL h$^{-1}$ and detected by a Bruker Daltonics 9.4 T Apex III FT-ICR MS modified with a heated metal capillary built in-house. 500 scans were accumulated to produce a signal-to-noise ratio of 4 : 1.

§ Use of ammonium-based buffers, although they are compatible with electrospray mass spectrometry, was avoided during the reaction of beSOD with cisplatin to prevent possible exchange with ligands on platinum during this reaction period. Under the conditions used in this study, Fig. 2 clearly shows multiple binding sites. The isotope fitting and $^{15}$N data show that it is the ammine ligands that are retained after the usual activation of cisplatin via aquation.


12 In a recent study of lysozyme platinated with cisplatin the ammine ligands were also retained on bound Pt: A. Casini, G. Mastrobuoni, C. Temperini, C. Gabbiani, S. Francese, G. Moneti, C. T. Supuran, A. Scozzafava and L. Messori, *Chem. Commun.*, 2007, 156-158.
Identification of Clusters from Reactions of Ruthenium Arene Anticancer Complex with Glutathione Using Nanoscale Liquid Chromatography Fourier Transform Ion Cyclotron Mass Spectrometry Combined with $^{18}$O-Labeling

Fuyi Wang, Stefan Weidt, Jingjing Xu, C. Logan Mackay, Pat R. R. Langridge-Smith, and Peter J. Sadler

School of Chemistry, University of Edinburgh, Edinburgh, United Kingdom
Institute of Chemistry, Chinese Academy of Sciences, Beijing, China
Department of Chemistry, University of Warwick, Coventry, United Kingdom

Reactions of the anticancer complex $[(\eta^6\text{-}\text{bip})\text{Ru(en)}\text{Cl}]^+$ (where bip is biphenyl and en is ethylenediamine) with the tripeptide glutathione ($\gamma$-L-Glu-L-Cys-Gly; GSH), the abundant intracellular thiol, in aqueous solution give rise to two ruthenium cluster complexes, which could not be identified by electrospray mass spectrometry (ESI-MS) using a quadrupole mass analyzer. Here we use Fourier transform ion cyclotron mass spectrometry (nanoLC-FT-ICR MS) to identify the clusters separated by nanoscale liquid chromatography as the tetranuclear complex $[(\eta^6\text{-}\text{bip})\text{Ru(GS}O_2\text{)}]_4^{2-}$ (2) and dinuclear complex $[(\eta^6\text{-}\text{bip})\text{Ru(GS}O_2\text{)}_2]^{3-}$ (3) containing glutathione sulfinate ($\text{GS}O_2\text{-}$) ligands. Use of $^{18}$OH$_2$ showed that oxygen from water can readily be incorporated into the oxidized glutathione ligands. These data illustrate the power of high-resolution MS for identifying highly charged multinuclear complexes and elucidating novel reaction pathways for metallodrugs, including ligand-based redox reactions. (J Am Soc Mass Spectrom 2008, 19, 544-549) © 2008 American Society for Mass Spectrometry

Complexes of the second row, group 8 transition-metal ruthenium are of medical interest. Two Ru$^{10}$ complexes are currently undergoing clinical trials as anti-cancer agents, and Ru$^{11}$ arene complexes have shown promising activity in model cancer systems [1]. Ruthenium has seven isotopes, making the mass isotopic pattern of ion peaks of ruthenium-containing compounds characteristic, yet complicated [2-5]. In our previous work [2, 3, 7-9], mass spectra acquired by electrospray ionization (ESI-MS) equipped with a quadrupole mass analyzer allowed unambiguous assignment of singly charged ion peaks of mononuclear ruthenium arene complexes as well as their adducts with amino acids, peptides, and DNA. However, low-resolution ESI-Q MS was unable to identify two multinuclear ruthenium clusters, which arose from the reaction of the ruthenium arene anticancer complex $[(\eta^6\text{-}\text{bip})\text{Ru(}\text{en)}\text{Cl}]^+$ (where bip is biphenyl and en is ethylenediamine) with the tripeptide glutathione ($\gamma$-L-Glu-L-Cys-Gly; GSH) in aqueous solution [3]. Glutathione is present in almost all cells at millimolar concentrations and can detoxify some transition-metal ions.

In this work, we have applied nanoscale liquid chromatography-Fourier transform ion cyclotron mass spectrometry (nanoLC-FTICR MS) to identify these clusters and $^{18}$OH$_2$ derivatization to determine whether oxygen from solvent becomes incorporated into oxidized glutathione found in the products.

Experimental

Materials

$[(\eta^6\text{-}\text{bip})\text{RuCl(}\text{en})][PF_6]$ (1) was synthesized as described elsewhere [10, 11]. Glutathione (GSH, reduced) and disodium hydrogen phosphate were purchased from Sigma (Dorset, UK), sodium dihydrogen phosphate from Aldrich (Dorset, UK), trifluoroacetic acid (TFAH) from Acros (Geel, Belgium) (Andover, MA) $^{18}$O-labeled
Results and Discussion

In unbuffered solution (pH ca. 3) and at 310 K, the ruthenium arene anticancer complex [(η⁶-bip)Ru(en)Cl]⁺ (1) reacted with 10 mol equivalent glutathione (γ-L-Glu-L-Cys-Gly, GSH) to give two di-ruthenium glutathione complexes as the main products (Figure S1, which can be found in the electronic version of this article), of which the di-ruthenium triply-S bridged product centered at 13.99 min has been identified previously by conventional LC-ESI MS and NMR [3]. The ESI-Q MS showed that the fraction centered at 17.25 min contains a di-ruthenium glutathione sulfenate/sulfinate complex (Figures S1 and S2). However, the concentration of multinuclear ruthenium clusters in the fraction eluted from 10.93 to 12.24 min was too low to allow good ESI MS analysis (Figures S1 and S2). The fraction centered at 17.25 min was also collected and...
concentrated for ESI-Q MS and NMR experiments. NMR results (Figure S3) show that only one set of proton resonances is observed for the two biphenyl ligands, suggesting that the two ruthenium centers in this complex are equivalent. No resonances are observed for the protons of the chelated ethylenediamine(ene) ligand, indicating that the en ligands have been displaced from ruthenium by oxidized GSH ligands (Figure S3). The NMR sample was diluted with 1:1 H2O/CH3CN and analyzed by ESI-Q MS with various cone voltages in an attempt to obtain further structural information. However, the low-resolution of the mass spectra did not allow unambiguous identification of the di-ruthenium product either, although several fragmented ions observed with a cone voltage of 50 V (Figure S4) appeared to correspond to the release of one and two sulfinate (GSO3-) ligands from the parent di-ruthenium complexes.

The di-ruthenium product was also detectable in the reaction mixture of complex 1 with GSH under argon and physiologically-relevant conditions (Ru:GSH 0.02:5 mM, in 22 mM NaCl, pH 7 at 310 K for 48 h; Figure S5). Since the glutathione adduct may be involved in the biological mechanism of action of this ruthenium arene anticancer complex, ESI-FT-ICR MS experiments were performed to identify the adducts.

To identify unambiguously the multinuclear ruthenium glutathione product by MS, 5 μL of a reaction mixture of higher concentration of complex 1 with GSH (2:20 mM) in unbuffered aqueous solution was incubated at 310 K for 48 h and separated by a nanoscale C18 reverse phase LC. The eluent was introduced to the ESI-FT-ICR MS using the nanoMate. Two TIC peaks were observed as shown in Figure 1. The partial mass

Figure 3. Mass spectra for fractions corresponding to LC peak 2 (retention times 23.3, 22.6, and 21.6, see Figure 1) from the reaction of [(η6-bip)Ru(en)Cl]2+ and GSH 2:20 mM in water (pH ~ 3) at 310 K for 48 h.

Figure 4. Isotopic models (dots, for which the values of x and y correspond to the m/z value and intensity of the respective isotopic ion peak) and mass spectra (lines) for fragment ions of arene ruthenium sulfinate glutathione adducts: (a) [(η6-bip)Ru(GSO2)2]2+ - Glu - 50 + 4H]2+ ([C20H22O2S1N11Ru112Ru'2]2+) and (b) [(η6-bip)Ru(GSO2)2]2+ - 30 + 10H]2+ ([C20H22O2S1N11Ru112Ru'2]2+). Possible structures for the adducts are shown in Chart 2.
Chart 1. (a) GS(O)H (glutathione sulfenic acid) and (b) GS(O)₂H (glutathione sulfinic acid). As ligands for ruthenium, the sulfur atom is deprotonated and at neutral pH then carries an overall charge of 2− for both sulfenate (GSO) and sulfinate (GSO₂) groups.

spectra for the fractions at specific retention times are shown in Figures 2 and 3, and the full scan spectra for the fractions eluting at 17.2 and 22.6 min in Figures S6 and S7. For the fractions of the first LC peak, a series of doubly charged ions is detected within the range m/z 1050–1110, with a mass difference between the neighboring ion clusters of 16 Da, i.e., an O atom (Figure 2). Corresponding to the doubly charged ion centered at m/z 1083.6502 for the fraction eluting at 17.2 min, a triply charged ion centered at m/z 722.4247 was observed (Figure S6). Similarly, a series of doubly charged ion peaks is detected within the range 900–930 for the fractions of the second LC peak; the difference between the neighboring ion clusters is also 16 Da, indicative of the loss or gain of an O atom between them (Figure 3). The full scan spectrum (Figure S7) for the fraction eluting at 22.6 min shows a triply charged ion centered at m/z 1217.1663 assignable to the dimer of the doubly charged ion centered at m/z 913.1206.

FT-ICR MS has previously been applied to the unambiguous determination of the oxidation state of iron in a metalloprotein by matching experimental and theoretical isotopic abundance mass distribution [12]. The isotopic simulations shown in Figure 4 suggest that the two LC peaks contain mixtures of tetra- and di-ruthenium sulfinate/sulfenate (GSO₂H/GSOH see Chart 1) glutathione complexes, respectively. The possible structures for the two main species are shown in Chart 2. The isotopic simulations (Figure 4) suggest that in the tetra-ruthenium adduct, two of the RuIV ions from the starting complex may be reduced to RuII, and in the di-ruthenium complex, both of the RuIV ions may be reduced to RuII. The biphenyl proton signals in the 1H NMR spectrum shown in Figure S8 for the fraction eluting from 10.90 to 12.50 min are too broad to observe, appearing to support the possibility that the ruthenium atoms in this cluster may be oxidized from RuIII to RuII, which is paramagnetic (although partial or complete spin-pairing is possible). The IR spectra of both of the adducts showed two bands at 855 and 920 cm⁻¹, evidence for the formation of S–O single bonds in S–O–H or S–O–M models. These data will stimulate further investigations of the nature of the postulated Ru–Ru, Ru–S, Ru–S–Ru, and Ru–S–O–Ru bonds, as well as the oxidation state of Ru [e.g., by X-ray absorption analysis].

The apparent stabilization of low oxidation states of ruthenium by oxidized sulfur ligands is surprising but may be aided by the formation of strong hydrogen bonds involving sulfinato oxygens under the acidic conditions used for chromatography and mass spectrometry (0.1% TFAH in the mobile phase) as has been reported for Ni(II)/Ni(I) in the nickel sulfinate/sulfenate complexes [13]. The presence of RuIII and hydride ligands (shift of two electrons from RuII or 2RuI to H⁺) cannot readily be ruled out since RuIII is known to form stable hydride complexes. However no 1H NMR
Table 1. The mass-to-charge ratios of the most abundant isotopomer of the ion fragments from the ruthenium sulfinate/sulfenate glutathione adducts $\left[\left(\eta^6{\text{-bip}}\right)\text{Ru(GS}_2\text{O}_2\text{)}_2\right]^{2+}$ (2) and $\left[\left(\eta^6{\text{-bip}}\right)\text{Ru(GS}_2\text{O}_2\text{)}_2\right]^{3+}$ (3) detected by direct infusion ESI-FT-ICR MS

<table>
<thead>
<tr>
<th>Adduct*</th>
<th>m/z (obs)</th>
<th>Ion</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>722.4247</td>
<td>(2 - Glu$^2$ - 50 + 5H)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>781.7766</td>
<td>(2 - 2O + 5H)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>787.1078</td>
<td>(2 - O + 5H)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>792.4406</td>
<td>(2 + 5H)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>1083.4502</td>
<td>(2 + 4H - Glu - 50)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>1094.1453</td>
<td>(2 + 4H - Glu - 40)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>1103.6351</td>
<td>(2 + 4H - Glu - 30)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>1108.1169</td>
<td>(unknown)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>1173.1748</td>
<td>(2 + 4H - 20)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>1181.1743</td>
<td>(2 + 4H - O)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>1189.1709</td>
<td>(2 + 4H)$^{2+}$</td>
</tr>
<tr>
<td>3</td>
<td>556.7662</td>
<td>(3 - Glu - Gly$^2$ + 11H)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>565.7668</td>
<td>(3 - Glu - 3O + 11H)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>609.1334</td>
<td>(3 - 3O + 11H)$^{3+}$</td>
</tr>
<tr>
<td></td>
<td>665.0892</td>
<td>(3 - Glu - Gly - GS$_2$O$_2$ + 10H)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>834.6312</td>
<td>(3 - Glu - Gly + 10H)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>848.1573</td>
<td>(3 - Glu - 3O + 10H)$^{2+}$</td>
</tr>
<tr>
<td></td>
<td>913.1215</td>
<td>(3 - 3O + 10H)$^{2+}$</td>
</tr>
</tbody>
</table>

*For proposed structures of adducts see Chart 2 and for mass spectra see Figures S9 and S10.

**Indicates loss of a Glu residue HO$_2$CCH(NH$_2$)CH$_2$CO (C$_5$H$_5$O$_2$N).

**H indicates gain of a proton.

**Indicates loss of a Gly residue NHCH$_2$CO$_2$H (C$_2$H$_4$O$_2$N).

peaks characteristic of hydride ligands (with negative chemical shifts) were detected. Also different species may be stabilized in the gas phase after desolvation than in solution.

The two fractions separated from the reaction mixture of complex 1 with GSH (2:20 mM) were collected using conventional HPLC and analyzed by ESI-FT-ICR MS. A series of doubly- and triply charged ions was observed and their m/z values and assignments are listed in Table 1; the mass spectra are shown in Figures S9 and S10. The results agree with the nanoLC-MS measurements, confirming that the first LC peak, shown in Figure 1, is a mixture of tetra-ruthenium glutathione sulfinate/sulfenate complexes, and the second, a mixture of di-ruthenium glutathione sulfinate/sulfenate complexes.

Our previous work has shown that the monoruthenium thiolate adduct formed from the reaction of complex 1 with GSH under physiologically-relevant conditions as described above is readily oxidized to the sulfenate adduct in the presence of O$_2$ [3]. However, under argon, the same reaction gave rise to a small amount of the di-ruthenium sulfinate adduct as shown in Figure S5, implying that, under these conditions, water may be involved as an oxygen donor during formation of the oxidized thiolate ligands in the ruthenium glutathione adducts, perhaps with accompanying reduction of Ru$^{11}$. Next we used isotopic labeling to test whether oxygen from solvent water can become incorporated into the sulfenate/sulfinate ligands in the products.

A reaction mixture of complex 1 with GSH (2:20 mM) in unbuffered 66.6% $^{18}$O-labeled water incubated at 310 K for 48 h was analyzed using the nanoLC-nanoESI-FT-ICR MS method described above. Two peaks similar to those shown in Figure 1 appeared in the TIC chromatogram (data not shown). The partial mass spectra of the fractions eluting at 18.2 and 22.6 min are shown in Figure 5, and the full scan spectra in Figures S11 and S12. For the first fraction, a doubly charged ion centered at m/z 1084.6487 and a triply charged ion centered at m/z 723.4261 (Figure 5 and Figure S11) were detected. Apart from the doubly charged ion centered at m/z 915.1388 and the triply charged ion at m/z 1219.1722, a quadruply charged ion centered at m/z 1371.6784 became detectable for the second fraction eluting at 22.6 min (Figure 5 and Figure S12), which appears to correspond to a trimer of the doubly charged ion at m/z 915.1388. The isotopic simulations (Figure 6) indicate that the two fractions contain tetra- and di-ruthenium glutathione sulfinate/sulfenate complexes, respectively, for which the oxygen atoms in the sulfinate and sulfenate ligands contain $^{18}$O and therefore arise from solvent water.

Conclusions

The glutathione adduct of the anticancer complex $\left[\left(\eta^6{\text{-bip}}\right)\text{Ru(en)Cl}_2\right]$ readily undergoes oxidation of coordinated glutathione to give rise to two multinuclear complexes.
complexes as the main products. These were unambiguously identified as di-ruthenium and teta-ruthenium glutathione sulfinate complexes. The MS data suggest that the initial Ru(I) may have undergone reduction to Ru(II) in these products. This possibility will stimulate further work to elucidate the oxidation state of ruthenium in these adducts and their coordination geometries. With 18O-labeled water as solvent, FT-ICR MS analysis indicated that oxygen atoms in sulfenate and sulfinate products can arise from water, implying that water can be involved in oxygen atom insertion into S-O bonds in the sulfenate and sulfinate ligands. Interestingly, the S-O bonds in cysteinyl sulfenate/sulfinate ligands in enzymes are thought to be involved in oxygen atom insertion into substrates and in biological signaling processes [14–18]. This work illustrates the power of FTMS coupled with nanoscale liquid chromatography and isotopic labeling for elucidating pathways for reactions of metalloproteins with intracellular biomolecules.

Acknowledgments

The authors thank RCUK (Interdisciplinary Research Centre RASOR), the Chinese Academy of Sciences, and Oncosense Ltd. for support. Dr. Holm Petzold (University of Edinburgh) for helpful discussion, and ORS for a scholarship for JX.

References


Sensitive, Specific, and Quantitative FTICR Mass Spectrometry of Combinatorial Post-Translational Modifications in Intact Histone H4

C. Logan Mackay,*+‡ Bernard Ramsahoye,‡ Karl Burgess,§ Ken Cook,‖ Stefan Weidt,† James Creanor,‡ David Harrison,‡ Pat Langridge-Smith,‖+ Ted Hupp,‡+ and Larry Hayward‡

SI-RFACMS, School of Chemistry, University of Edinburgh, Edinburgh, U.K., EH9 3JU, Institute of Genetics and Molecular Medicine, Division of Cancer Biology, University of Edinburgh, Crewe Road South, Edinburgh, U.K., EH4 2XR, Henry Welcome Functional Genomics Facility, Joseph Black Building, University of Glasgow, Glasgow, U.K., G12 8QQ, Dionex Ltd., 4 Albany Court, Albany Court Industrial Estate, Camberley, U.K., and RASOR Interdisciplinary Research Collaboration in Proteomic Technologies, University of Glasgow, Glasgow, U.K., G12 8QQ

We describe a quantitative Fourier transform ion cyclotron resonance mass spectrometric (FTICR MS) analysis of the relative proportions of post-translational modification states (PTMs) of core histones in cultured cells and tissues. A novel preseparation process using a monolithic column interfaced to a 12 T FTICR MS equipped with electron capture dissociation (ECD) yields very high mass accuracy spectra, allowing direct assignment of the PTMs present in the dominant modification states of intact H4, resolving a well recognized ambiguity between trimethylation and acetylation states. By eliminating preseparation, we also obtain a highly quantitative analysis of the distribution of H4 PTMs. Rapid, extensive, and reversible effects on PTMs induced by a histone deacetylase inhibitor indicate that H4 and other core histones are accessible to modification throughout the chromatin, not just in regions of active transcription. These methods provide tools for analysis of the histone code and its role in chromatin function.

In eukaryotes, nuclear DNA exists in a nucleoprotein complex called chromatin, the fundamental unit of which is the nucleosome consisting of a central histone octamer core (containing two molecules each of histone H2A, H2B, H3, and H4) around which the DNA is wrapped twice. In the nucleus, strings of nucleosomes are packaged into higher order chromatin structures regulated at several spatial scales, controlling access of enzymes involved in replication, transcription, and DNA repair.1-4 Remodeling of chromatin is therefore required during these processes.5 The precise structural basis of chromatin remodeling remains poorly defined but the association of specific and dynamic histone post-translational modifications (PTMs) with specific functional states of chromatin is well established.6-8 Modification specific antibodies have been derived to study histone PTMs including acetylation, methylation, phosphorylation, or ubiquitination of specific residues. These antibodies can be very sensitive and specific,9,10 but the quantitative extent to which a given modification applies is difficult to determine. Also, while chromatin immunoprecipitation and reprecipitation techniques can be used to detect different modifications in the vicinity of each other,11,12 these techniques are laborious and cannot readily detect multiple different modifications present on the same molecule, a potentially important functional signature. For example the pattern of multiple modifications in individual histones, and their association in nucleosomes, appears to be central to epigenetic regulation of DNA function in health and disease,13-14 providing a combinatorial "histone code" specifying chromatin function.15 A sensitive, specific, and quantitative method to systematically characterize histone PTMs

10.1021/ac702452d CCC: $40.75 © 2008 American Chemical Society
Published on Web 05/06/2008

Analytical Chemistry, Vol. 80, No. 11, June 1, 2008 4147
Figure 1. Mean (n = 4) MALDI-TOF mass spectra of acid extracted histones from untreated asynchronously growing wild type (dots) or p53 null (solid line) HCT116 colorectal cancer cells. Inset: mean (n = 4) with error bars representing the standard error of the mean for MALDI-TOF mass spectra of histone H4, including post-translational modifications, from untreated asynchronously growing wild type (dots) or p53 null (solid line) HCT116 colorectal cancer cells. Note the apparent superposition of noncovalent sulfate adduct species on top of the underlying PTM distribution.

Figure 2. 9.4 T broadband FTICR mass spectrum of the charge state distributions of the core histones present in a typical acid extraction from HCT116 cells, illustrating the reduced charge states and high background associated with contamination by low molecular weight organic compounds. Inset: 9.4 T broadband FTICR mass spectrum of the charge state distributions of the core histones present in a typical LC fraction of an acid extract from HCT116 cells, illustrating the broader charge states distribution and reduced background associated with elimination of low molecular weight organic compounds.

and especially their combinations in the individual histone molecule would therefore be of value. The various covalent histone modifications are catalyzed by families of enzymes which add or remove the relevant modification in a highly regulated process. In the case of histone acetylation, acetyl groups are added by histone acetyl transferases and removed by histone deacetylases. Trichostatin A (TSA) inhibits histone deacetylases at concentrations which are otherwise well tolerated by living cells, making it a useful drug for manipulating histone acetylation status experimentally in cell culture.

Recent advances in mass spectrometry (MS) provide direct methods for detecting histone PTMs by characteristic mass shifts in the intact proteins or their constituent peptides, the latter generated by physical fragmentation or by enzymatic digestion.

To date, these methods have tended to use preseparation of acid extracted histone mixtures, or their digested peptide products, using liquid chromatography. Prior digestion yields fragments of smaller mass amenable to mass analysis on a wider variety of mass spectrometers but may dissociate coexistent PTMs present in different parts of the intact protein and frequently also results in incomplete coverage of the primary protein sequence when...

Figure 3. 9.4 T compound FTICR mass spectrum of an isolated region of the 7+ charge state of histone H4, showing comparison with the simulated isotope distribution of the corresponding PTM states. Mass accuracy for each species is less than 2 ppm.

Figure 4. A 9.4 T compound FTICR mass spectrum of the 7+ charge state of histone H4 without ISA treatment, showing the observed post-translational modification distribution for the dominant N-terminally acetylated doubly methylated species. 9.4 T compound FTICR mass spectrum of the 7+ charge state of histone H4 following 3 h of 500 nM TSA treatment, showing the observed post-translational modification distribution for the dominant N-terminally acetylated, quadruply lysine acetylated, and doubly methylated species.
Figure 5. Time course of evolution of 9.4 T FTICR mass spectrum of the 7+ charge state of histone H4 during treatment of HCT116 cells with TSA. Dose response of 9.4 T FTICR mass spectrum of the 7+ charge state of histone H4 following 3 h of treatment of HCT116 cells with TSA. Time course of evolution of 9.4 T FTICR mass spectrum of the 7+ charge state of histone H4 after removal of the drug following 3 h treatment of HCT116 cells with 500 nM TSA.

Figure 6. Mean and standard error of mean (n = 3) for integrated abundance of each H4 ion species (every species is N-terminally acetylated) for p53 null (solid bars) and wild type (open bars) HCT116 without TSA. Mean and standard error of mean (n = 3) for integrated abundance of each H4 ion species (every species is N-terminally acetylated) for p53 null (solid bars) and wild type (open bars) HCT116 after 3 h of 500 nM TSA.

particular peptide species are detected poorly in the mass spectrometer. Chromatographic preseparation simplifies the observed mass spectra improving the specificity of assignment of ion species but may degrade the quantitative reproducibility of the technique. In these cases, chemical derivitization has often been used to improve quantitation. Mass analysis of intact proteins using high mass resolution MS, for example, Fourier transform ion cyclotron resonance (FTICR), avoids the problems of prior digestion and identifies coexistent PTMs in the individual protein molecule by characteristic mass shifts. In this paper, we further develop these FTICR MS techniques in order to determine the sequence and relative extent to which dominant PTMs occur in individual histone H4 molecules under varying experimental conditions.

By eliminating preseparation, we obtain a global view of all four core histones and their PTMs. Focusing on histone H4, we demonstrate a technique that offers rapid analysis and relative quantitation for the distribution of histone H4 PTMs in untreated cultured cancer cells and following histone deacetylase (HDAC) inhibitor (trichostatin A) treatment. To address the specificity of PTM assignments, we develop a novel preseparation process using a monolithic PF-DVB column (Dionex, Sunnyvale, CA) interfaced to a 12 T Apex-Qe Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with electron capture dissociation (ECD) yielding good sensitivity and.


Analytical Chemistry, Vol. 80, No. 11, June 1, 2008 4149
were treated for the indicated time with the desired concentration (1.1%). Following exposure, cells were washed with PBS and harvested or refed with drug-free media first, then resuspended in H2O, and clarified by centrifugation. The supernatant was discarded, and the pellets were washed in 1% 50 mM HCl in acetone and then in pure acetone overnight. The acetone supernatant was discarded, and the pellets were washed in 1% 50 mM HCl in acetone and then in pure acetone overnight. The acetone supernatant was discarded. 

**METHODS**

**Acid Extraction of Histones.** Wild type and p53 null HCT116 colon cancer cell lines (gift from B. Vogelstein, J. Hopkins) were cultured to approximately 60-80% confluence (10-15 million cells per 5 million cells, incubated for 15 min in eppendorfs on ice, and nuclei pelletcd by centrifuga-

**Figure 7.** nLC fraction (3 μL) 12 T FTICR mass spectra of the 13+ charge state of histone H4 (treated with 10 nM TSA for 3 h) illustrating the isolated region for ECD fragmentation. ECD fragmentation spectrum of the quadrupolar isolated N terminally acetylated and doubly methylated species of histone H4. N-terminal containing ions designated “c” and C-terminal containing ions designated “z”. Fragmentation pattern as identified by Prosight PTM (mass accuracy <10 ppm) and confirmed by Mascot.

**Figure 8.** nLC fraction (3 μL) 12 T FTICR mass spectra of the 13+ charge state of histone H4 (treated with 250 nM TSA for 1 h) illustrating the isolated region for ECD fragmentation. ECD fragmentation spectrum of the quadrupolar isolated N terminally acetylated and quadruply lysine acetylated, doubly methylated species of histone H4. N-terminal containing ions designated “c” and C-terminal containing ions designated “z”. Fragmentation pattern as identified by Prosight PTM (mass accuracy <10 ppm) and confirmed by Mascot.

---

**Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS).** Matrix-assisted laser desorption ionization time-of-flight mass spectrometry provided an overview of the protein species present in the acid extract and the reproducibility of their purity and abundance. An amount of 1 μL of unfractionated protein extract was spotted in triplicate on a MALDI plate with 1 μL of sinapinic acid in 48: U.K.), using approximately 1 mL per 5 million cells, incubated for 15 min in eppendorfs on ice, and nuclei pelleted by centrifugation at 800g for 10 min. Nuclear pellets were acid extracted using 1 mL of 0.2 M H2SO4 (incubated for 1 h at 4 °C) and centrifuged for 15 min at 13 000 rpm. Protein was precipitated by adding trichloroacetic acid (TCA) to a final volume of 20%. The supernatant was discarded, and the pellets were washed in 1% 50 mM HCl in acetone and then in pure acetone overnight. The acetone supernatant was discarded by centrifugation, and the resulting proteinaceous film containing histones was air-dried thoroughly, then resuspended in H2O, and clarified by centrifugation. The volume was adjusted to give a protein content of the supernatant approximately 1.5 μg/μL by Bicinchoninic Assay (BCA) assay. Sample quality was determined by 1D SDS-PAGE, and the presence of the core histones H4, H3, H2A, and H2B confirmed by Coomassie staining and Western blotting and by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS).
The spectra were batch processed in Data Explorer (ABI) using default settings of advanced background correction, noise filtering, and 5 point Gaussian smoothing. Multiple independent experiments \((n = 4)\) were performed for each experimental condition with an average over 5–8 replicates for each experiment, normalized across the \(m/z\) region of interest \((11\,000–17\,000)\), allowing calculation of mean spectra with standard errors.

9.4 Tesla FTICR MS of Intact Unfractionated Histones. Mass analyses were acquired using a modified 9.4 T Apex II Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonics, Billerica, MA) coupled to a nanoelectrospray ionization (nESI) enabled TriVersa Nanomate (Advion Bioscience, Ithaca, NY) used for direct infusion of acid extracted histone samples between 50 and 100 nL/min, without chromatographic preseparation. The samples were diluted to a final concentration of 0.15 \(\mu g/\mu L\) (histones \(\sim 2 \mu M\)) in electrospray buffer, consisting of 47.5:47.5 methanol (Riedel-De Haën) and \(H_2O\) with 5% formic acid. The 9.4 T Apex II FTICR incorporates an in-house built heated stainless steel desolvation capillary operating under instrument control at 160 °C. Desolvated ions were transmitted to a 6 cm Infinity cell Penning trap. Typical chamber base pressure was \(\sim 2 \times 10^{-11}\) mbar. Trapped ions were excited (frequency chirp 48–500 kHz at 100 steps of 25 \(\mu s\)) and detected between \(m/z\) 250 and 3000 for \(\sim 0.5\) s to yield a broadband 512 kword time-domain data. Each spectrum was the average of \(\sim 200\) scans. Time-domain ICR data was acquired using an AQR console. The acquisition rate was typically \(\sim 1\) scan/s. Fast Fourier transforms and subsequent analyses were performed using DataAnalysis (Bruker Daltonics, Billerica, MA). External calibration was achieved using cesium perfluoroheptanoic acid (CS PFHA) clusters providing a mass accuracy of less than 2 ppm. Isotopic distribution simulation was achieved using the Simulate Isotopic Pattern module of DataAnalysis (Bruker Daltonics, Billerica, MA).

12 Tesla FTICR MS of Intact Fractionated Histone H4. Top-down fragmentation of isolated histone H4 acetylation states required the development of a monolithic column based liquid chromatographic technique. This was achieved on an Ultimate 3000 system (Dionex, Sunnyvale, CA), equipped with a monolithic PicoSep (500 \(\mu m \times 50\) mm) analytical column (Dionex, Sunnyvale, CA). The column oven was maintained at 60 °C to minimize back pressure. Samples containing 1 \(\mu g\) of total histones were centrifuged (16 100g for 2 min) prior to injection onto the column. Solutions A and B consisted of 2:97.5 and 80:19.5 acetonitrile–water with 0.05% formic acid, respectively. Samples were injected onto the analytical column, washed with buffer A for 5 min, followed by a 20 min linear gradient elution \((29 \mu L/min)\). Online/off line sample introduction was achieved using the TriVersa Nanomate operating in positive ion mode, facilitating simultaneous nESI and fraction collection. A passive split ratio of \(\sim 100:1\) and a nESI potential of 1.45 kV was used. Separations were performed and 10 s fractions collected, resulting in fraction volumes of 3 \(\mu L\), collected in a 96 well microtiter plate and stored at \(\sim 20 \degree C\) for further analysis. Data were acquired on a 12 T Apex Qe FTICR (Bruker Daltonics, Billerica, MA) equipped with mass resolving quadrupole and electron capture dissociation. Desolvated ions were transmitted to a 6 cm Infinity cell Penning trap. Trapped ions were excited (frequency chirp 48–500 kHz at 100 steps of 25 \(\mu s\)) and detected between \(m/z\) 600 and 2000 for 0.5 s to yield broadband 512 kword time-domain data. Each spectrum was the sum of 200 mass analyses. Typical chamber base pressure was \(\sim 9 \times 10^{-11}\) mbar. Time-domain FTICR data was acquired using an AQS console. The acquisition rate was typically \(\sim 1\) scan/s. Fast Fourier transforms and subsequent analyses were performed using DataAnalysis (Bruker Daltonics, Billerica, MA). Multiple charge states could be observed in this way for each of the major core histone species. For MS/MS experiments, particular ion species were isolated using the mass resolving quadrupole and ECD fragmentation was utilized. Typically 1.7 A was applied to the dispenser cathode filament (Heatwave Technologies, McMinnville, OR), 20 V to the lens, and a pulse of 6 ms. Unlike the more conventional collision induced dissociation (CID) where the collision with gas molecules causes the breakage of peptide amide bonds results in N-terminal b and C-terminal y ions, ECD results from the capture by the protein of an electron resulting in the rapid cleavage of the amine backbone bond to produce N-terminal c ions and C-terminal z ions \((\text{or c}^+ \text{ and } z^+ \text{ ions})\). Fragment ion data were analyzed using the sophisticated numerical annotation procedure (SNAP) (Bruker Daltonics, Billerica, MA) and either Prosight PTM (The Kelleher Group, Urbana, IL) or Mascot (Matrix Science Ltd., U.K.).

RESULTS

Acid Extraction of Nuclear Pellets Reproducibly Yields High Purity Core Histones. Western blotting and Coomassie staining of 1-D SDS gels indicated the presence in cellular acid extracts of all four (H2A, H2B, H3, and H4) core histones in high abundance and purity. The linker histone H1 was also observed.

This was confirmed by MALDI-TOF MS (Figure 1). The MALDI spectra show abundant clusters of peaks in the expected \(m/z\) regions for all four intact core histone species, including variants and PTMs, for both the wt (dashed line) and p53 null (solid line) cells. H1 was also observed at lower abundance in the expected \(m/z\) region \((21–22\text{ kDa})\). As a result of the relatively unbiased ionization properties of MALDI, the technique provides a robust means of assessing reproducibility of the purification technique, although the mass resolution of MALDI-TOF MS is not sufficient to unambiguously assign individual variants or PTMs. It demonstrates minimal contamination by species with masses other than those expected for histones. A number of unknown contaminating proteins are visible in the spectra at low abundances (for example in the region of \(m/z\) 12 500).

The \(m/z\) 11 200–11 800 region of the spectra (Figure 1 inset), shows the major PTMs and adducts observed for H4. The extraction technique and MALDI-MS method yielded reasonably reproducible spectra from wild type (dashed line) and p53 null (solid line) HCT116 colon cancer cells \((n = 4)\), which were

---


Analytical Chemistry, Vol. 80, No. 11, June 1, 2008 4151
Assignment of Combinatorial H4 PTMs by Comparison of FTICR MS Spectra with Calculated Isootope Distributions. For unfractionated histone extracts, the charge state envelope observed on the 9.4 T FTICR MS was narrow, presumably due to ion suppression in the electrospray process from competition with contaminating small molecular weight organic species and proteins present in the samples. This is supported by the observation that fractions containing histones collected after an LC separation and reinfused display a typical, much broader charge state envelope centered close to m/z 1000 (Figure 2 inset). Nevertheless at least one charge state could be observed for each of H2A, H2B, H3, and H4 (Figure 2). We focus in this paper on histone H4 for which the 7+ and 8+ charge states were observed. The spectra were externally calibrated with reference to Cs PFHA clusters. The predominant species of H4 in both the wt and p53 untreated asynchronously growing HCT116 cells was identified as a doubly methylated moiety, with N-terminal methionine cleaved off and N-terminal acetylation present. These assignments were supported by comparison with calculated isotope distributions (Figure 3). It was found that although the peak centers of the spectra generated by isotopic simulation software did not vary, the ion’s abundance was very sensitive to the simulation parameters such as peak width. Other combinations of methylations and acetylations could also be provisionally assigned by the presence of characteristic mass shifts for methylation and acetylation (Figure 4A). Our assignments are in agreement with published data, but in common with published assignments, are subject to residual ambiguity for example between trimethylation or acetylation. We address this residual ambiguity below. Notably no species consistent with noncovalent sulfate adduction were found that although the peak centers of the spectra generated by acid extractions were performed for each of four experimental conditions (wild type or p53 null and TSA treatment for 3 h) and the 9.4 T FTICR MS analysis repeated on each unfractionated sample. The resulting spectra were cross-calibrated and normalized over the spectral region encompassing the H4 region in the 7+ charge state in m/z range 1600–1650. The relative abundance for each assigned PTM was estimated by integration of the isotope distribution in a 2 m/z interval, centered on the peak, and means and standard errors (n = 3) calculated for each experimental condition. The basal distribution among PTM states and the shift to higher acetylation induced by TSA proved highly reproducible, with statistically indistinguishable effects (within the standard error of the mean) evident in wild type and in p53 null HCT116 (Figure 6).

Electron Capture Dissociation of Isolated Acetylation States of Histone H4. Unambiguous assignment of the PTMs of histone H4 was achieved by chromatographic preseparation and ECD fragmentation of the reinfused fractions. Initially, ECD fragmentation was performed on the entire region of a single charge state of H4. The resulting fragmentation patterns provided ambiguous assignments as every possible combination of five acetylation sites and dimethylation of Lys20 was evident at comparable probability. Therefore, it was necessary to selectively enhance, using the instrument’s mass resolving quadrupole, the one acetylation (Figure 7) and five acetylation species (Figure 8) with their associated variously methylated subspecies resulting in unambiguous assignment of these modification states. By virtue of the accuracy of the measurement of the fragment masses, the well-known ambiguity of assignment to trimethylation or acetylation was resolved. We noted an absence of detectable Arg3 dimethylation and Lys20 trimethylation.

DISCUSSION

The FTICR MS developments we describe in this paper allow the unambiguous assignment of histone PTMs on the basis of high mass accuracy alone by “top-down” FTICR MS/MS and provide a highly reproducible assay offering relative quantification, under varying experimental conditions, of the distribution of
combinatorial PTMs in intact histones (exemplified by H4) in a global acid extract from cell nuclei. We recognize that histone acetylation status has been shown to affect electrospray ionization efficiency and that therefore absolute quantification will in the future require the incorporation of isotopically labeled standards in the sample for each acetylation state. Nevertheless because the ionization efficiency for any given acetylation state is independent of the experimental conditions under which cells are grown, the method does allow measurement of changes in the ratio of abundance of acetyl states under differing experimental conditions.

Each spectrum represents the signal derived from 3 pmol of histones at most, equivalent to the histone content of 2000 cells. However our current upstream liquid handling methods require a starting extract derived from at least 1 million cells. The future interface with nanoscale liquid handling methods in development will allow the study of material derived from several orders of magnitude fewer cells. We detect no significant difference between wild type and p53 null HT116 colon cancer cells in global histone modifications in the basal state or when treated with the HDAC inhibitor TSA. Our assay has potential utility in pharmacodynamic assessment of this new class of anticancer agent in vivo. We note that the rapid and dramatic global effects of TSA on H4 acetylation status indicate that the enzymes responsible for H4 acetylation and deacetylation must have access to H4 throughout the chromatin and not merely to H4 in regions of active gene transcription, as these active regions contain only a small percentage of the total cellular H4. It may be that undesirable nonspecific effects of HDAC inhibitors could be reduced by targeting HDACs involved specifically in active transcription, thereby eliminating this global effect on bulk chromatin. We also note an absence of detectable Arg3 dimethylation and Lys20 trimethylation in our bulk chromatin extracts, suggesting that these known H4 modifications, which are detectable by highly sensitive antibodies, must be present only in small regions of chromatin of special function. Our method will have future application to analysis of the role of the histone code in chromatin function, by analysis of subfractions of chromatin of defined function.

ACKNOWLEDGMENT
The authors thank RASOR Interdisciplinary Research Center and Cancer Research U.K. for support.

Received for review November 30, 2007. Accepted March 19, 2008.