# Exploring Gas-Phase Protein Conformations by Ion Mobility-Mass Spectrometry

Peter Allen Faull


Doctor of Philosophy by Research

The University of Edinburgh

2009

*For my parents, Deirdre & John*

*Science may set limits to knowledge,*

*but should not set limits to imagination.*

*From, "A History of Western Philosophy", by Bertrand Russell*
*British author, mathematician, & philosopher (1872 – 1970)*

# Acknowledgements

Over the course of the past four years, many people have earned an acknowledgement here. My thanks and eternal gratitude are directed foremost to my supervisor, Dr. Perdita "Tuning" Barran. Her help, guidance, knowledge, passion and ability to produce an ion signal from hours of noise (and very occasionally, the reverse) have provided me with a solid grounding in the dark arts of mass spectrometry and ion mobility-mass spectrometry. Perdi, thank you.

I would like to thank the members of the Perdita Barran Research Group 2005 to 2009 predominantly for the superb science performed and the scientific help in discussions, presentations and overcoming obstacles, but Journal Club cakes/coffee stick in my mind even clearer. To Drs. Bryan McCullough, Hannah Florance, Wutharth Chin, Holger Husi, Hayden Eastwood and Andrew Stopford, thank you for support and guidance in the earlier years of my Ph.D. To Dr. Bridgette Duncombe, Judith Nicholson, Roland Chu, Yana Berezovskaya, Harriet Cole and Stefan Eßwein, thank you for everything in later years. I must say a personal thank you to Martin De Cecco for improving my shattered self-confidence during some tough experiments by letting me win some squash matches, and a huge thanks to Dr. Stefan Weidt for his computer expertise and saving me on numerous occasions.

A special thank you must be said to an eccentric gentleman that I never believed I could become such close friends with, but am so glad I have: Mr. Jason Michael Drosos Kalapothakis. Not only did he challenge me continuously with the science of my work, he challenged me in all things I did too! My thanks to his contribution in this work for performing numerous in silico calculations for me, and in attempting to show me how to perform them myself – to little avail. Ευχαριστώ Kaetze.

My thanks to the engineers and scientists at Waters Micromass: Steve Pringle, Jason Wildgoose and Nick Tomczyk. All three were helpful and supportive throughout my time in Edinburgh and at the end of the phone when needed. My thanks to everyone who have assisted me with hardware support i.e. fixing equipment when it has decided enough is enough: Stuart and Davey in the mechanical workshop; Donald in the electronic workshop; and my thanks to world-traveller Douglas Munro in the physics electronic workshop, whose over-cautious use of electrical protection showed me that if it's not broken, it still may need to be fixed. Thanks must also be given to the University of Edinburgh staff who have made my Ph.D possible: the great guys in stores; and to Annette, Rona and Amanda in the departmental office.

Thanks go to the people who kindly provided me with protein samples, as my biological skills still leave something to be desired. To Dr. Christoph Schmidt for Factor H protein; to Dr. Hannah Florance for p53 protein; and to Judith Nicholson for AGR2 protein and the protein aptamers.

I thank all my friends in Edinburgh, Sheffield and London who have supported me and understood my absence at events due to, “The Ph.D”. They have helped me through the difficult times and reminded me why I am striving to do what I do. Thanks to my colleagues at John Lewis, Edinburgh, during my weekend “downtime”, and to the guys at Inverleith Hockey Club who welcomed me to top-level hockey when I first moved to Edinburgh.

Finally, I thank my wonderful family. My mother and father, Deirdre and John, who have had to encourage and support not one Ph.D student but two, and have done so with amazing calm and soothing advice. Their love and guidance have made all things possible. To my sister, Andrea, who submitted her Ph.D in August 2009 and understands first hand what it’s like to work into the wee hours. And to the lucky brother, Gavin Nicholas, who will never have to suffer a Ph.D but was always there when needed.

# Declaration

This thesis is submitted in part fulfilment of the requirement for the degree of Doctor of Philosophy at the University of Edinburgh. Unless otherwise stated, this work is my own and has not been previously submitted, in whole or in part, for any degree at this or any other university.

…….. Peter Allen Faull

August 2009

# Abstract

Analysis and characterisation of biomolecules using mass spectrometry has advanced over the past decade due to improvements in instrument design and capability; relevant use of complementary techniques; and available experimental and *in silico* data for comparison with cutting-edge research. This thesis presents ion mobility data, collected on an in-house modified QToF mass spectrometer (the MoQTOF), for a number of protein systems.

Two haemoproteins, cytochrome *c* and haemoglobin, have been characterised and rotationally-averaged collision cross-sections for a number of multimeric species are presented. Intact multiply-charged multimers of the form $[xCyt\ c + nH]^{z+}$ where x = 1 (monomer), x = 2 (dimer) and x = 3 (trimer) for cytochrome *c* have been elucidated and for species with $x \geq 2$, reported for the first time. Fragment ions possibly attributed to a novel fragmentation mechanism, native electron capture dissociation, are reported with a brief discussion into their possible production from the dissociation of the gas-phase dimer species.

Haemoglobin monomer globin subunits, dimers and intact tetramer have been successfully transferred to the gas phase, and their cross-sections elucidated. Comparisons with *in silico* computational data have been made and a discussion of the biologically-active tetramer association/dissociation technique is presented.

Three further proteins have been studied and their gas-phase collision cross-sections calculated. Two regions of the large Factor H (fH) complement glycoprotein, fH 10-15 and fH 19-20, have been characterised for the first time by ion mobility-mass spectrometry. Much work using nuclear magnetic resonance spectroscopy has previously been achieved to produce structural information of these protein regions, however further biophysical characterisation using mass spectrometry may aid in greater understanding of the interactions these two specific regions have with other biomolecules.

The DNA-binding core domain of the tumour suppressor p53 has been characterised and cross-sections produced in the presence and absence of the zinc metal ion that may control the domain's biological activity. Within this core domain, p53 inactivation mutations have been shown to occur in up to 50% of human cancers, therefore the potential exists to further cancer-fighting activity through research on this region. Anterior Gradient-2 (AGR2) protein facilitates downregulation of p53 in an as yet unclear mechanism. Recent work using peptide aptamers has demonstrated that this downregulation can be disrupted and levels of p53 restored. Collision cross-sections for six peptide aptamers have been calculated, as well as cross-sections for multimers of AGR2 protein. A complex between one aptamer with the protein has also been elucidated.

Use of the commercially available Synapt HDMS ion mobility-mass spectrometer at Waters MS Technologies Centre (Manchester, UK) allowed data to be collected for both Factor H protein regions and for the DNA-binding core domain of p53. Data are compared in the appropriate chapters with data collected using the MoQTOF.

# Abbreviation List

| | |
|---|---|
| 3D | Three-dimensional |
| µL | Microlitre; 1 µL = 1 x $10^{-6}$ litres |
| AGR2 | Anterior gradient-2 |
| ATD | Arrival time distribution |
| BIRD | Blackbody infrared radiative dissociation |
| CD | Circular dichroism |
| CID | Collision-induced dissociation |
| cm | Centimetre; 1 cm = 0.01 metres |
| Da | Dalton; 1 Da = 1.6605 x $10^{-27}$ kilograms |
| DC | Direct current |
| DNA | Deoxyribonucleic Acid |
| e | Electron charge; e = 1.602 × $10^{-19}$ C |
| EHSS | Exact hard-spheres scattering |
| ESI | Electrospray ionisation |
| FAIMS | Highfield asymmetric waveform ion mobility spectrometry |
| fH | Factor H |
| FT-ICR | Fourier transform ion cyclotron resonance |
| GAG | Glycosaminoglycan |
| GlcNAc | N-acetylglucosamine |
| Hb | Haemoglobin |
| HDMS | High Definition Mass Spectrometry |
| i.d. | Inner diameter |
| IM-MS | Ion mobility-mass spectrometry |
| kDa | Kilodalton; 1 kDa = 1000 daltons |
| kV | Kilovolt; 1 kV = 1000 volts |
| L | Litre |
| MALDI | Matrix-assisted laser/desorption ionisation |
| MDM2 | Mouse double minute 2 |
| mL | Millilitre; 1 mL = 0.001 litres |
| MoQTOF | Mobility quadrupole time-of-flight |
| ms | Millisecond; 1 ms = 0.001 seconds |
| MS | Mass spectrometry |
| $M_R$ | Molecular weight |
| *m/z* | Mass-to-charge ratio |
| Nano-ESI | Nano-electrospray ionisation |
| NMR | Nuclear magnetic resonance |
| o.d. | Outer diameter |
| PA | Projection approximation |
| QTOF | Quadrupole time-of-flight |
| RCSB | Research Collaboratory for Structural Bioinformatics |
| RF | Radio frequency |
| RNA | Ribonucleic acid |
| SID | Surface-induced dissociation |
| SORI | Sustained off-resonance irradiation |
| SRIG | Stacked ring ion guide |
| tATD | Total arrival time distribution |
| Td | Townsend number; 1 Td = $10^{-17}$ V $cm^2$ |
| TIC | Total ion chromatogram |
| TM | Trajectory method |
| TOF | Time-of-flight |
| T-wave | Travelling wave |
| TWIG | Travelling wave ion guide |
| TWIMS | Travelling wave ion mobility spectrometry |
| UV | Ultraviolet |

| | |
|---|---|
| VMD | Visual Molecular Dynamics (software) |
| XRC | X-ray crystallography |

# Chapter 1. Introduction

The Twentieth Century provided a wealth of scientific achievements which promoted the necessity for interdisciplinary research – that which requires study and contribution from more than one scientific discipline *i.e.* combinations of biology, chemistry, physics, mathematics and engineering. This concept has evolved and is now recognised in the Twenty-first Century as essential in academic and industrial research fields, since collaboration between centres of excellence aims to bring greater knowledge, clarity and ideally conclusion to scientific endeavours.

## 1.1 DNA Discovery – The Key to Protein Studies

The work contained within this Thesis is concerned with the structure of a single family of molecules, proteins, which are integral to a huge number of biological processes. Their production *in vivo* is now understood to be ultimately governed by one molecule in particular which was a topic of contested debate before its infamous structure elucidation in the 1950s. The post-World War II years yielded incredible discoveries and the scientific journal, *Nature*, described 1953 as an, "*annus mirabilis* for science," due to publications from Watson & Crick;[1] Franklin & Gosling;[2, 3] and Avery, MacLeod & McCarty[4] all concerning the structure and function of deoxyribose nucleic acid, DNA. This molecule was shown to be the material of inheritance passed between generations of organisms and is the key initial reagent required for protein production. From the realisation that the DNA molecule dictates the blueprint (the *genome*) for all biological processes of an organism, the field of genomics expanded and within a quarter of a century, the first DNA-based genome (total complement of genes) of the bacteriophage Φ X174 was sequenced in its entirety.[5] From this and other successes, the Human Genome Project began in 1990 and by April 2003, was declared complete.[6]

Genes are distinct coding and non-coding regions of DNA that are acted upon by numerous molecules to produce a single protein. This link from the sequence of DNA bases (aromatic heterocyclic molecules of guanine, alanine, thymine and cytosine) within a gene to the primary sequence of amino acids in a protein that it encodes provides the initial information necessary to study proteins. Unfortunately, the primary sequence of a protein does not provide any information for its higher order structure and once produced, will fold and change shape to assume a functional three-dimensional (3D) structure governed by the amino acids in the polypeptide chain.

## 1.2 Protein Structure

Proteins are biological polymers composed of amino acid monomer units which are enzymatically linked by peptide bonds in a condensation-type reaction. These bonds are inherently strong due to their planar nature and delocalisation of electrons between the C=O double bond and the electron lone pair on the N-H nitrogen atom. Amino acids all contain an N-terminal amine ($-NH_2$) functional group

and a C-terminal carboxylic acid (-COOH) functional group connected to a central alpha carbon (CH) as shown in Figure 1.1. The fourth valence position of the alpha carbon differs for each amino acid and confers reactivity and character such as acidity, basicity, aromaticity, solubility, hydrophobicity and ionisability. The simplest amino acid, glycine, contains a single hydrogen atom at this position and is the only achiral naturally occurring amino acid. A table depicting amino acid structure and naming conventions may be found in Appendix A.

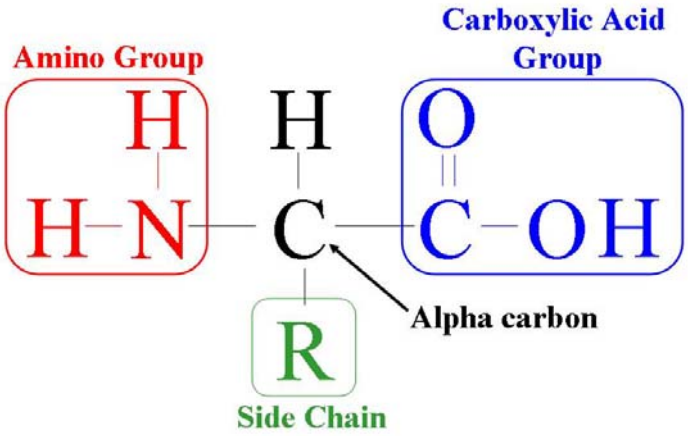



*Figure 1.1 Generalised structure of an amino acid depicting the amino (red), carboxylic acid (blue) & functional side chain (green) groups.*

Protein primary structure (the sequence of amino acids) is dictated by the DNA sequence of the relevant gene and production is a multi-step process. Briefly, the process of transcription enzymatically produces (*via* RNA polymerase) a single complementary messenger ribonucleic acid (mRNA) strand from the parent DNA strand in the nucleus. This is transported into the cytoplasm and combines with the ribosome (the protein-synthesising machinery). Amino acids are transported to the ribosome where they are chemically linked together in sequence *via* peptide bonds to form a nascent protein. This process of protein biosynthesis, *translation*, involves reading the information from the mRNA molecule. Amino acids are coded by three base pairs (trinucleotide or "codon"), with codons for initiation and termination processes providing control for correct protein production.[7] In solution at neutral pH, the majority of amino acids exist as dipolar ions called zwitterions due to protonation of the amino group and deprotonation of the carboxylic acid group. At low pH, the carboxylic acid group is not dissociated and the amino group will be protonated, resulting in a cationic molecule. As the pH is increased, the carboxylic acid functional group will lose a proton ($pKa \sim 2$) and the zwitterionic form will dominate until approximately pH 9 when the amino group will also lose a proton forming an anionic molecule.

Protein secondary structure is the arrangement of the polypeptide backbone into structural domains by the action of hydrogen bonding networks (*i.e.* formation of ordered short-range interactions). Successive addition of hydrogen bonding strengthens structures to form helices (such as $\alpha$, $\pi$ and $3_{10}$)

if there are amino acids present in the sequence that have a propensity of helical formation. Other forms of secondary structure include sheets (parallel and antiparallel) and turns. Non-covalent interaction between secondary structural elements, as well as disulfide-bridge formation, lead to the creation of tertiary structure where proteins are now able to adopt conformations capable of performing a physiological role.[8] Quaternary structure is achievable through interaction of multiple (two or more) protein chains (*e.g.* tetrameric haemoglobin) to form a complex arrangement of non-covalent interactions. Finally, quinary structure[9] is recognised as the packing of molecules within crystals or into virus shells, as well as helical quinary structures arising from monomeric units having characteristic tertiary structure, but little or no discernible quaternary structure.

## 1.3 Protein Folding

Proteins adopt specific conformations to perform their desired physiological role. The nature of this folding is an on-going field of interest. A simple experiment[10] performed by Christian Anfinsen in the late 1960s on ribonuclease A showed that enzyme activity had not been compromised after it had been denatured from its native state, then allowed to refold under native conditions. This refolding to assume a state of activity comparable to the initial native protein allowed Anfinsen to conclude that protein folding must be dominated by the non-covalent interactions between the amino acids, thereby dictated by the primary sequence. Shortly after Anfinsen's work was published, a theoretical biochemist, Cyrus Levinthal, suggested[11] that due to the huge number of degrees of freedom in an unfolded protein chain, a denatured protein molecule could adopt a similarly huge number of conformations. If a single polypeptide chain is imagined that is composed of one hundred amino acids, and that each amino acid has two possible conformations (for simplicity), the entire polypeptide chain would have $2^{100}$ possible conformations. Upon initiating protein refolding, the protein would then adopt every possible conformation. If one picosecond for each conformational state was allowed, the time required for the folding process would be $10^{18}$ seconds ($\sim 10^{10}$ years), which is a time greater than the age of the universe. Proteins are known that fold in on the microsecond time scale (such as the small protein, the tryptophan cage)[12] therefore this could not be the case. Cooperative interactions between amino acid residues during protein folding must overcome this problem and with each additional folding step, the following step is coordinated correctly. Other molecules assist in protein folding (chaperones) and enzymes exist to correct misfolded proteins, thereby allowing protein folding to occur in biologically-relevant time scales.

## 1.4 Mass Spectrometry

Mass spectrometry (MS) has become an indispensible analytical technique in the Twenty-first century with applications ranging from detecting and identifying chemicals in space[13] to much more simple and closer-to-home applications of identifying chemical compounds produced by, perhaps, a chemistry undergraduate. An excellent number of reviews on the subject of studying biomolecules by

mass spectrometry have been written.[14-17] Instruments vary in design due to the number of companies now producing specific mass spectrometers, but all perform the same job: to measure the mass-to-charge ratio (*m/z*) of an ionised entity; whether single ion or massive macromolecule. Research[18] carried out nearly one hundred years ago (in 1913), on ionised neon particles and their movement in magnetic and electric fields, led Sir Joseph John "J. J." Thomson to observe two light patches on a photographic collecting plate. Thomson's discovery of two neon isotopes (Ne-20 and Ne-22) was not only the first evidence for stable element isotopes, but also the first example of mass spectrometry. A gifted teacher, Thomson also received the 1906 Nobel Prize in Physics, "*in recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases*". Mass spectrometry is therefore soon to celebrate its centenary year, and Thomson would most likely recognise a modern mass spectrometer as the key components are still present (a source, an analyser and a detector). The developments that this technique has undergone, through both iterative improvements and great leaps of scientific ingenuity (that garnered their discoverers with Nobel Prizes themselves) proves that mass spectrometry, in its second century, should continue to flourish.

### 1.4.1 Ionisation Sources

The work in this Thesis presents mass spectrometry analyses of biological molecules, therefore the focus will be on instrumentation and techniques that are presently used in this field. To subject a biological sample to mass spectrometry analysis, it must be in ionised form and in the gas phase. This is achieved through ionisation using two techniques. The 2002 Nobel Prize in Chemistry was shared between Koichi Tanaka, John Fenn[19] and Kurt Wüthrich, "*for their work in developing methods to enable identification and structural analysis of biological macromolecules"* as they brought the two techniques described next to the wider scientific community. Another Nobel Prize for mass spectrometry to add to Thomson's.

#### 1.4.1.1 MALDI

Matrix-assisted laser desorption/ionisation (MALDI) and soft laser desorption (SLD – rarely implemented in modern biomolecule studies) were developed by Karas, Hillenkamp and Tanaka in the late 1980s.[20-24] These techniques make use of sample ablation from a solid target using a laser pulse. This produces ionised gas-phase molecules, which can then be transferred for analysis by mass spectrometry. Until the work of Hillenkamp *et al*., the sample was ablated in solid state and produced a large amount of fragmentation. The key was in the discovery that mixing the sample analyte with an organic compound (matrix) that strongly absorbs within the laser wavelength, produced significantly less fragmentation and extended the mass range of analytes that could be interrogated. Analytes are mixed with a saturated solution of the matrix - commonly used examples being 2,5-dihydrobenzoic acid or α-cyano-4-hydroxycinnamic acid. These compounds absorb strongly in the UV region with nitrogen lasers operating at $\lambda$ = 337 nm typical for MALDI experiments. Matrices should sublime at

room temperature, thereby encouraging co-crystallisation of the matrix and analyte to form optimal crystals on a target spot for ablation. Upon laser irradiation (experiment initiation), matrix excitation is thought to occur upon irradiating the target spot, with the majority of the laser energy transferred to the matrix. Subsequent analyte ionisation by gas-phase proton transfer (and, possibly to a smaller extent, ion-molecule reactions, preformed ion desorption and photoionisation) is currently the accepted mechanism of gas-phase ion formation.[25, 26] Singly-charged ions are most abundantly formed in the process, promoting the use of MALDI for large molecule analysis. However, biomolecules over 100 kDa in size interrogated by MALDI often result in broad mass peaks typically in the order of several hundreds Daltons. This may cause difficulty in accurately assigning peaks to calculated and expected species, however judging by the number of publications in the most popular mass spectrometry journals, MALDI is continuing to be used a solid and robust analytical technique.

#### 1.4.1.2 Electrospray Ionisation for Mass Spectrometry

Pioneered by Dole[27] in 1968 and subsequently developed by Fenn[28-30] in 1984, electrospray ionisation (ESI) is currently the method of choice for most biological mass spectrometrists as it produces *in vacuo* multiply-charged ions for both small and simple to large and complex solution-phase species. It is considered a mild ionisation technique as it does not destroy biological analyte molecules and therefore has proven itself invaluable in biological mass spectrometry.

In ESI, small charged droplets are formed due to the action of an applied potential difference between the capillary and the counter electrode (most usually the mass spectrometer sample cone). This potential is typically 1 to 4 kV and enriches the capillary tip with positively charged ions (under positive mode ionisation conditions). Charging leads to droplets with an elongated meniscus, a Taylor Cone, and releases solvent droplets containing charged analyte ions. Solvent evaporation (arrow 1 in Figure 1.2) due to heating of the source region and application of a stream of desolvation gas (frequently nitrogen) decreases the droplet size, and due to charge conservation, Coulombic forces overcome the liquid surface tension at a critical diameter, leading to fission of smaller charged droplets (arrow 2 on Figure 1.2). Repeated evaporation and fission events lead to very small charged droplets which are the gas-phase ion precursors.[15]

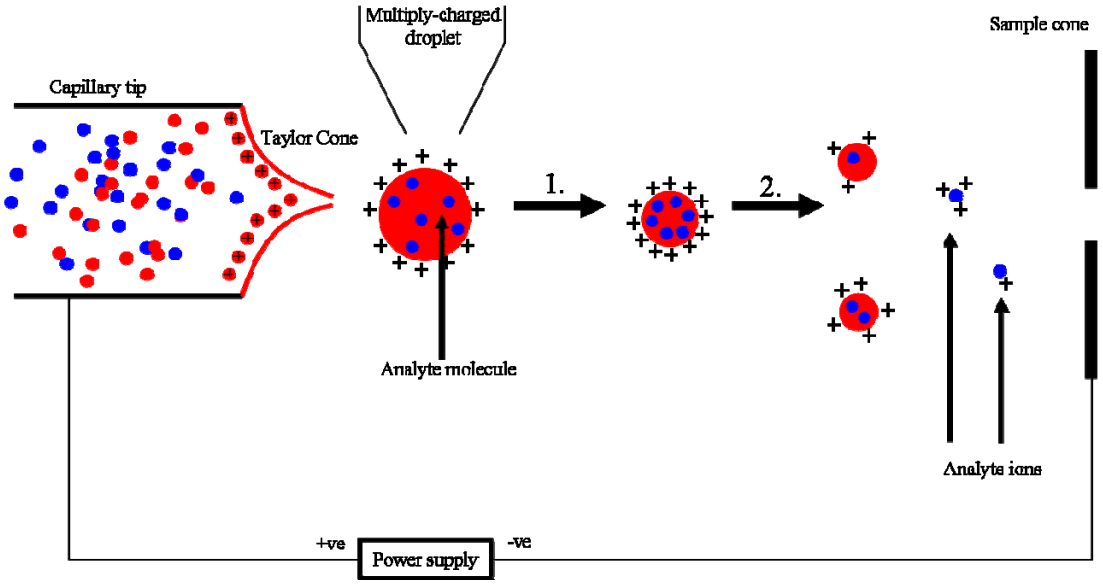



*Figure 1.2; Electrospray ionisation schematic showing formation of gas-phase multiply-charged analyte ions from solution-phase species. Mechanism 1. (annotated arrow) indicates solvent evaporation from the multiply-charged analyte-solvent droplet. Mechanism 2. (annotated arrow) indicates that the droplet has reached the Rayleigh limit and droplet surface tension is overcome by Coulombic force, leading to fission of charged analyte droplets which enter the mass spectrometer via the sample cone. Schematic produced by author using N. B. Cech and C. G. Enke diagram[31] as a guide.*

The diameter at which a droplet of given charge and solvent will undergo fission (Rayleigh limit) was investigated[32] in the Nineteenth century was described using equation 1.1: $q$ = droplet charge; $\varepsilon_0$ = environment permittivity; $\gamma$ = surface tension; and $D$ = droplet diameter.

$$q^2 = 8\pi^2 \varepsilon_0 \gamma D^3 \qquad \text{Equation 1.1}$$

Formation of solvent-free gas-phase ions is thought to differ between small ions and larger native-like proteins and protein complexes. Small ions may be produced by the ion evaporation model[33] where droplets reaching a certain size eject charged analyte ions from it due to Coulombic repulsion. Large ions are formed by the charge residue model[27] whereby droplet fission occurs until there is no longer solvent present on the analyte molecule. This mechanism is now widely accepted to be the most dominant form of ion formation from the ESI process for all analyte ion sizes.[34]

### 1.4.1.3 Nano-Electrospray Ionisation for Mass Spectrometry

Nano-ESI was developed in the mid-1990s by Wilm and Mann[35, 36] and promotes the use of small analyte volumes, without compromising signal intensity, due to the formation of smaller droplets for desolvation. Glass capillaries are pulled to a fine tip of 1 – 50 µm in diameter and a potential (lower than ESI; 1 – 2 kV) is applied either through the use of a metal (typically platinum) wire inserted into the analyte and in contact with the source housing, or *via* conductive metal-coated (platinum or gold)

capillaries. Sample flow rates are significantly decreased in nano-ESI with 10 μL of sample allowing a spraying time of more than three hours. Nano-ESI is more tolerant of salt contamination than traditional ESI,[36, 37] and the desolvation process is a gentler, more reliable method for introducing labile protein assemblies into the gas phase intact.[16]

### 1.4.2 Mass Analysers for Mass Spectrometry

Two mass analysers will be discussed due to their use within the work of this Thesis – the quadrupole and the time-of-flight. This region of the instrument separates gas-phase ions on the basis of their *m/z* ratio. Other mass analysers such as Fourier transform ion cyclotron resonance cells or quadrupole ion traps are also available, but will not be discussed further.

#### 1.4.2.1 The Quadrupole Mass Analyser

Quadrupole mass analysers are utilised in the vast majority of laboratories due to their inexpensive cost (both to purchase and to maintain), their ease of use and their ability to function as both a selective mass filter or as a full mass analyser. Developed by Paul[38] in 1957 (who was awarded a share in the Nobel Prize in Physics in 1989 for this development), quadrupole mass analysers are composed of four parallel cylindrical or hyperbolic rods, arranged as two sets of opposing electrode pairs, to which a fixed direct current (DC) voltage is applied along with an alternating radio frequency (RF) field. Opposite rods are in-phase with each other with respect to their RF field, and the adjacent rod pair are 180$^{o}$ out-of-phase producing one pair of negative rods and one pair of positive rods with respect to RF ground. Ions entering the quadrupole assembly are ultimately either stabilised or destabilised by the quadrupole field and are transmitted successfully, or strike the electrodes and are lost. This oscillating trajectory is determined by the inter-rod spacing (which decreases as the number of rods increases in assemblies such as hexapoles or octapoles); the applied potential amplitude, and the RF frequency. These determine the quadrupole electric field an ion experiences in the analyser region, and is calculated by equation 1.2[39]: $\Phi$ = total potential on the rods; $U$ = DC potential; $V$ = $RF_{zero\text{-}to\text{-}peak}$ voltage; $\omega$ = angular frequency of the RF (with $\omega = 2\pi\nu$, $\nu$ = frequency).

$$+\Phi_0 = +(U + V\cos\omega t) \quad \text{and} \quad -\Phi_0 = -(U + V\cos\omega t) \qquad \text{Equation 1.2}$$

A schematic of a quadrupole mass analyser is presented in Figure 1.3 showing the rod pairing (labelled + and – with respect to potential). An ion of defined *m/z* value will be stabilised and transported at a unique set of values for U, V and ω. If these values are fixed, only that ion will be successfully transmitted through the analyser. The quadrupole is thus acting as a mass filter. This is exploited in instruments that have multiple quadrupole assemblies aligned consecutively, allowing selection of a single gas-phase species to be subsequently fragmented by collision-induced dissociation (CID) and the ion fragments analysed in the next quadrupole region. If the V value (RF

amplitude) is scanned through a range of values, multiple ions of different *m/z* value (within a determined range) will be transmitted successfully.

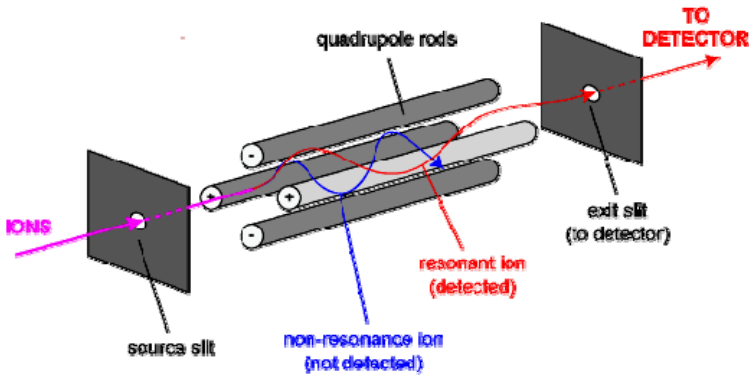



*Figure 1.3 Schematic of a quadrupole mass analyser. The resonant ion (red) traces a stable "corkscrew-like" trajectory through the analyser and is detected. The non-resonance ion (blue) is not detected due to its unstable trajectory, colliding with the rods and is lost.*

### 1.4.2.2 The Time-of-Flight Mass Analyser

First described by Stephens[40] in 1946, time-of-flight (TOF) analysers were integrated into commercially available instruments when Wiley and McLaren demonstrated their practical usage nine years later.[41] They operate by accelerating a packet of ions over a fixed distance using a potential difference. The ion packet enters a field-free drift region where individual ion velocity is related to its *m/z* value. This value can be elucidated as it is directly proportional to the ions flight time ($t_f$) through a fixed distance. Equations 1.3 to 1.5 show the relationship between ion acceleration potential and flight time. Equation symbols relate to: ion velocity ($v$); mass ($m$); kinetic energy ($E_k$); flight time ($t_f$); acceleration potential ($V$); ionic charge ($q$); integer ion charge ($z$); electron charge ($e$); and flight distance ($d_f$);

$$E_k = \frac{1}{2} m v^2 \qquad \text{Equation 1.3}$$

$$E_k = qV = zeV \qquad \text{Equation 1.4}$$

$$t_f = \sqrt{\frac{m}{z}\left(\frac{d_f^{\,2}}{2eV}\right)} \qquad \text{Equation 1.5}$$

Equation 1.3 provides clarity that at a specific kinetic energy, ion velocity is mass dependent. Equation 1.4 relates kinetic energy proportionally to ion charge in the acceleration region, with an ion of $z = 2^+$ receiving twice as much energy as the same ion with only $z = 1^+$. Combining these two equations and solving for flight time yields equation 1.5.

Many modern instruments are composed of reflectron TOF analyser regions, however early instruments used linear analyser regions where ions would be accelerated into a cylindrical flight tube using a pulsed high potential of several kilovolts on a pusher plate. Ions would separate according to their *m/z* value in the field-free region, and be detected at the opposite end. Upon detection, the next set of ions could be pulsed into the tube. It became apparent that an ions vicinity relative to the pusher acceleration plate affected the potential applied to it *i.e.* close proximity to the pusher plate provided a greater potential difference than an ion further away. This kinetic energy difference would cause two identical *m/z* ion species to have different flight times, arriving at the detector in quick succession and thereby reducing resolution. This issue has been overcome through use of a series of grids and ring electrodes used to create a retarding field – a reflectron – which was first described by Mamyrin[42] in 1973. It is positioned at the opposite end of the field-free drift region from the pusher plate, and used to reflect ions back through the flight tube to a detector located adjacent to the pusher plate. Upon acceleration from the pusher plate, ions with high kinetic energy will penetrate the reflectron field further than ones possessing lower energy. Tuning of the reflectron field therefore allows ions with the same *m/z* value but differing kinetic energies to have the same flight time within the TOF region and be detected simultaneously, providing increased resolution. Figure 1.4 shows a reflectron setup as used for a MALDI experiment.[43]

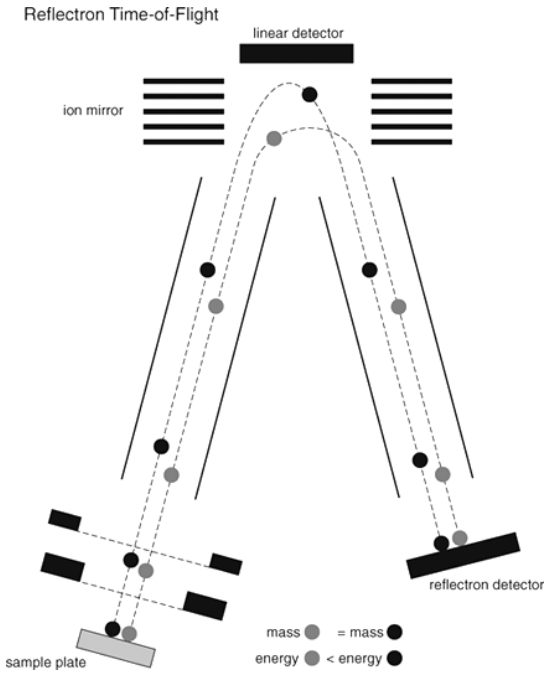



*Figure 1.4.[43] Reflectron time-of-flight region showing two ions of equal mass but differing kinetic energy arriving together at the detector due to implementation of a reflectron retardation field.*

### 1.4.3 Detectors for Mass Spectrometry

Ion detection is the final stage of a mass spectrometry experiment whereby a signal relating to the intensity of the ions present must be produced. Three detectors common to instruments containing quadrupole and/or time-of-flight analysers are discussed.

#### 1.4.3.1 The Electron Multiplier

Ions from the mass analyser are accelerated and collide with the surface of a first conversion dynode, producing electrons. [39, 44] These electrons are further accelerated to collide with a second dynode, producing more electrons. This process is repeated until the final dynode is reached; thereby producing a significant amplification of signal by up to $10^6$ greater intensity than the ions that struck the initial dynode. The signal is detected as an ion current with Figure 1.5 showing a schematic of an electron multiplier.

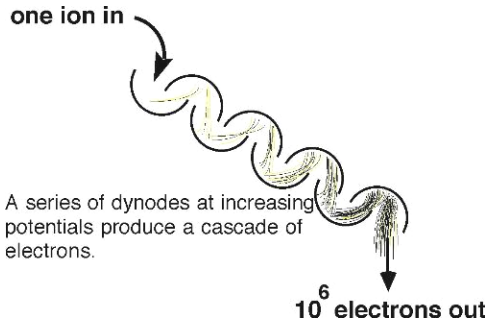



*Figure 1.5. An electron multiplier detection system showing that one ion striking the initial dynode can produce a signal of up to $10^6$ electrons at the other end.*

#### 1.4.3.2 The Photomultiplier Tube

A photomultiplier[39] detects generated photons from a phosphor screen. These photons have been generated by electron collisions with the phosphor screen; the electrons themselves being secondary particles ejected from a conversion dynode upon ions colliding with it, similar to the electron multiplier.

#### 1.4.3.3 The Microchannel Plate

The instrument used in this Thesis has a microchannel[39, 45] plate detector at the head of the TOF region. The plate is made from glass with many cylindrical channels bored into it, each channel functioning as an effective electron multiplier. A single ion entering one channel, colliding with the internal walls and ejecting electrons can produce over one thousand electrons at the channel end. This is a very sensitive detector and sudden changes in pressure or condensed moisture within the channels can damage it. Figure 1.6 shows a multichannel plate detector schematic.

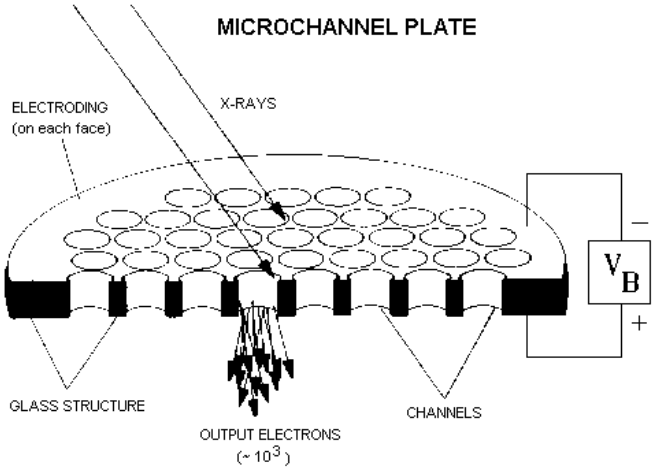



*Figure 1.6.[45] Microchannel plate schematic for ion detection and amplification.*

### 1.4.4 Further Mass Spectrometry Techniques

Further mass spectrometry techniques are available, and exist due to the ability for mass spectrometry to not only measure the mass of the intact species under investigation, but also to selectively dissociate the gas-phase species and to measure the ions produced (product ions) from precursor ions.

#### 1.4.4.1 Collision-Induced Dissociation

Early collision experiments were conducted using triple quadrupole assemblies[46, 47] however most current experiments are performed using quadrupole time-of-flight (QTOF) tandem instruments as shown in Figure 1.7.

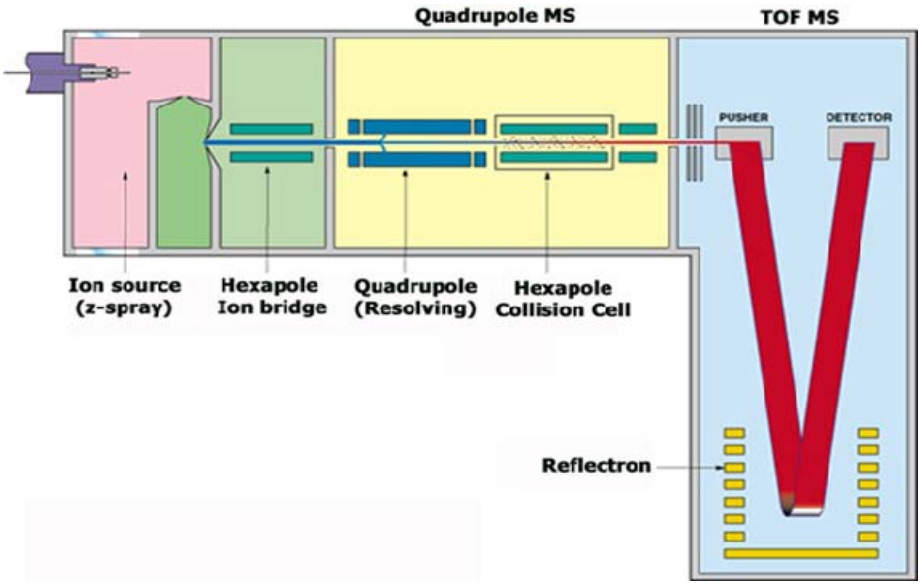



*Figure 1.7. Waters QTOF 1 instrument schematic showing the tandem quadrupole and time-of-flight analyser regions separated by a collision cell.*

A collision cell separates the quadrupole and TOF analyser regions and allows ions to be mass selected in the quadrupole region, subsequently activated in the collision cell by colliding them with an inert buffer gas, and the product ions analysed in the TOF analyser. Collision cell activation is frequently achieved by ion-neutral collisions[48] (with gases such as argon, helium or nitrogen) and stems from initial research by Tang *et al.* in 1994.[49] This work was termed nozzle-skimmer induced dissociation as ESI-generated ions were accelerated in the interface region between the source and initial quadrupole assembly. Further refinement led to the technique of collision-induced dissociation. A number of excellent reviews exist on the subject.[50, 51] Upon each collision with a buffer gas molecule, a small amount of an ions energy is converted from kinetic into internal modes.[52, 53] Large protein complexes undergo thousands of such small heating events over tens of microseconds.[16] Collisional activation is a complex process and Benesch[54] has written an excellent insight into the distinct events that occur depending on the amount of internal energy available including: cleaning, restructuring, unfolding, dissociating, and fragmenting. Fragmentation is typically observed through cleavage of the amide bond on the polypeptide backbone of proteins, allowing relatively facile characterisation of mass spectra produced due to the abundant production of *b* and *y* fragment ions. In 1984, Roepstorff and Fohlmann[55] introduced a naming convention for product ions formed from backbone cleavages. In this proposed scheme, $a_n$ and $x_n$ ions are generated by $C_\alpha$-$C_{carbonyl}$ bond cleavage; $b_n$ and $y_n$ ions are generated by $C_{carbonyl}$-N (the amide bond) cleavage; and $c_n$ and $z_n$ ions are generated by N-$C_\alpha$ cleavage with the subscript *n* denoting the number of amino acids contained within the ion, as shown in Figure 1.8.

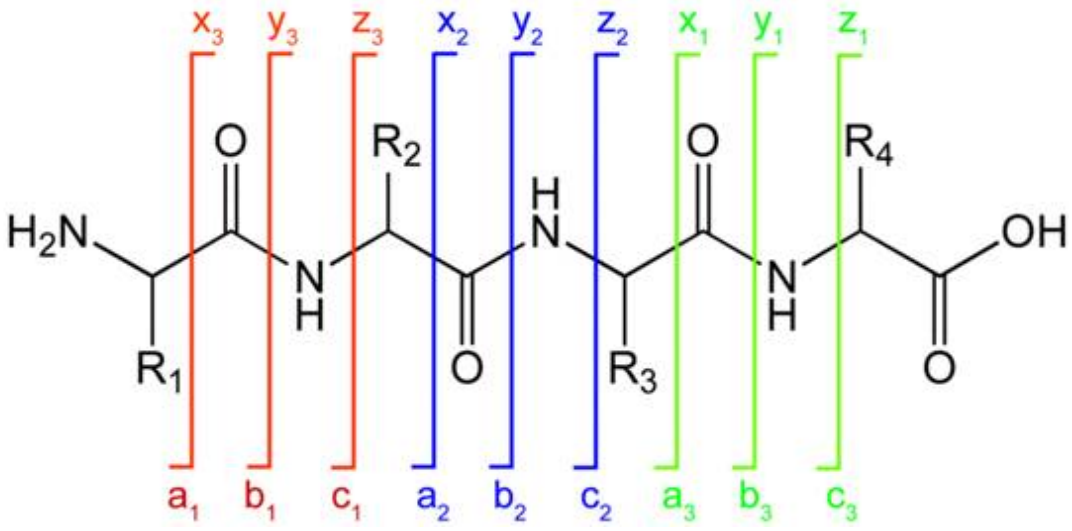



*Figure 1.8. Fragments produced by gas-phase collision techniques in proteins.*

### 1.4.4.2 Surface-Induced Dissociation

Incorporation of a surface into a mass spectrometer for ion activation was pioneered by Cooks *et al.* from the late 1970s.[56] Gas-phase analyte ions are collided with a surface producing fragment ions,

neutralised precursor molecules, sputtered surface atoms, and ion-surface reaction products. Many different surface materials have been used including graphite and diamond, but more common are hydrocarbon, fluorocarbon, and functionalised alkanethiolate self-assembled monolayer films on gold.[57] There is evidence that SID may provide more structural information than CID for large biomolecules[58] based upon increased complex dissociation *via* higher energy charge fission processes.

## 1.5 Other Bioanalytical Tools Used to Complement Mass Spectrometry Protein Studies

A number of additional bioanalytical tools are available to study protein structure. When used to complement mass spectrometry experimental findings, further information of the biological system under investigation can be produced that perhaps could not have been achieved through mass spectrometry alone. X-ray crystallography (XRC) and nuclear magnetic resonance (NMR) are two techniques exploited by nearly all leading groups in elucidating protein structure.

### 1.5.1 X-ray Crystallography & Nuclear Magnetic Resonance

Classified as high resolution techniques, XRC and NMR can provide accurate structural information but pose a number of difficulties for protein chemists. XRC involves crystallisation of a good quality defracting crystal which can take a number of months to perfect. Once achieved, the crystal is interrogated by diffraction of x-rays by electron density surrounding atoms within the crystalline cell. This electron density map is then deconvoluted and the inter-atomic distances (resolution-dependent) of the protein atoms can be realised.[59] NMR spectroscopy uses magnetic fields to interrogate the electromagnetic properties of protein nuclei both individually and their interaction with neighbouring nuclei. A connectivity map between neighbouring nuclei is produced and a protein 3D structure can be constructed. Use of nuclear Overhauser-effect spectroscopy experiments allow correlation between two protons if they are within 5Å of one another. Protons may be on adjacent atoms, or close to one another in space due to tertiary structure.[60] NMR experiments require large amounts of computational time; protein concentrations in the millimolar range; use of non-biological solvents and work best if the sample has been enriched with isotopes of $^{15}N$ and $^{13}C$.

### 1.5.2 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy may be used as a probe of protein secondary and tertiary structure and is common place as a low resolution and relatively simple technique in protein structural studies. CD[61] utilises the differential absorption of left and right circularly polarised light by chromophores present on amino acid residues (such as aromatic rings), the polypeptide backbone (amide bonds) and disulfide-bridges. These chromophores are capable of absorbing UV light at a specific frequency and the intensity of this absorption is dependent upon the conformation adopted by the chromophore within the protein structure, thereby providing an indication of the structural

environment. Far UV absorption (λ = 170 – 250 nm) is used to indicate the presence of secondary structure in proteins by backbone amide absorption; whereas near UV absorption (λ = 250 – 400 nm) is used as a probe of protein tertiary structure through amino acid side chain absorption by phenylalanine, tryptophan and tyrosine residues. UV and visible spectroscopy (λ = 200 – 600 nm) was utilised in structural research of bovine haemoglobin.[62] Far-UV CD showed that haemoglobin has a primarily α-helical structure based upon the negative ellipticity measured at 222 nm. Upon decreasing the pH of the sample, significant reduction of the helical content and adoption of a more disordered structure was observed. Near-UV measurements (λ = 400 – 430 nm) displayed a dramatic decrease in signal at lower pH indicating a significant degree of protein unfolding leading to loss of the hydrophobic pocket that stabilises a functional ligand.

## 1.6 Molecular Modeling – *in silico* Approaches to Protein Structure

Molecular modeling covers a wide range of approaches used to calculate, analyse and visualise molecular properties.[63] Computational (*in silico*) studies of protein structure have enabled comparisons to be made with experimental data for a wide dynamic range of biomolecules from small peptides[64] and glycosaminoglycans[65] to large biomolecules.[66] XRC and NMR biomolecule structures are available from the Research Collaboratory for Structural Bioinformatics[67] (RCSB) to download and modify as required. Using MOBCAL[68, 69] – a program to calculate mobilities – these structures can be modelled and computational collision cross-sections produced using three different methods.

### 1.6.1 The Projection Approximation Method

Developed by Mack[70] in 1925, this method treats all atoms within a polyatomic molecule as hard spheres. Interactions between the molecule and buffer gas are treated as simple hard-sphere collisions. Collision cross-sections using this method are calculated by averaging the molecule cross-section over all possible orientations. Mack produced beeswax models of organic compounds and mounted them on a moveable stand. Using a bright lantern, the shadow of the molecule would be projected onto a paper screen, and the stand moved to change the orientation of the molecule thereby changing the shadow projected. He was then able to sum the shadows produced to obtain a collision cross-section for the molecule. Now performed *in silico*, this method has been updated[71] to include a temperature-dependent term which includes an approximation of the long range interactions involved in the collisions.

### 1.6.2 The Exact Hard-Sphere Scattering Method

An improvement on the projection approximation method by including the interaction temperature and collision nature allowed Shvartsburg and Jarrold[68] to produce the exact hard-sphere scattering method. The orientationally-averaged momentum transfer cross-section is calculated by determining

the scattering angles between the incoming buffer gas atom trajectory and the departing buffer gas atom trajectory. From this, the collision integral ($\Omega$) is produced by averaging the momentum transfer cross-section over the relative molecule and buffer gas velocities. Collision integrals obtained are larger than those estimated from the projection approximation.

### 1.6.3 The Trajectory Method

The Trajectory Method is the most computationally expensive, requiring significantly longer time scales than either the projection approximation or exact hard-sphere scattering (EHSS) methods. It further refines the EHSS method with consideration of the long range interactions between ions and buffer gas.[69, 72] This is achieved using a potential that attempts to describe all interactions between the buffer gas and all polyatomic atoms in the ion. This includes Leonard-Jones and ion-induced dipole interactions. Cross-sections obtained using this method are the most accurate, however provide values that are within a few percent of those obtained by EHSS.[66]

## 1.7 Ion Mobility-Mass Spectrometry

When the further separation dimension of ion mobility spectrometry was added to the existing technology of mass spectrometry fifteen years ago,[73] few people could have predicted the explosive impact it would have. Figure 1.9 represents a search for, “ion mobility-mass spectrometry” (IM-MS) in PubMed[74] from January 1992 to August 2009 showing the number of journal articles available over the seventeen year period.

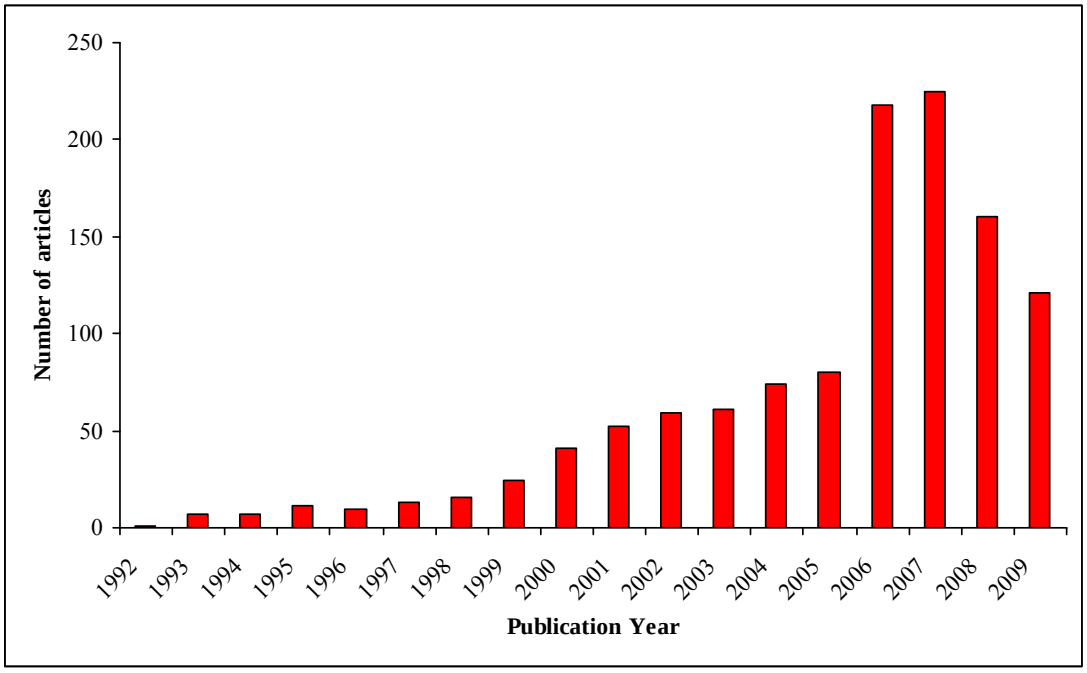



*Figure 1.9. PubMed search for "ion mobility-mass spectrometry" from January 1992 to August 2009. The first commercially available instrument was launched in 2004, perhaps explaining the increase in papers in subsequent years.*

Numerous excellent reviews[75-79] exist on the topic of IM-MS providing detail on features such as: macromolecular structures in solution and the gas phase; mobility measurement, peak shape and resolving power; instrumentation; combination with other techniques such as liquid chromatography; and ion formation. Ion mobility is fundamentally based upon the behaviour of ions moving through gases. Mason and McDaniel[80] co-wrote a book in 1988 that covers, in greater detail than the scope of this Thesis, this behaviour.

### 1.7.1 Ion Mobility Theory

Ion mobility ($K$; $cm^2\ s^{-1}\ V^{-1}$) is determined by an ions drift velocity ($v_d$; $cm\ s^{-1}$) under the influence of an electric field gradient ($E$; $V\ cm^{-1}$).[75, 76, 80] Ions are typically pulsed into a chamber filled with a buffer gas of known pressure ($P$; Torr), temperature ($T$; K) and length ($L$; cm) and the time taken ($t_d$; seconds) for them to traverse the chamber is measured. The electric field gradient provides the electrostatic force necessary to move the ions through the chamber; however collisions with the buffer gas retard this movement and ions reach a constant drift velocity. This relation is provided in equation 1.6:

$$v_d = KE \qquad \text{Equation 1.6}$$

If the time taken to traverse a drift cell of length $L$ (cm) is $t_d$ (seconds), then:

$$K = \frac{L}{t_d E}$$ Equation 1.7

To allow and promote comparison between instruments and laboratories, ion mobility is frequently expressed as reduced mobility ($K_0$; $cm^2\ s^{-1}\ V^{-1}$) corrected to standard conditions of pressure ($P_0$; 760 Torr) and temperature ($T_0$; 273.15 K):

$$K_0 = K \frac{T_0}{T} \frac{P}{P_0}$$ Equation 1.8

Substituting $K$ in equation 1.8 with the expression in 1.7, and replacing the electric field $E$ term with $V/L$ where $V$ is the drift voltage (V) applied across the mobility region provides equation 1.9:

$$K_0 = \frac{L^2}{t_d V} \frac{T_0}{T} \frac{P}{P_0}$$ Equation 1.9

Within the drift chamber, an ions energy is determined by the ratio of the electric field $E$, to the buffer gas density number, $N$. At low $E/N$ values when drift velocities are small compared with thermal velocities, mobility is independent of field strength and ions will not align in the electric field. This is known as the low-field limit. Conversely, the high-field limit will have ions with drift velocities greater than thermal velocities and alignment in the field is likely to occur. The instrument used in this Thesis operates in the low-field limit ($\leq$ 10 Td; 1 Td = $10^{-17}$ V $cm^2$) and therefore can be used to deduce structural information.[76] This is achieved by elucidating the rotationally-averaged collision cross-section ($\Omega$; expressed in $Å^2$) *via* the relationship[81] with mobility (reduced) in equation 1.10: $z$, integer ion charge; $e$, electron charge ($1.602 \times 10^{-19}$ C); $\mu$, reduced mass of the ion pair (kg); $k_B$ is the Boltzmann constant ($1.381 \times 10^{-23}$ J $K^{-1}$); and $T$, temperature (K).

$$\Omega = \frac{3ze}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{0.5} \frac{1}{K_0}$$ Equation 1.10

The rotationally-average collision cross-section, Ω, provides information about the shape and size of the ion which is dictated by the charge, reduced mass and mobility at a given drift gas temperature and pressure.

### 1.7.2 Development and Application of Ion Mobility Mass Spectrometry

Ion mobility spectrometry studies began in the 1970s under the name plasma chromatography[81, 82] and by the late 1970s,[83] structural isomers of polycyclic aromatic hydrocarbons could be separated by their

differing mobilites. Cobalt ($Co^+$ and $Co^{2+}$) ions were measured[84] in helium gas by Kemper and Bowers, who would continue to be field-leaders in ion mobility research. The Bowers group (University of California, Santa Barbara, U.S.A.) continue to investigate metal ion mobilities, but within complexes carrying pendant chromophores.[85] Ion mobility-mass spectrometers all contain a source to generate gas-phase ions; a drift chamber for mobility separation; a mass analyser; and a detector. Differences between instruments and techniques arise concerning whether the mobility region is a drift tube (linear ion mobility) or uses a travelling wave-based radio frequency-only stacked ion ring guide system. These will now be discussed.

#### 1.7.2.1 Drift Tube-Based Ion Mobility-Mass Spectrometry Instruments

Bowers and Kemper developed[86] a hybrid double focussing mass spectrometer in 1990 that contained a high pressure drift reaction cell. Cobalt mobilities of both ground state and metastable ions were investigated and new mobilities in helium were reported. Cell pressure could be varied up to 2 Torr; the cell could be cooled or heated; and exiting mobility separated ions were analysed using a quadrupole mass spectrometer. The drift cell set up used is the original copper-block design that is the basis for the mobility cell used in the MoQTOF instrument within experiments of this Thesis. Addition of ion funnels[87] to more recent instruments compress divergent ion streams leaving the source capillary down to a small diameter and transfer ions to the drift cell from the ion source without the use of high acceleration fields, thereby avoiding high-energy ion-neutral collisions. These are examples of low resolution ion mobility where ions have been injected into a drift tube with a buffer gas pressure less than 10 Torr at low drift fields.[88] Bower's research group continues to use instruments such as these to study the amyloid-β protein[89, 90] but have recently developed[91] a new, higher resolution instrument that contains a two metre long drift tube followed by quadrupole and TOF analysers (Figure 1.10). The instrument source can be equipped with either nano-ESI or MALDI ionisation sources, and the linear drift field is maintained with a series of 290 stainless steel drift rings with 100 mm o.d. × 75 mm i.d. connected by 100 kΩ resistors. A maximum drift voltage of 10 kV can be achieved, with gas pressures of 12 to 15 Torr. Electrical discharge problems can be reduced by addition of a small amount of $SF_6$ gas. Resolution of ~110 ($t_a/\Delta t$; angiontension II, $2^+$ peak) is observed when the drift region is filled with helium. The instrument is capable of separating molecules with cross-sections which differ by ~1%, and has been used to differentiate inverse peptides (H-SDGRG-OH vs. H-GRGDS-OH); differentiate optical isomers ([Tyr-dAla-Gly-Phe-Leu] and [Tyr-dAla-Gly-Phe-dLeu] molecules); and to study monomeric and multimeric structures of bradykinin.

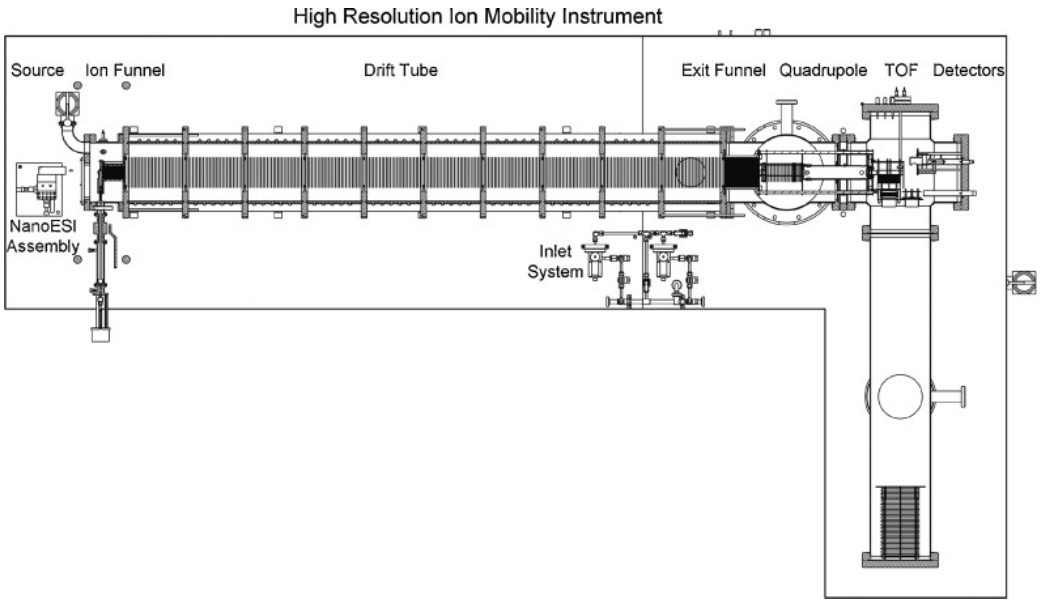



*Figure 1.10.[91] A new high resolution mass spectrometer with both nano-ESI and MALDI sources, a two metre long drift field and quadrupole and TOF mass analysers capable of significantly increased resolution ($t_a/\Delta t$ ~110).*

Instrument development is a strength of David Clemmer's group (Universirty of Indiana, U.S.A.) They were the first to design an ion mobility device with a TOF analyser post-IMS,[92] and further modified the instrument to include a quadrupole and collision cell between the drift tube and TOF regions.[93] They further refined the instrument by adding an ion trap prior to the drift tube to allow ion storage before injection into the mobility region.[94] In the past few years, the Clemmer group have turned their focus towards developing instruments with multiple consecutive ion mobility regions, allowing for $IMS^n$ ion separation. Their instrument[95, 96] incorporates a three metre long drift tube with several ion funnels and activation regions along the drift axis. The drift tube can be operated as a single long instrument, or divided into three independent drift regions. It is also possible to select an ion of specified mobility in the first region, and collisionally activate it. The fragment ions formed, or new gas-phase conformations adopted, can be separated in the second low-field region before exposure to a second high-field region for activation and analysis by MS. Electrosprayed ubiquitin and insulin chain B proteins were investigated using this instrument setup,[96] and many new peaks identified as isobaric ions were successfully separated on the basis of their mobility. The results also tentatively suggested that some ion types fell into "families" of structure classes (helical or globular motifs). Proteome profiling using nano-liquid chromatography coupled to this instrument have also been researched to separate and characterise a complicated peptide mixture from digestion of soluble proteins extracted from human urine;[97] to analyse and compare the proteomes of the heads of fruit flies;[98] and to study the proteins present in whole embryo and adult fruit fly heads.[99]

The most recent instrument and technique development from the Clemmer group, Overtone Mobility Spectrometry,[100, 101] allows a linear drift tube to be operated as a continuous filter. Ions generated by electrospray enter a drift tube with segmented drift regions (5.8 cm in length; 22 segments in total

with one replaced in the middle of the instrument with an ion funnel; ~2.25 Torr helium buffer gas) and the drift fields imposed are modulated at a frequency that allows only those ions having mobilities that are resonant with the experimental conditions to be transmitted through all drift regions. A uniform 10 V $cm^{-1}$ DC field is applied across each segment using a combination of applied voltages and 57 kΩ resistors in series. Drift fields are applied to adjacent pairs of segments to create a sawtooth voltage gradient down the axis of the drift tube. A second sawtooth voltage gradient is offset from the first by one segment and is initiated upon turning off the first voltage gradient.[100] Alternation of these fields is referred to as the drift field application frequency and is the time that one field is operated at whilst the other is off for the same amount of time *i.e.* at a 2 kHz drift field application frequency, one phase will operate for 500 µs whereas the other will be off for 500 µs. The two phases are alternatively operated and turned off to propel the ions through the drift tube. Figure 1.11 is a schematic of the instrument.

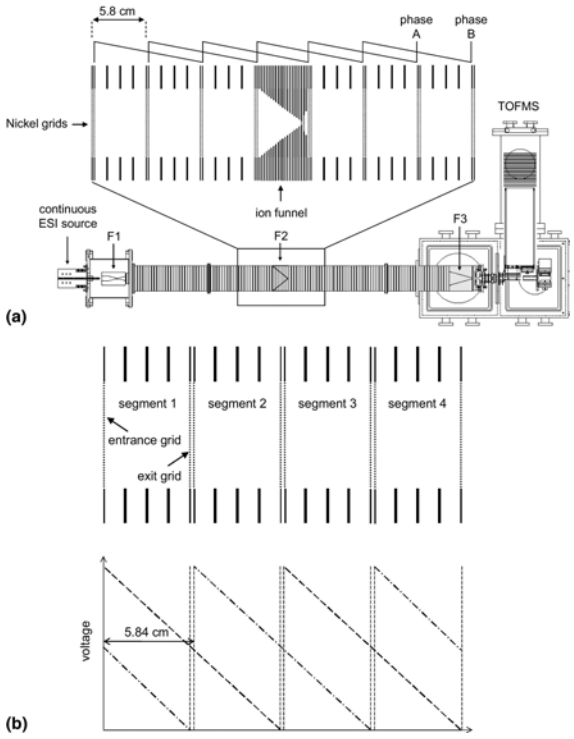



*Figure 1.11.[100] Schematic diagram of the overtone mobility spectrometer drift tube. (a). The instrument incorporates an ESI ion source, three ion funnels (labelled F1 to F3), a segmented drift region, and a time-of-flight mass spectrometer. Upon application of a linear field (and gating ions at F1) it is possible to record traditional ion mobility mass spectra. The middle region of the instrument is divided into segments, as shown in detail in the inset of the Figure. This region consists of 22 sections separated by nickel grids. The drift field of each section is controlled by a two-phase circuit. When the segmented region is operated at a specific field application frequency, it is possible to continuously transmit ions of a specific mobility.*
*(b) Schematic diagram showing the details of the segmented drift tube (top) and the overlap of the voltage gradients on the drift tube for the two sawtooth waves (bottom). The dashed line corresponds to the first sawtooth wave (phase A) and the dashed-dotted line corresponds to the second sawtooth (phase B). Here the application of the two phases is modulated at different frequencies to allow for the transmission of different ions through the drift tube. For this drift tube setup, the drift field is kept at 10 V $cm^{-1}$. At this drift field, the space between two sections (between two adjacent grids) consists of 114.4 V, giving a field of −477 V $cm^{-1}$. The region containing the negative field essentially acts as an ion elimination region.*

### 1.7.2.2 Synapt High Definition Mass Spectrometry Instrument: Using Travelling Wave Ion Mobility Technology

Waters Micromass Technologies (Atlas Park, Manchester, U.K) released the first commercially available ion mobility-mass spectrometer, the Synapt HDMS, in 2004 (Figure 1.12). Until then, instruments were built in-house and were tailored for the needs of the particular research group and experiments could potentially be difficult to replicate between groups with different experimental hardware.

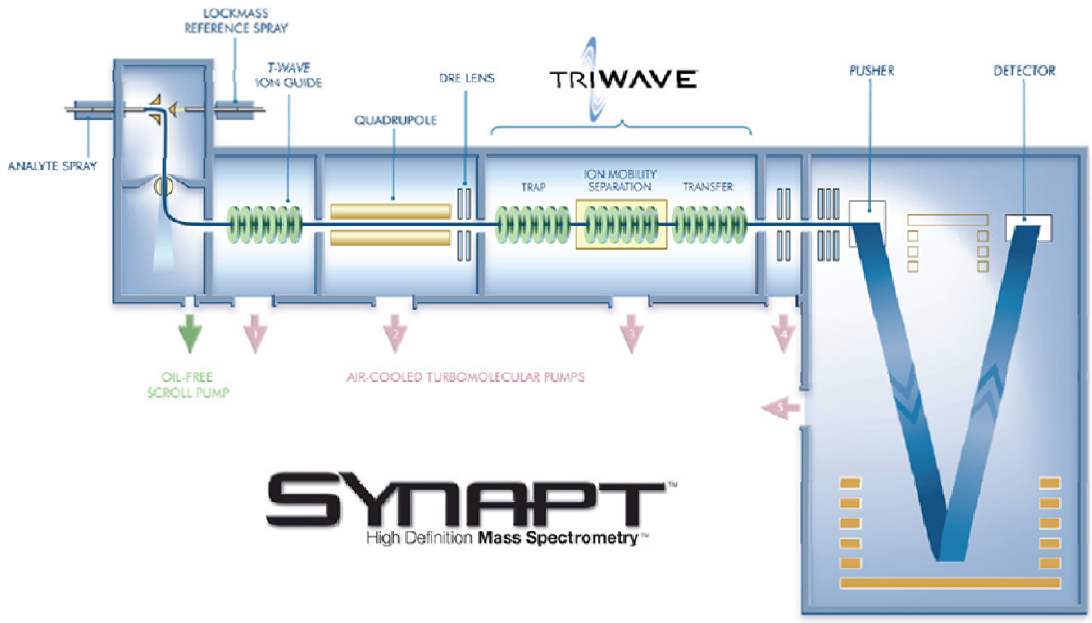



*Figure 1.12.[102] A schematic diagram of the Synapt HDMS instrument.*

Use of an RF-only stacked ring ion guide (SRIG) to generate a radially-confining effective potential well[103] allows ions to be trapped at elevated pressures. The SRIG is composed of a series of ring electrodes (Figure 1.13) with adjacent rings having opposite RF voltage phases applied to them. Ions are propelled through the SRIG by superimposing a DC potential on the RF voltage and then switching it to an adjacent electrode. Sequential movement of the potential moves ions on a "travelling wave" on which ions can "surf" (Figure 1.14), thereby converting the SRIG to a travelling wave ion guide (TWIG). The three main parameters that determine successful ion transport are the wave height (V), wave velocity ($ms^{-1}$) and gas pressure. At elevated nitrogen gas pressure (~1 mbar), ion separation within the travelling wave can occur based on their mobility and due to collisions with the buffer gas. Collisions will cause ions with lower mobility to be displaced over the wave crest to the previous potential well, and will reside there until the next "wave" carries them forward. This is referred to as travelling wave ion mobility spectrometry (TWIMS). The TWIMS setup in the Synapt instrument consists of 61 electrode pairs with a pulse repeat pattern of six pairs. The mobility region is 185 mm in length; operates at nitrogen pressures up to 1 mbar; and travelling waves of up to 25 V are used with wave velocities of 200 to 600 $ms^{-1}$. TWIG technology has been applied to one region prior (the trap region) and one post to (the transfer region) the mobility TWIMS region. The trap region is composed of TWIGs; 33 electrode pairs with RF applied to all but the last electrode, which is DC-only with a variable voltage so as to allow periodic gating of ions into the mobility TWIMS region. So as to ensure mobility separated species are maintained into the TOF region, the transfer TWIG (33 electrode pairs) has a continual 1 to 2V, 300 $ms^{-1}$ wave applied to it.

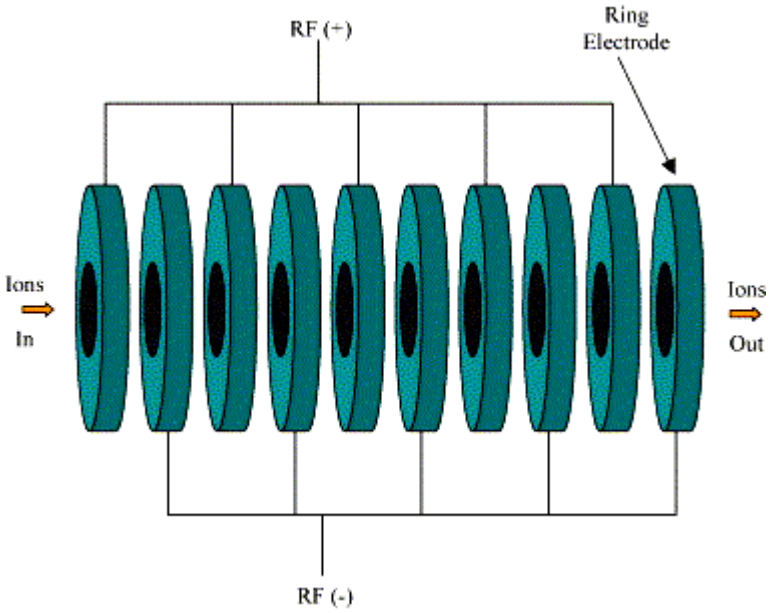



*Figure 1.13.[104] A stacked ring ion guide (SRIG) with opposite phase RF fields applied to adjacent electrodes so as to radially confine ion species.*

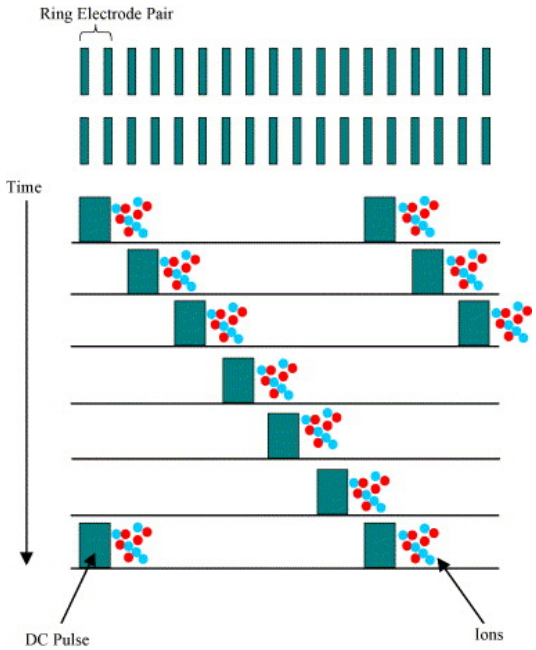



*Figure 1.14.[104] TWIG operation for ion propulsion in the presence of a background gas.*

Synapt instruments (or modified instruments similar to the Synapt) are used within a number of research groups to look at drug and pharmaceutical molecules,[105, 106] peptides,[104, 107] and large biological molecules.[66, 108, 109] Data obtained on the Synapt instrument must be calibrated against known drift time ion mobility-mass spectrometry data,[110] preferably with similar arrival times and masses.[109, 111] Normalised collision cross-sections corrected for charge and reduced mass are plotted against corrected arrival times which exclude the time spent outside the TWIMS region. This creates a calibration curve with a power-series fit. The calibration allows an estimation of a molecules cross-

section provided that the mobilities (corrected arrival times) lie within the mobility range observed for the calibrant used, irrespective of the calibrant cross-section size range.[112]

### 1.7.2.3 High-field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS)

High-field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) is an atmospheric pressure mobility technology capable of separating ions depending upon their differing mobilities in high and low field, with the first paper produced in the early 1990s.[113] FAIMS technology operates by applying alternating strong and weak electric fields to ions drifting in a gas.[114] These fields are applied to electrodes which are either two flat plates, or concentric cylinders. In the flat plate setup, an asymmetric waveform (Figure 1.15) is applied to one plate and the other plate is at a fixed DC voltage.

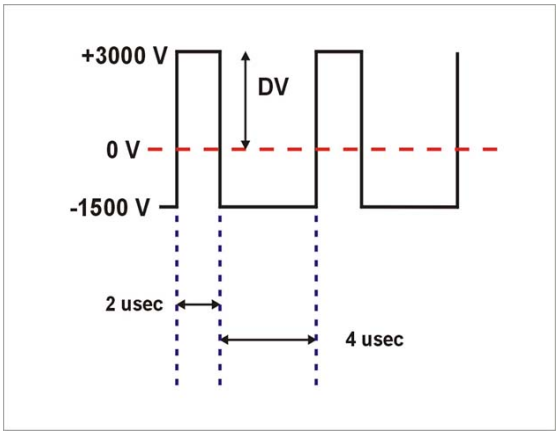



*Figure 1.15.[115] Asymmetric waveform utilised in FAIMS technology. DV is the dispersion voltage.*

The asymmetric waveform has a high applied (dispersion) voltage for an initial short period, followed by a lower, opposite polarity voltage for a longer period. The product of time and voltage for each of these two periods is equal. This waveform is applied for an extended length of time and causes a net drift of ions towards one plate, causing ions to move in a sawtooth motion. The drift is counteracted by the application of a compensation voltage, causing the drifting ion to move in a direction opposite to its drift motion and therefore does not collide with the electrode. Manipulation of the dispersion and compensation voltages permits transmission of a subset of the original mixture of ions, with all other ions colliding with one of the electrodes. Ions transmitted under particular voltage conditions have comparable change in mobility from high field to low field conditions, but they do not necessarily have the same low field mobility.[114] In cylindrical geometry FAIMS, similar principles apply however ions are focussed to a greater degree due to drifting to an optimal radial location between the concentric cylinders, promoting greater transmission and greater sensitivity. Bradykinin peptide,[116] ubiquitin[117] and cytochrome *c*[118] protein conformers have been successfully separated using FAIMS technology. Due to low field mobilities not being attainable, collision cross-sections cannot be

elucidated from FAIMS alone. However, FAIMS has been used to separate gas-phase ubiquitin conformers produced by ESI and then sampled by a triple quadrupole mass spectrometer where energy-loss measurements were used to determine collision cross-sections.[119] Not only were multiple conformers elucidated for a number of charge states, but cross-sections also compared favourably with literature values.

## 1.8 Studying Intact Biomolecules Using Mass Spectrometry & Ion Mobility-Mass Spectrometry

Interdisciplinary research on biological molecules has assisted in the evolution and development of mass spectrometry as a bioanalytical tool. A number of excellent reviews[120-124] and papers[78, 125] show that studying large biomolecular conformations, protein structure and function, and protein-ligand interactions using ESI-MS and IM-MS provides extremely worthwhile, interesting and relevant experimental data, revealing much about the intrinsic properties of the molecules in the absence of bulk solvent. An area of intense interest and expansion is intact protein mass spectrometry – the ability to transfer and characterise an intact protein frequently containing multiple components, from solution phase into the gas phase.[126] Transfer into the gas phase allows potential separation of hydration interactions and intramolecular interactions, and allows them to be examined independently.[124] Research[16, 127-129] suggests that a number of factors need to be optimised for this to occur: samples need to be of an adequate concentration (1 nM to 100 μM) so as to avoid unwanted aggregation, but concentrated enough to overcome mass spectrometer sensitivity and ion loss issues; biological samples are to be present in a buffer solvent that stabilises non-covalent interactions with minimal salt content; instrument source pressure needs to be controllable as elevated source pressures ($> \sim 1 \times 10^{-1}$ mbar) above those used in traditional mass spectrometry ($< \sim 1$ to $5 \times 10^{-2}$ mbar) have been shown to assist in efficient transfer of large macromolecular species for MS detection;[15, 126, 130, 131] and utilising nano-electrospray ionisation to produce charged biomolecules from solution that are efficiently transferred to the gas phase without significant disruption to non-covalent interactions.[15] Correct charge state assignment of native electrosprayed protein complexes is also crucial for large species due to the number of possible protonation events and inclusion of small adducts.[132] Understanding the above factors has allowed large macromolecular assemblies in excess of 1 MDa to be analysed intact by MS,[133] and invites further promising research into whole protein mass spectrometry.

Recently, Breuker[134] studied the temporal evolution (picoseconds to minutes) of the native protein structure of cytochrome *c* during and after transfer into the gas phase. She observed that the process can involve a number of processes including side-chain collapse, unfolding, and refolding into new, non-native structures. Charged side-chain collapse occurs on the picoseconds timescale and actually stabilises the native fold, producing transiently-stabilised protein ions. Loss of hydrophobic interactions and subsequent dissociation of electrostatic bonds occurs on the millisecond time scale,

with new non-covalent bond formation occurring on the second to minute time scale. These findings help to further understand the complex number of events occurring from electrospraying proteins to analysis. Ruotolo and Robinson[128] reviewed recent findings for monomeric proteins, as well as multiprotein assemblies in the gas phase that suggest many features of protein structure (compact structure, elements of secondary structure, hydrogen bonding interactions and quaternary structure) can be retained in the absence of solvent.[127, 135] They state that the results may not imply the precise interactions present in the native state of a protein complex are maintained, but overall size and many binding interactions interrogated for solution-phase proteins are found to be transferred successfully to the gas phase.

Cytochrome *c* (Chapter 3) has been extensively used in MS and IM-MS studies[118, 136-152] with original work into protein conformation studies by IM-MS performed in 1995 by Clemmer, Hudgins and Jarrold.[136] They observed a number of conformational structures including a native-like compact form (A); partially unfolded structures with an open haem crevice (B); an unfolded coil that retains α-helical structures (C); a typical random coil structure without secondary or tertiary structure (D); and a near linear conformation (E). This range of conformations was detected through observation of the $z = 7^+$ charge state IM-MS drift time measurements and reveals the interesting number of structures the protein may adopt upon transfer to the gas phase. Spraying cytochrome *c* from a number of solvent conditions produced a wide range of charge states with multiple conformers observed at the same charge state. Clemmer's group have a database of cross-sections[153] for a range of species including tryptic digest peptides, oligonucleotides, oligosaccharides, dendrimers, and small organic ions. Many research groups utilise the protein cross-sections from the database as absolute values for calibration of data (as discussed in the Synapt HDMS section). Protein cross-sections for ubiquitin, cytochrome *c*, lysozyme and apomyoglobin are available. Figure 1.16 is a plot of cytochrome *c* collision cross-sections obtained from the database showing the number of conformations observed at various charge states, as well as the calculated cross-sections for a native and extended conformation.[154]

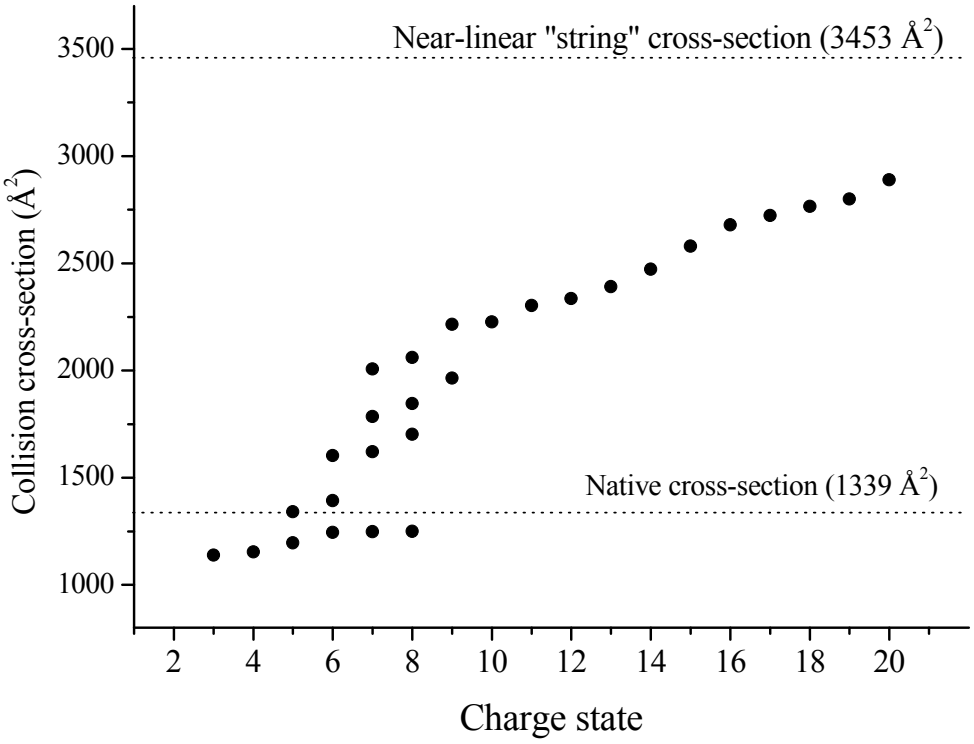



*Figure 1.16.[154] Plot of database cytochrome c cross-sections as a function of charge state showing multiple resolved conformations at discrete charge states. Dotted lines indicate cross-sections calculated using the exact hard-spheres scattering method for the native structure and an extended "string" structure.*

The native structure cross-section was determined using the exact hard-spheres scattering method. The most compact conformations are slightly smaller than this calculated cross-section, indicating the gas-phase protein being slightly more compact than in solution. An extended (near-linear) "string" cross-section can be obtained by setting all Φ and Ψ angles to 180$^o$ (except those between Cys14 and Cys17 where the haem group is covalently bound), and adjusting the torsion angles near the proline residues.[154]

The Robinson research group (University of Cambridge, U.K.) focus on transferring large macromolecular assemblies into the gas phase and studying structure and dissociation mechanisms including: the chaperonin GroEL 14mer (~800 kDa);[130] the small heat shock protein *Ta*HSP16.9 (~200 kDa);[155] the heteromeric adenosine 5′-triphosphate (ATP)-binding cassette transporter $BtuC_2D_2$ (~129.5 kDa);[156] the eleven-membered ring topology of the *trp* RNA binding protein, TRAP (~88 kDa);[135] amyloid-β protein tetramer and dodcamer oligomers;[157] and the tetrameric complex transthyretin (~56 kDa;).[111, 158] They have contributed significantly to the field of macromolecular structures by MS and IM-MS and have modified instruments[16] such as to reduce the quadrupole analyser RF frequency to permit selection and transmission of high *m/z* ions. They have produced two

superb *Nature Protocols*: one for determining the stoichiometry and interactions of macromolecular assemblies by mass spectrometry,[159] that "walks through" procedures, capillary preparation, common pitfalls and troubleshooting for analysing large biological molecules. The other protocol[111] is invaluable for users of the Synapt HDMS instrument as it describes data collection and interpretation for large protein assemblies using IM-MS.

The Heck research group (University of Utrecht, The Netherlands) have also modified a QTOF instrument[131] to improve its performance in studying macromolecular assemblies by mass spectrometry (Figure 1.17).

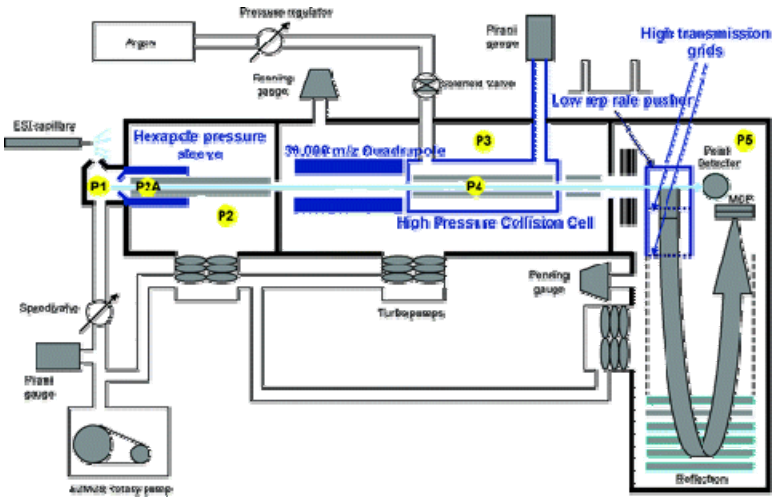



*Figure 1.17.[131] Heck's modified QTOF1 allowing large macromolecular protein assemblies to be studied by mass spectrometry.*

Pressures in the four regions have been measured: P1, $1 \times 10^{1}$ mbar; P2A, $8 \times 10^{-3}$ mbar; P2; $4 \times 10^{-3}$ mbar; P3, $6.7 \times 10^{-4}$ mbar; P4, $1.5 \times 10^{-2}$ mbar; and P5, $2 \times 10^{-6}$ mbar. Two transmission meshes were changed from 1000 to 200 lines $inch^{-1}$, leading to a three-fold sensitivity enhancement in the TOF region. Using this instrument, they have interrogated: the gas-phase dissociation of the $GroEL_{14}$ oligomer (801 kDa) upon successful transmission at collision cell pressures of $2 \times 10^{-2}$ mbar; the gas-phase stability of the GroEL-GroES chaperonin complex (~875 kDa); flavin binding to vanillyl alcohol oxidase (~509 kDa); the gas-phase dissociation of bovine, porcine and human haemoglobins (~64.5 kDa);[160] the GroEL-gp31 complex (877kDa);[126] the gas-phase stability of a number of GroEL substrates;[161] and the thermal disassembly of co-chaperonins GroES and gp31 (heptamer assemblies of ~72 kDa and ~84 kDa respectively).[162]

The Scrivens research group (University of Warwick, U.K.) have recently utilised mass spectrometry to study proteins[66] and larger protein complexes including haemoglobin and some variants;[109, 163, 164] and also utilise TWIMS ion mobility-mass spectrometry as a filter for phosphorylated peptides in tryptic digests;[165] using desorption ionisation coupled to TWIMS technology to study pharmaceutical

products (composed of codeine, paracetamol, benzocaine of polyethylene glycol).[105] The group is active in comparing experimental cross-sections from their TWIMS instruments to absolute drift cell measurements, with scope to improve the calibration between measurements not obtained by drift cell experiments. They also perform *in silico* computations to compare mass spectrometry collision cross-sections to those obtained from NMR and XRC crystal data (section 1.6).

## 1.9 Chapter 1. References


1. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** 171, (4356), 737-738.

2. Franklin, R. E.; Gosling, R. G., Molecular Configuration in Sodium Thymonucleate. *Nature* **1953,** 171, (4356), 740-741.

3. Franklin, R. E.; Gosling, R. G., Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* **1953,** 172, (4369), 156-157.

4. Avery, O. T.; MacLeod, C. M.; McCarty, M., Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: Induction of Transformation by a Desoxyribonucleic Acid fraction isolated from Pneumococcus Type III. *J. Exp. Med.* **1944,** 79, (2), 137-158.

5. Sanger, F.; Air, G. M.; Barrell, B. G.; Brown, N. L.; Coulson, A. R.; Fiddes, J. C.; Hutchison, C. A.; Slocombe, P. M.; Smith, M., Nucleotide sequence of bacteriophage [phi]X174 DNA. *Nature* **1977,** 265, (5596), 687-695.

6. Casey, D. K., Human Genome Program, U.S. Department of Energy, Genomics and Its Impact on Science and Society: A 2008 Primer. **2008**.

7. Voet, D.; Voet, J. G., *Biochemistry*. Second ed.; John Wiley & Sons, Inc.: Chichester, U.K., 1995.

8. Branden, C.; Tooze, J., *Introduction to Protein Structure*. Second ed.; Garland Publishing: New York, U.S.A., 1999.

9. Edelstein, S. J., Patterns in the quinary structures of proteins. Plasticity and inequivalence of individual molecules in helical arrays of sickle cell hemoglobin and tubulin. *Biophys. J.* **1980,** 32, (1), 347-360.

10. Anfinsen, C. B., Principles that Govern the Folding of Protein Chains. *Science* **1973,** 181, (4096), 223-230.

11. Levinthal, C., Are There Pathways for Protein Folding? *J. Chim. Phys. Et Phys. Chim. Biol.* **1968,** 65, (1), 44-45.

12. Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J., Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs. *J. Am. Chem. Soc.* **2002,** 124, (44), 12952-12953.

13. Johnson, P. V.; Beegle, L. W.; Kim, H. I.; Eiceman, G. A.; Kanik, I., Ion Mobility Spectrometry in Space Exploration. *Int. J. Mass Spectrom.* **2007,** 262, (1-2), 1-15.

14. Wyttenbach, T.; Bowers, M. T., Intermolecular Interactions in Biomolecular Systems Examined by Mass Spectrometry. *Annu. Rev. Phys. Chem.* **2007,** 58, (1), 511-533.

15. Heck, A. J. R.; van den Heuvel, R. H. H., Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **2004,** 23, (5), 368-389.

16. Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V., Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chem. Rev.* **2007,** 107, (8), 3544-3567.

17. Domon, B.; Aebersold, R., Mass Spectrometry and Protein Analysis. *Science* **2006,** 312, (5771), 212-217.

18. Thomson, J. J., *Rays of Positive Electricity and Their Application to Chemical Analyses*. Longmans, Green and Company: London, U.K., 1913.

19. Fenn, J. B., Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chem. Int. Ed.* **2003,** 42, (33), 3871-3894.

20. Tanaka, K.; Waki, H.; Ido, Y.; Yoshida, Y.; Yoshida, T. In *Detection of High Mass Molecules by Laser Desorption time-of-flight Mass Spectrometry*, Second Japan-China Joint Symposium on Mass Spectrometry, Osaka, Japan, 1987; Osaka, Japan, 1987; pp 185-188.

21. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T., Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988,** 2, (8), 151-153.

22. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988,** 60, (20), 2299-2301.

23. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* **1991,** 63, (24), 1193A-1203A.

24. Hillenkamp, F.; Karas, M.; James, A. M., Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. In *Methods Enzymol.*, Academic Press: 1990; Vol. 193, pp 280-295.

25. Glückmann, M.; Karas, M., The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization. *J. Mass Spectrom.* **1999,** 34, (5), 467-477.

26. Krüger, R.; Karas, M., Formation and fate of ion pairs during MALDI analysis: anion adduct generation as an indicative tool to determine ionization processes. *J. Am. Soc. Mass Spectrom.* **2002,** 13, (10), 1218-1226.

27. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., Molecular Beams of Macroions. *J. Chem. Phys.* **1968,** 49, (5), 2240-2249.

28. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** 246, (4926), 64-71.

29. Yamashita, M.; Fenn, J. B., Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* **1984,** 88, (20), 4451-4459.

30. Wong, S. F.; Meng, C. K.; Fenn, J. B., Multiple charging in electrospray ionization of poly(ethylene glycols). *J. Phys. Chem.* **1988,** 92, (2), 546-550.

31. Cech, N. B.; Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **2001,** 20, (6), 362-387.

32. Rayleigh, J. W. S., *Philos. Mag.* **1882,** 14, 184-186.

33. Iribarne, J. V.; Thomson, B. A., On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **1976,** 64, (6), 2287-2294.

34. Iavarone, A. T.; Williams, E. R., Mechanism of Charging and Supercharging Molecules in Electrospray Ionization. *J. Am. Chem. Soc.* **2003,** 125, (8), 2319-2327.

35. Wilm, M. S.; Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Processes* **1994,** 136, (2-3), 167-180.

36. Wilm, M.; Mann, M., Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996,** 68, (1), 1-8.

37. Juraschek, R.; Dülcks, T.; Karas, M., Nanoelectrospray--more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* **1999,** 10, (4), 300-308.

38. Paul, W.; Steinwedel, H. S., Ein Neues Massenspektrometer Ohne Magnetfeld. *Z. Naturforsch* **1957,** 69, (8a), 448.

39. de Hoffman, E.; Stroobant, V., *Mass Spectrometry Principles and Application*. Second ed.; Wiley: Paris, 1999.

40. Stephens, W. E., A Pulsed Mass Spectrometer with Time Dispersion. *Phys. Rev.* **1946**, (69), 691.

41. Wiley, W. C.; McLaren, J. B., Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **1955,** 16, 1150.

42. Mamyrin, B. A.; Karataev, V. I.; Schmikk, D. V.; Zagulin, V., The Mass-Reflectron, a New Nonmagnetic Time of Flight Mass Spectrometer with High Resolution. *Sov. Phys. JETP* **1973,** 37, 45.

43. Castleberry, C. M.; Chou, C. W.; Limbach, P. A., Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry of Oligonucleotides. *Curr. Nucleic Acid Chem.* **2008,** 33, (10), 1-21.

44. Siuzdak, G., *Mass Spectrometry for Biotechnology*. Academic Press: New York, 1996.

45. Zombeck, M. V. http://hea-www.harvard.edu/HRC/HomePage.html.

46. Ganem, B.; Li, Y. T.; Henion, J. D., Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J. Am. Chem. Soc.* **1991,** 113, (16), 6294-6296.

47. Ganem, B.; Li, Y. T.; Henion, J. D., Observation of noncovalent enzyme-substrate and enzyme-product complexes by ion-spray mass spectrometry. *J. Am. Chem. Soc.* **1991,** 113, (20), 7818-7819.

48. Jennings, K. R., The changing impact of the collision-induced decomposition of ions on mass spectrometry. *Int. J. Mass Spectrom.* **2000,** 200, (1-3), 479-493.

49. Tang, X. J.; Brewer, C. F.; Saha, S.; Chernushevich, I.; Ens, W.; Standing, K. G.; Chait, B. T., Investigation of protein-protein noncovalent interactions in soybean agglutinin by electrospray ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1994,** 8, (9), 750-754.

50. Bowers, M. T.; Marshall, A. G.; McLafferty, F. W., Mass Spectrometry: Recent Advances and Future Directions. *J. Phys. Chem.* **1996,** 100, (31), 12897-12910.

51. Brodbelt, J. S., Analytical applications of ion-molecule reactions. *Mass Spectrom. Rev.* **1997,** 16, (2), 91-110.

52. Shukla, A. K.; Futrell, J. H., Tandem mass spectrometry: dissociation of ions by collisional activation. *J. Mass Spectrom.* **2000,** 35, (9), 1069-1090.

53. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **2004,** 39, (10), 1091-1112.

54. Benesch, J. L. P., Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (3), 341-348.

55. Roepstorff, P.; Fohlmann, J., Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectom.* **1984,** 11, 601.

56. Cooks, R. G.; Terwilliger, D. T.; Ast, T.; Beynon, J. H.; Keough, T., Surface modified mass spectrometry. *J. Am. Chem. Soc.* **2002,** 97, (6), 1583-1585.

57. Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L., Surface-Induced Dissociation of Small Molecules, Peptides, and Non-Covalent Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2008,** 19, (2), 190-208.

58. Wysocki, V. H.; Jones, C. M.; Galhena, A. S.; Blackwell, A. E., Surface-Induced Dissociation Shows Potential to Be More Informative Than Collision-Induced Dissociation for Structural Studies of Large Systems. *J. Am. Soc. Mass Spectrom.* **2008,** 19, (7), 903-913.

59. Blundell, T. L.; Johnson, L. N., *Protein Crystallography*. Academic: New York, 1976.

60. Wuthrich, K., Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* **1990,** 265, (36), 22059-22062.

61. Kelly, S. M.; Jess, T. J.; Price, N. C., How to Study Proteins by Circular Dichroism. *Biochim. et Biophys. Acta* **2005**, (1751), 119.

62. Griffith, W. P.; Kaltashov, I. A., Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry. *Biochem.* **2003,** 42, (33), 10024-33.

63. Kuczera, K., *Molecular Modeling in Peptide and Protein Analysis, review article in the Encyclopedia of Analytical Chemistry.* John Wiley & Sons, Ltd.: Chichester, U.K., 2000; p 5894-5930.

64. Wyttenbach, T.; von Helden, G.; Bowers, M. T., Gas-Phase Conformation of Biological Molecules: Bradykinin. *J. Am. Chem. Soc.* **1996,** 118, (35), 8355-8364.

65. Jin, L.; Barran, P. E.; Deakin, J. A.; Lyon, M.; Uhrin, D., Conformation of glycosaminoglycans by ion mobility mass spectrometry and molecular modelling. *Phys. Chem. Chem. Phys.* **2005,** 7, (19), 3464-71.

66. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H., Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **2008,** 22, (20), 3297-304.

67. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucl. Acids Res.* **2000,** 28, (1), 235-242.

68. Shvartsburg, A. A.; Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996,** 261, (1-2), 86-91.

69. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996,** 100, (40), 16082-16086.

70. Mack, E., Average Cross-sectional Areas of Molecules by Gaseous Diffusion Methods. *J. Am. Chem. Soc.* **1925,** 47, (10), 2468-2482.

71. von Helden, G.; Wyttenbach, T.; Bowers, M. T., Inclusion of a MALDI ion source in the ion chromatography technique: conformational information on polymer and biomolecular ions. *Int. J. Mass Spectrom. Ion Processes* **1995,** 146-147, 349-364.

72. Shvartsburg, A. A.; Hudgins, R. R.; Dugourd, P.; Jarrold, M. F., Structural Elucidation of Fullerene Dimers by High-Resolution Ion Mobility Measurements and Trajectory Calculation Simulations. *J. Phys. Chem. A* **1997,** 101, (9), 1684-1688.

73. Wittmer, D.; Chen, Y. H.; Luckenbill, B. K.; Hill, H. H., Electrospray Ionization Ion Mobility Spectrometry. *Anal. Chem.* **1994,** 66, (14), 2348-2355.

74. http://www.ncbi.nlm.nih.gov/pubmed/. In 2009.

75. Creaser, C. S.; Griffiths, J. R.; Bramwell, C. J.; Noreen, S.; Hill, C. A.; Thomas, C. L. P., Ion mobility spectrometry: a review. Part 1. Structural analysis by mobility measurement. *The Analyst* **2004,** 129, (11), 984-994.

76. Clemmer, D. E.; Jarrold, M. F., Ion Mobility Measurements and their Applications to Clusters and Biomolecules. *J. Mass Spectrom.* **1997,** 32, (6), 577-592.

77. Bohrer, B. C.; Merenbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E., Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem.* **2008,** 1, (1), 293-327.

78. Kaddis, C. S.; Loo, J. A., Native Protein MS and Ion Mobility: Large Flying Proteins with ESI. *Anal. Chem.* **2007,** 79, (5), 1778-1784.

79. Verbeck, G. F.; Ruotolo, B. T.; Sawyer, H. A.; Gillig, K. J.; Russell, D. H., A Fundamental Introduction to Ion Mobiltiy Mass Spectrometry Applied to the Analysis of Biomolecules. *J. Biomol. Tech.* **2002,** 13, (2), 56-61.

80. Mason, E. A.; McDaniel, E. W., *Transport Properties of Ions in Gases*. Wiley: New York, 1988.

81. Revercomb, H. E.; Mason, E. A., Theory of plasma chromatography/gaseous electrophoresis. Review. *Anal. Chem.* **1975,** 47, (7), 970-983.

82. Cohen, M. J.; Karasek, F. W., Plasma Chromatography - a New Dimension for Gas Chromatography and Mass Spectrometry. *J. Chromatogr. Sci.* **1970,** 8, (6).

83. Hagen, D. F., Characterization of isomeric compounds by gas and plasma chromatography. *Anal. Chem.* **1979,** 51, (7), 870-874.

84. Kemper, P. R.; Bowers, M. T., State-selected mobilities of atomic cobalt ions. *J. Am. Chem. Soc.* **1990,** 112, (8), 3231-3232.

85. Baker, E. S.; Bushnell, J. E.; Wecksler, S. R.; Lim, M. D.; Manard, M. J.; Dupuis, N. F.; Ford, P. C.; Bowers, M. T., Probing Shapes of Bichromophoric Metalâ^'Organic Complexes Using Ion Mobility Mass Spectrometry. *J. Am. Chem. Soc.* **2005,** 127, (51), 18222-18228.

86. Kemper, P. R.; Bowers, M. T., A hybrid double-focusing mass spectrometer--high-pressure drift reaction cell to study thermal energy reactions of mass-selected ions. *J. Am. Soc. Mass Spectrom.* **1990,** 1, (3), 197-207.

87. Wyttenbach, T.; Kemper, P. R.; Bowers, M. T., Design of a new electrospray ion mobility mass spectrometer. *Int. J. Mass Spectrom.* **2001,** 212, (1-3), 13-23.

88. Valentine, S. J.; Counterman, A. E.; Clemmer, D. E., Conformer-Dependent Proton-Transfer Reactions of Ubiquitin Ions. *J. Am. Soc. Mass Spectrom.* **1997,** 8, (9), 954-961.

89. Murray, M. M.; Bernstein, S. L.; Nyugen, V.; Condron, M. M.; Teplow, D. B.; Bowers, M. T., Amyloid β Protein: Aβ40 Inhibits Aβ42 Oligomerization. *J. Am. Chem. Soc.* **2009,** 131, (18), 6316-6317.

90. Teplow, D. B.; Lazo, N. D.; Bitan, G.; Bernstein, S. L.; Wyttenbach, T.; Bowers, M. T.; Baumketner, A.; Shea, J.; Urbanc, B.; Cruz, L.; Borreguero, J.; Stanley, H. E., Elucidating Amyloid beta-Protein Folding and Assembly: A Multidisciplinary Approach. *Acc. Chem. Res.* **2006,** 39, (9), 635-645.

91. Kemper, P. R.; Dupuis, N. F.; Bowers, M. T., A new, higher resolution, ion mobility mass spectrometer. *Int. J. Mass Spectrom.* **2009,** In Press, Corrected Proof.

92. Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E., Three-Dimensional Ion Mobility/TOFMS Analysis of Electrosprayed Biomolecules. *Anal. Chem.* **1998,** 70, (11), 2236-2242.

93. Lee, Y. J.; Hoaglund-Hyzera, C. S.; Srebalus Barnes, C. A.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E., Development of high-throughput liquid chromatography injected ion mobility quadrupole time-of-flight techniques for analysis of complex peptide mixtures. *J. Chrom. B* **2002,** 782, (1-2), 343-351.

94. Valentine, S. J.; Koeniger, S. L.; Clemmer, D. E., A Split-Field Drift Tube for Separation and Efficient Fragmentation of Biomolecular Ions. *Anal. Chem.* **2003,** 75, (22), 6202-6208.

95. Koeniger, S. L.; Merenbloom, S. I.; Valentine, S. J.; Jarrold, M. F.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E., An IMS-IMS Analogue of MS-MS. *Anal. Chem.* **2006,** 78, (12), 4161-4174.

96. Merenbloom, S. I.; Koeniger, S. L.; Valentine, S. J.; Plasencia, M. D.; Clemmer, D. E., IMS-IMS and IMS-IMS-IMS/MS for Separating Peptide and Protein Fragment Ions. *Anal. Chem.* **2006,** 78, (8), 2802-2809.

97. Moon, M. H.; Myung, S.; Plasencia, M.; Hilderbrand, A. E.; Clemmer, D. E., Nanoflow LC/Ion Mobility/CID/TOF for Proteomics: Analysis of a Human Urinary Proteome. *J. Proteome Res.* **2003,** 2, (6), 589-597.

98. Taraszka, J. A.; Gao, X.; Valentine, S. J.; Sowell, R. A.; Koeniger, S. L.; Miller, D. F.; Kaufman, T. C.; Clemmer, D. E., Proteome Profiling for Assessing Diversity: Analysis of Individual Heads of Drosophila melanogaster Using LC-Ion Mobility-MS. *J. Proteome Res.* **2005,** 4, (4), 1238-1247.

99. Taraszka, J. A.; Kurulugama, R. T.; Sowell, R. A.; Valentine, S. J.; Koeniger, S. L.; Arnold, R. J.; Miller, D. F.; Kaufman, T. C.; Clemmer, D. E., Mapping the Proteome of Drosophila melanogaster: Analysis of Embryos and Adult Heads by LC-IMS-MS Methods. *Journal of Proteome Research* **2005,** 4, (4), 1223-1237.

100. Kurulugama, R. T.; Nachtigall, F. M.; Lee, S.; Valentine, S. J.; Clemmer, D. E., Overtone Mobility Spectrometry: Part 1. Experimental Observations. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (5), 729-737.

101. Valentine, S. J.; Stokes, S. T.; Kurulugama, R. T.; Nachtigall, F. M.; Clemmer, D. E., Overtone Mobility Spectrometry: Part 2. Theoretical Considerations of Resolving Power. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (5), 738-750.

102. http://www.waters.com.

103. Luca, A.; Schlemmer, S.; Cermak, I.; Gerlich, D., On the combination of a linear field free trap with a time-of-flight mass spectrometer. *Rev. Sci. Instrum.* **2001,** 72, (7), 2900-2908.

104. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H., An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **2007,** 261, (1), 1-12.

105. Williams, J. P.; Scrivens, J. H., Coupling desorption electrospray ionisation and neutral desorption/extractive electrospray ionisation with a travelling-wave based ion mobility mass spectrometer for the analysis of drugs. *Rapid Commun. Mass Spectrom.* **2008,** 22, (2), 187-196.

106. Howdle, M. D.; Eckers, C.; Laures, A. M. F.; Creaser, C. S., The Use of Shift Reagents in Ion Mobility-Mass Spectrometry: Studies on the Complexation of an Active Pharmaceutical Ingredient with Polyethylene Glycol Excipients. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (1), 1-9.

107. Kaur-Atwal, G.; Weston, D. J.; Green, P. S.; Crosland, S.; Bonner, P. L. R.; Creaser, C. S., Analysis of tryptic peptides using desorption electrospray ionisation combined with ion mobility spectrometry/mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007,** 21, (7), 1131-1138.

108. Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V., Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chem. Int. Ed. Engl.* **2007,** 46, (42), 8001-4.

109. Scarff, C. A.; Patel, V. J.; Thalassinos, K.; Scrivens, J. H., Probing Hemoglobin Structure by Means of Traveling-Wave Ion Mobility Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (4), 625-631.

110. Database, C. C.-s., http://www.indiana.edu/~clemmer.

111. Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V., Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **2008,** 3, (7), 1139-52.

112. Shvartsburg, A. A.; Smith, R. D., Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **2008,** 80, (24), 9689-9699.

113. Buryakov, I. A.; Krylov, E. V.; Nazarov, E. G.; Rasulev, U. K., A new method of separation of multi-atomic ions by mobility at atmospheric pressure using a high-frequency amplitude-asymmetric strong electric field. *Int. J. Mass Spectrom. Ion Processes* **1993,** 128, (3), 143-148.

114. Guevremont, R., High-field Asymmetric Waveform Ion Mobiltiy Spectrometry (FAIMS). *Can. J. Anal. Sci. Spectrosc.* **2004,** 49, (3), 105-113.

115. www.faims.com. **2009**.

116. Purves, R. W.; Barnett, D. A.; Ells, B.; Guevremont, R., Gas-phase conformers of the [M + 2H](2+) ion of bradykinin investigated by combining high-field asymmetric waveform ion mobility spectrometry, hydrogen/deuterium exchange, and energy-loss measurements. *Rapid Commun. Mass Spectrom.* **2001,** 15, (16), 1453-6.

117. Purves, R. W.; Barnett, D. A.; Guevremont, R., Separation of protein conformers using electrospray-high field asymmetric waveform ion mobility spectrometry-mass spectrometry. *Int. J. Mass Spectrom.* **2000,** 197, (1-3), 163-177.

118. Shvartsburg, A. A.; Li, F.; Tang, K.; Smith, R. D., Characterizing the Structures and Folding of Free Proteins Using 2-D Gas-Phase Separations: Observation of Multiple Unfolded Conformers. *Anal. Chem.* **2006,** 78, (10), 3304-3315.

119. Purves, R. W.; Barnett, D. A.; Ells, B.; Guevremont, R., Investigation of bovine ubiquitin conformers separated by high-field asymmetric waveform ion mobility spectrometry: cross section measurements using energy-loss experiments with a triple quadrupole mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2000,** 11, (8), 738-745.

120. Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **1997,** 16, (1), 1-23.

121. Breuker, K., The study of protein-ligand interactions by mass spectrometry - a personal view. *Int. J. Mass Spectrom.* **2004,** 239, (1), 33-41.

122. Kaltashov, I. A.; Eyles, S. J., Studies of biomolecular conformations and conformational dynamics by mass spectrometry. *Mass Spectrom. Rev.* **2002,** 21, (1), 37-71.

123. Winston, R. L.; Fitzgerald, M. C., Mass spectrometry as a readout of protein structure and function. *Mass Spectrom. Rev.* **1997,** 16, (4), 165-179.

124. Jarrold, M. F., Peptides and Proteins in the Vapor Phase. *Annu. Rev. Phys. Chem.* **2000,** 51, (1), 179-207.

125. Kaddis, C. S.; Lomeli, S. H.; Yin, S.; Berhane, B.; Apostol, M. I.; Kickhoefer, V. A.; Rome, L. H.; Loo, J. A., Sizing Large Proteins and Protein Complexes by Electrospray Ionization Mass Spectrometry and Ion Mobility. *J. Am. Soc. Mass Spectrom.* **2007,** 18, (7), 1206-1216.

126. van Duijn, E.; Bakkes, P. J.; Heeren, R. M.; van den Heuvel, R. H.; van Heerikhuizen, H.; van der Vies, S. M.; Heck, A. J., Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat. Methods* **2005,** 2, (5), 371-6.

127. Sobott, F.; McCammon, M. G.; Hernandez, H.; Robinson, C. V., The flight of macromolecular complexes in a mass spectrometer. *Philos. Transact. A Math Phys. Eng. Sci.* **2005,** 363, (1827), 379-89; discussion 389-91.

128. Ruotolo, B. T.; Robinson, C. V., Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **2006,** 10, (5), 402-8.

129. Benesch, J. L. P.; Robinson, C. V., Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Curr. Opin. Struct. Biol.* **2006,** 16, (2), 245-251.

130. Sobott, F.; Robinson, C. V., Characterising electrosprayed biomolecules using tandem-MS--the noncovalent GroEL chaperonin assembly. *Int. J. Mass Spectrom.* **2004,** 236, (1-3), 25-32.

131. van den Heuvel, R. H. H.; van Duijn, E.; Mazon, H.; Synowsky, S. A.; Lorenzen, K.; Versluis, C.; Brouns, S. J. J.; Langridge, D.; van der Oost, J.; Hoyes, J.; Heck, A. J. R., Improving the Performance of a Quadrupole Time-of-Flight Instrument for Macromolecular Mass Spectrometry. *Anal. Chem.* **2006,** 78, (21), 7473-7483.

132. Liepold, L.; Oltrogge, L. M.; Suci, P. A.; Young, M. J.; Douglas, T., Correct Charge State Assignment of Native Electrospray Spectra of Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (3), 435-442.

133. Rostom, A. A.; Fucini, P.; Benjamin, D. R.; Juenemann, R.; Nierhaus, K. H.; Hartl, F. U.; Dobson, C. M.; Robinson, C. V., Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, (10), 5185-5190.

134. Breuker, K.; McLafferty, F. W., Stepwise evolution of protein native structure with electrospray into the gas phase, 10e-12 to 10e-2 s. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** 105, (47), 18145-18152.

135. Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V., Evidence for macromolecular protein rings in the absence of bulk water. *Science* **2005,** 310, (5754), 1658-61.

136. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., Naked Protein Conformations: Cytochrome c in the Gas Phase. *J. Am. Chem. Soc.* **1995,** 117, (40), 10141-10142.

137. Jurchen, J. C.; Williams, E. R., Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. *J. Am. Chem. Soc.* **2003,** 125, (9), 2817-2826.

138. Thalassinos, K.; Slade, S. E.; Jennings, K. R.; Scrivens, J. H.; Giles, K.; Wildgoose, J.; Hoyes, J.; Bateman, R. H.; Bowers, M. T., Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *Int. J. Mass Spectrom.* **2004,** 236, (1-3), 55-63.

139. Jurchen, J. C.; Garcia, D. E.; Williams, E. R., Further studies on the origins of asymmetric charge partitioning in protein homodimers. *J. Am. Soc. Mass Spectrom.* **2004,** 15, (10), 1408-1415.

140. Chen, Y.-L.; Collings, B. A.; Douglas, D. J., Collision cross sections of myoglobin and cytochrome c ions with Ne, Ar, and Kr. *J. Am. Soc. Mass Spectrom.* **1997,** 8, (7), 681-687.

141. Badman, E. R.; Myung, S.; Clemmer, D. E., Evidence for Unfolding and Refolding of Gas-Phase Cytochrome c Ions in a Paul Trap. *J. Am. Soc. Mass Spectrom.* **2005,** 16, (9), 1493-1497.

142. Li, Y. T.; Hsieh, Y. L.; Henion, J. D.; Ganem, B., Studies on heme binding in myoglobin, hemoglobin, and cytochrome c by ion spray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1993,** 4, (8), 631-637.

143. Babu, K. R.; Moradian, A.; Douglas, D. J., The methanol-induced conformational transitions of beta-lactoglobulin, cytochrome c, and ubiquitin at low pH: a study by electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2001,** 12, (3), 317-28.

144. Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E., Monitoring structural changes of proteins in an ion trap over approximately 10-200 ms: unfolding transitions in cytochrome c ions. *Anal. Chem.* **2001,** 73, (24), 6000-7.

145. Breuker, K.; McLafferty, F. W., Native electron capture dissociation for the structural characterization of noncovalent interactions in native cytochrome c. *Angew. Chem. Int. Ed. Engl.* **2003,** 42, (40), 4900-4.

146. Breuker, K.; McLafferty, F. W., The thermal unfolding of native cytochrome c in the transition from solution to gas phase probed by native electron capture dissociation. *Angew. Chem. Int. Ed. Engl.* **2005,** 44, (31), 4911-4.

147. Steinberg, M. Z.; Elber, R.; McLafferty, F. W.; Gerber, R. B.; Breuker, K., Early Structural Evolution of Native Cytochrome c after Solvent Removal. *Chembiochem* **2008,** 9, (15), 2417-2423.

148. Morar, A. S.; Olteanu, A.; Young, G. B.; Pielak, G. J., Solvent-induced collapse of {alpha}-synuclein and acid-denatured cytochrome c. *Protein Sci.* **2001,** 10, (11), 2195-2199.

149. Janfelt, C.; Talaty, N.; Mulligan, C. C.; Keil, A.; Ouyang, Z.; Cooks, R. G., Mass spectra of proteins and other biomolecules recorded using a handheld instrument. *Int. J. Mass Spectrom.* **2008,** 278, (2-3), 166-169.

150. Henderson, S. C.; Valentine, S. J.; Counterman, A. E.; Clemmer, D. E., ESI/ion trap/ion mobility/time-of-flight mass spectrometry for rapid and sensitive analysis of biomolecular mixtures. *Anal. Chem.* **1999,** 71, (2), 291-301.

151. Zhao, T.; King, F. L., Direct Determination of the Primary Binding Site of Cisplatin on Cytochrome c by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (6), 1141-1147.

152. Faull, P. A.; Korkeila, K. E.; Kalapothakis, J. M.; Gray, A.; McCullough, B. J.; Barran, P. E., Gas-phase metalloprotein complexes interrogated by ion mobility-mass spectrometry. *Int. J. Mass Spectrom.* **2009,** 283, (1-3), 140-148.

153. Clemmer, D. E. http://www.indiana.edu/~clemmer.

154. Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **1997,** 119, (9), 2240-2248.

155. Benesch, J. L.; Sobott, F.; Robinson, C. V., Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry. *Anal. Chem.* **2003,** 75, (10), 2208-14.

156. Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V., Micelles Protect Membrane Complexes from Solution to Vacuum. *Science* **2008,** 321, (5886), 243-246.

157. Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T., Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **2009,** 1, (4), 326-331.

158. Sobott, F.; McCammon, M. G.; Robinson, C. V., Gas-phase dissociation pathways of a tetrameric protein complex. *Int. J. Mass Spectrom.* **2003,** 230, (2-3), 193-200.

159. Hernandez, H.; Robinson, C. V., Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protocols* **2007,** 2, (3), 715-726.

160. Versluis, C.; Heck, A. J. R., Gas-phase Dissociation of Hemoglobin. *Int. J. Mass Spectrom.* **2001,** 210-211, 637-649.

161. van Duijn, E.; Simmons, D. A.; van den Heuvel, R. H.; Bakkes, P. J.; van Heerikhuizen, H.; Heeren, R. M.; Robinson, C. V.; van der Vies, S. M.; Heck, A. J., Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J. Am. Chem. Soc.* **2006,** 128, (14), 4694-702.

162. Geels, R. B. J.; Calmat, S.; Heck, A. J. R.; van der Vies, S. M.; Heeren, R. M. A., Thermal activation of the co-chaperonins GroES and gp31 probed by mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008,** 22, (22), 3633-3641.

163. Williams, J. P.; Scrivens, J. H.; Green, B. N.; Farrar, L. M.; Sutcliffe, M., Hb Leeds [beta 56 (D7) Gly - Cys]: a new hemoglobin that aggrevates anemia in a child with beta (0) thalassemia trait. *Hemoglobin* **2007,** 31, (3), 367-373.

164. Williams, J. P.; Giles, K.; Green, B. N.; Scrivens, J. H.; Bateman, R. H., Ion mobility augments the utility of mass spectrometry in the identification of human hemoglobin variants. *Rapid Commun. Mass Spectrom.* **2008,** 22, (20), 3179-3186.

165. Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H., Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* **2008,** 81, (1), 248-254.

# Chapter 2. Experimental

*Chapter 2 describes the MoQTOF instrument utilised for ion mobility experiments including data acquisition and analysis. Schematics of the ion mobility-mass spectrometer are shown with experimental tuning parameters that produced optimal signal. Protein preparation and spray conditions are also detailed.*

## 2.1 The MoQTOF Instrument

The MoQTOF (Mobility Quadrupole Time-of-Flight)[1] instrument was built in-house by Drs. B. McCullough and P. Barran as part of a collaboration with Waters Micromass and Prof. M. Bowers (University of California, Santa Barbara, U.S.A.). Completed in late 2005 / early 2006, the instrument was initially used to study small peptide and protein systems ($M_R$ < ~8 kDa)[2] with aims to use the instrument on a wide variety of molecules over a large dynamic range. This Thesis presents work that has developed and extended instrument use to enable characterisation of a number of biomolecular systems. In order to look at more native, large macromolecules this has required extensive tuning of the source; voltage, electronic and pressure conditions; drift cell elements and parameters; and analyser and detector settings.

The layout of a standard QTOF instrument (Chapter 1) begins with a source region where sample ionisation occurs; a transfer region of ion optics for ion focussing; analyser(s) region(s) for *m/z* separation; and finally a detector region. Between the ion optic transfer and first (quadrupole) analyser region, the ion mobility drift cell chamber was created in the MoQTOF. A diagram showing all regions in the MoQTOF and labelled parts A to N is shown in Figure 2.1.

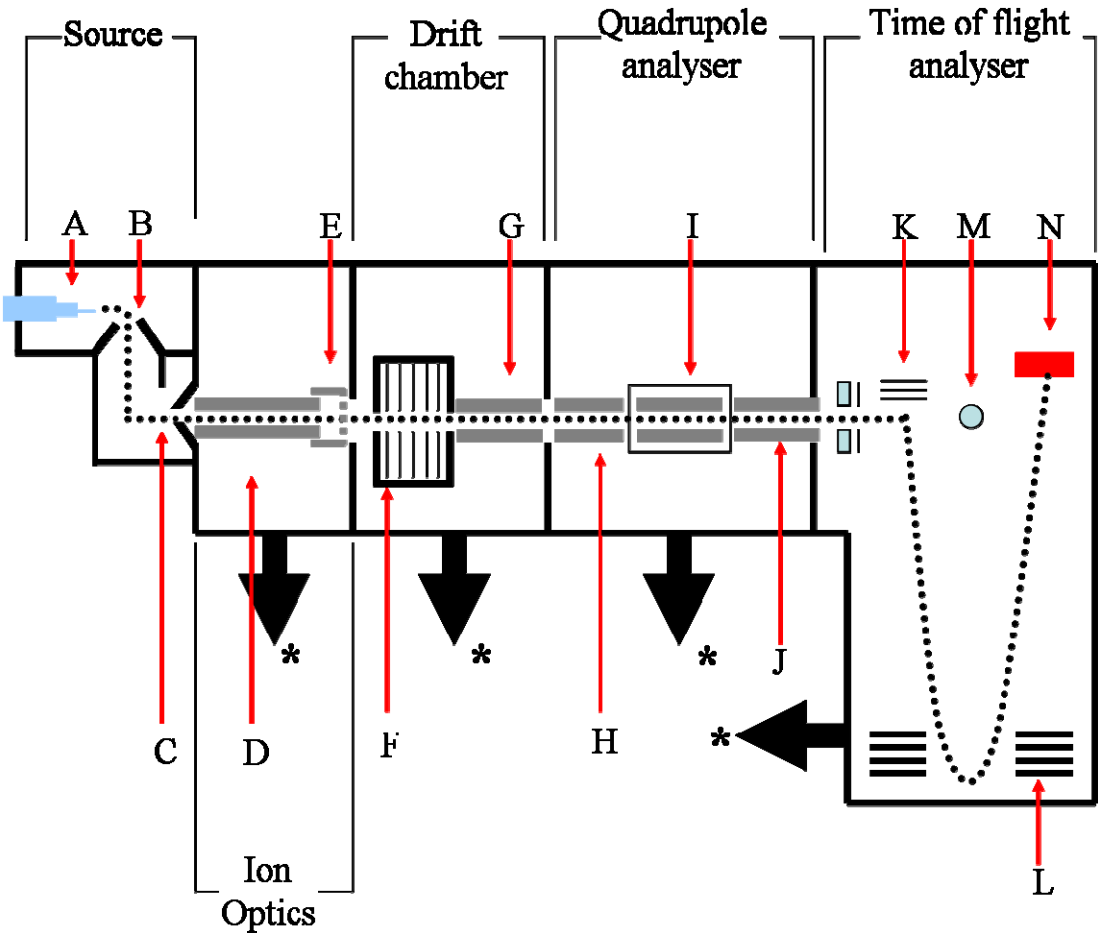



*Figure 2.1. Diagram of the MoQTOF showing source, ion optics, drift chamber, quadrupole analyser, and time of flight analyser regions. Labelled parts: A – electrospray capillary; B – source cone; C – extractor cone; D – pre-cell hexapole; E – top hat lens; F – drift cell; G – post-cell hexapole; H – pre-collision cell quadrupole; I – collision cell; J – post-collision cell hexapole; K – pusher stack; L – reflectron; M – point detector (for quadrupole analyser); N – MCP detector (for time of flight analyser); * vacuum regions pumped by turbomolecular pumps.*

A more detailed schematic of the MoQTOF is presented in Figure 2.2 with structures A to G labelled showing the linear arrangement of mass spectrometer components leading up to the time-of-flight analyser region.

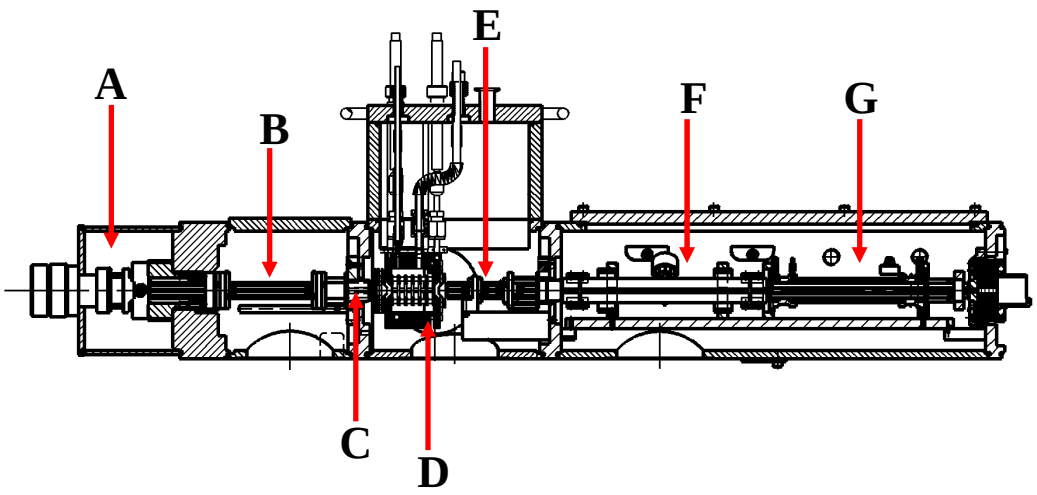



*Figure 2.2. Schematic of the MoQTOF instrument showing: A – source region; B – pre-cell hexapole; C – top hat lens; D - ion mobility copper drift cell; E – post-cell hexapole; F – quadrupole analyser; and G – collision cell.*

### 2.1.1 The Pre-Cell Hexapole and The Top Hat Lens

Electrospray ionisation produces a constant beam of ionised molecules.[3-5] Ion mobility-mass spectrometry characterisation requires discrete packets of ions to be pulsed into the drift cell. This is achieved by storing ions in the pre-cell hexapole by raising the voltage on a top-hat (TH1; typically held 30V above the H1 voltage) lens attached to the end of the pre-cell hexapole and pulsing this voltage (typical frequency range 40 to 80 Hz; typical pulse widths 10 to 40 µs) to allow ions to leave the storage region through a 6 mm orifice in TH1, and enter the drift cell.

### 2.1.2 The Drift Cell

The basic drift cell make-up is: 1) a three component Einzel lens before the cell; 2) a copper drift cell containing five drift rings; 3) molybdenum entrance and exit orifices; and 4) a copper end-cap with an exit lens proceeding it. Each of these four components will be discussed separately.

The Einzel lens is made up of three stainless steel lenses (L1, L2 and L3) with dimensions 40.6 mm outer diameter (o.d.) × 22.4 mm inner diameter (i.d.) × 3.8 mm width. Each lens is supplied voltage separately and are all mounted on ceramic rods with additional ceramic spacers so as to maintain electrical isolation and for optimal alignment. L1 acts as an initial guiding lens for ions from the pre-cell hexapole. L2 acts as a steering lens and has been subdivided into four sections to allow x and y beam directional steering. The voltage supplied to L2 is split on each pair ($x_{left}$ and $x_{right}$; $y_{top}$ and $y_{bottom}$) of sections using a variable resistor, with the sum of the section pair voltage equalling the value of L2. L3 acts as a focusing lens. It is adjacent to the cell entrance orifice and is tapered to a final inner diameter of 7.6 mm.

The drift cell is housed within the drift chamber of the MoQTOF instrument which is pumped to vacuum using a 500 L $s^{-1}$ Pfeiffer TMH520 turbomolecular pump (Pfeiffer Vacuum Ltd., Newport Pagnell, UK) backed by an Edwards E2M40 two stage oil sealed rotary vane vacuum pump (Edwards, Crawley, West Sussex, UK). The top hat lens projects into the drift chamber from the source chamber so as to come into close proximity with the Einzel lens. A short post-cell hexapole is housed within the drift chamber and transfers mobility separated ions from the drift cell into the quadrupole analyser region. The drift cell is flange-mounted on vertical stainless steel rods at the four corners of the cell. Voltage supplies for the drift cell components, heaters and thermocouples are *via* a feedthrough on the top flange. The main region of ion mobility separation occurs within the copper drift cell (length 65.5mm, width 88.9 mm and height 88.9 mm) due to the presence of five drift rings. These rings (30.5 mm o.d. × 15.2 mm i.d. × 3.2 mm) are mounted from the end cap on ceramic rods and separated by 5.2 mm ceramic spacers. They are connected in series by 1 MΩ resistors thereby maintaining a linear drift field as required for low field ion mobility measurements. The cell can be heated resistively using ten imbedded ceramic heaters (71 mm ceramic rods threaded with tantalum wire) in the cell walls, or cooled by a stream of nitrogen gas that passes through channels cut into the cell block. Temperature is monitored by a k-type thermocouple (read on an Omega CN1001TC thermocouple controller) attached to the top of the block.

Entrance and exit orifices (30.5 mm o.d. × 1.0 mm i.d.) are made from molybdenum discs and held in position by copper retainer rings (30.5 mm o.d. × 15.2 mm i.d. × 1.6 mm width).

The main cell body is electrically isolated from the copper end-cap by a ceramic spacer ring permitting these two regions to have a voltage applied to them to create the drift potential. The end-cap is the attachment point for the ceramic rods that support the drift rings. It can be heated *via* a pair of heaters (50 mm) or cooled by a stream of nitrogen gas through channels cut into the end-cap. A final lens (L4) is attached to the end-cap to focus ions leaving the cell into the post-cell hexapole.

### 2.1.3 Buffer Gas

Buffer gas (helium; 99.999%; BOC Gases, Surrey, UK) is passed through a moisture filter (to remove trace water, oxygen and hydrocarbons potentially present in the cylinder) and enters the drift cell top flange *via* a flexible stainless steel bellow. The pressure within the cell is measured using a MKS Baratron (MKS Instruments, Andover, Massachusetts, U.S.A.) providing a readout from 0 to 5 Torr.

## 2.2 Nano-Electrospray Ionisation

All experiments were performed using nano-electrospray ionisation[5] (nano-ESI) permitting smaller sample volumes to be consumed over longer time periods when compared with traditional electrospray ionisation *e.g.* 10 µL protein samples sprayed for up to three hours (~3.3 µL $hr^{-1}$ flow rate

or less) *versus* typically 80 µL $hr^{-1}$ flow rate in traditional ESI. As detailed in the Introduction Chapter, nano-ESI produces more desolvated ions and native-like charge states (lower charged) can be observed. Figure 2.3 shows mass spectra of 20 µM cytochrome *c* solution sprayed from 50 mM ammonium acetate (pH 6.85) from an ESI source (3.5 kV capillary voltage; black line) and the same sample diluted five fold to 4 µM from a nano-ESI source (1.5 kV capillary voltage; red line).

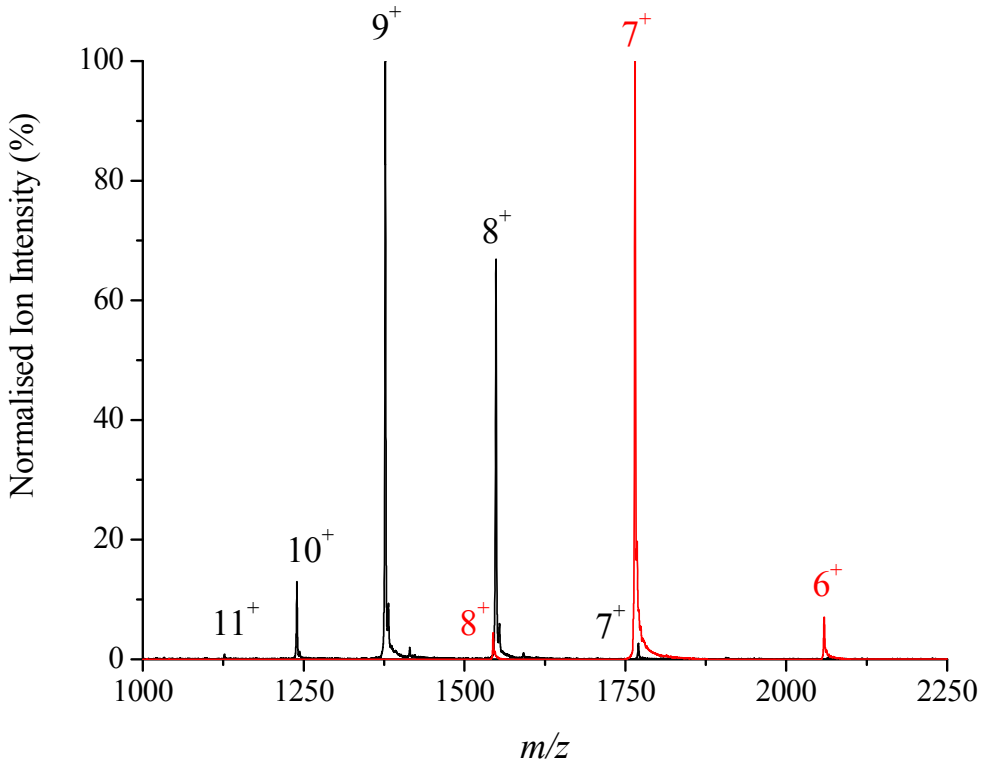



*Figure 2.3. Mass spectrum of native-like cytochrome c sprayed from ESI (3.5 kV capillary voltage, black line) and nano-ESI sources (1.5 kV capillary voltage, red line).*

Both species show narrow charge state distributions indicating compact conformers. The nano-ESI source produces only three charge states (z = $6^+$ to $8^+$) from a more dilute solution, whereas the ESI source produces five charge states (z = $7^+$ to $11^+$). The base peak of each spectrum differs by two charge states, an indicator that more native-like conditions can be obtained using nano-ESI due to the smaller capillary voltage applied to the capillary tip, as well as the possibility of using lower sample concentrations.

Capillary tips were produced in-house from 1.2 mm interior diameter borosilicate glass capillaries (Kwik-Fil, World Precision Instruments, Sarasota, Florida, U.S.A.). Tips were pulled using a Flaming/Brown Micropipette puller (Model P-97, Sutter Instrument Company, Novato, California, USA) to produce tips of length ~ 4 to 5 cm with end point tip diameters of 10 – 30 µm. By varying

temperature, speed and pulling time, an optimal tip-pulling program for generating protein signal was generated. This program was maintained for approximately one year before the heating filament was replaced, and the tip-pulling instrument was tuned to reproduce the original tip shape. It was possible to generate various programs which produced different tip shapes and lengths, permitting signal optimisation for individual samples. Nano-ESI also affords the use of lower capillary potentials to be applied to samples. This is of benefit when attempting to study native-like proteins where the use of the lowest capillary potential possible is sought. Typically, potentials less than 2.0 kV are required for studying systems sprayed by nano-ESI whereas a potential of more than 2.5 kV is required for ESI. The tip housing was secured to an adjustable xyz-axes stage permitting careful and reproducible positioning of the capillary tip with respect to the sample cone. The Micromass Z-spray source has an off-axis geometry, assisting ion sensitivity and transmission by restricting transfer of neutral species or solvent into the instrument. The ion flight path curves by 90$^{o}$ relative to the capillary into the sample cone and then is guided back through 90$^{o}$ to the extractor cone thereby tracing out the approximate "z" flight path that it derives its name from.

## 2.3 Source Conditions on The MoQTOF

Careful manipulation of mass spectrometer source conditions enables production of good signal intensity for the ion of interest. Manual control over a number of parameters such as source temperature, pressure, sample and extractor cones and the pre-cell hexapole and top hat lens assembly allows tuning to maximise signal. To assist desolvation, a source temperature range of 80 - 100$^{o}$C was used. With increasing biomolecular size, increasing pressure is required to transfer intact molecules to the gas phase.[6, 7] To achieve this, the instrument was modified to allow direct pumping of the region behind the instrument entrance (extractor cone) by throttling of the a source rotary pump (Edwards E28) through insertion of a Speedivalve (SP25, Edwards, Crawley, West Sussex, United Kingdom) in-line isolation valve to the vacuum line. When fully open, a source pressure of $8 \times 10^{-3}$ mbar could be achieved with the instrument at base pressure. This pressure is read using a Pirani gauge above an Edwards E28 roughing pump which backs the original QTOF instrument turbomolecular pumps. To assist pumping in the drift chamber region, an Edwards E2M40 rotary vane dual stage mechanical pump (Edwards, Crawley, West Sussex, United Kingdom) backs the drift chamber turbomolecular pump. When helium was present in the drift cell, a booster fitted to the E2M40 rotary is operated to maintain pressures in the quadrupole analyser region below ~$5.0 \times 10^{-6}$ mbar (base pressure in absence of helium, ~$1.0 \times 10^{-6}$ mbar) as measured by a Penning gauge. The drift chamber pressure is analysed by a Penning gauge and reads ~$5.0 \times 10^{-3}$ mbar in the presence of ~3.0 Torr of helium ($8.0 \times 10^{-6}$ at base pressure). The pressure in the TOF region is measured *via* a Penning gauge connected to the flight tube which typically reads ~$8.0 \times 10^{-7}$ mbar in the presence of ~3 Torr of helium within the drift cell (base pressure, ~$1.0 \times 10^{-7}$ mbar). Varying the source pressure to optimise large macromolecule signal could be increased above ~$8 \times 10^{-1}$ mbar, however strain on the turbomolecular pumps in the analyser and TOF regions occurred, thereby raising the pressure in these regions above

acceptable levels. The pressure in the TOF chamber must be kept below ~$5.0 \times 10^{-6}$ mbar so as to not damage sensitive equipment parts located in this region. These pressures are summarised in Table 2.1.

*Table 2.1. Instrument pressures at base pressure and with ~3.0 Torr of helium in the drift cell.*

| Region | Base pressure (mbar) | Pressure with ~3.0 Torr helium in the drift cell (mbar) |
|---|---|---|
| Source (Speedivalve fully open) | $8.0 \times 10^{-3}$ mbar | ~$2.0 \times 10^{-2}$ mbar |
| Drift chamber | $8.0 \times 10^{-6}$ mbar | ~$5.0 \times 10^{-3}$ mbar |
| Quadrupole Analyser | $1.0 \times 10^{-6}$ mbar | ~$5.0 \times 10^{-6}$ mbar |
| Time-of-Flight Analyser | $1.0 \times 10^{-7}$ mbar | ~$8.0 \times 10^{-7}$ mbar |

## 2.4 Mass Spectrometry Mode Settings

Prior to modifying the MoQTOF, the initial QTOF 1 instrument provided robust analytical power as a traditional mass spectrometer. A key component in the instrument development was to not lose the ability for the instrument to function as a "traditional" mass spectrometer *i.e.* when not in ion mobility mode. This was achieved by placing a short 10 cm flange-mounted hexapole between the pre- and post-cell hexapoles. The RF is supplied from the same generator as the post-cell hexapole, DC is supplied by an external power supply (at an equal voltage to the post-cell hexapole), and the Speedivalve throttle is opened fully. Low sample and extractor cone voltages, typically 50V and 5V respectively, and source temperatures of 80 – 100$^{o}$C provide sufficient desolvation of gas-phase ions.

## 2.5 Ion Mobility-Mass Spectrometry Mode Settings

All ion mobility electrical components are controlled from a power supply designed in-house. All voltages are set manually with reference to the voltage applied to the collision cell (CV; ground – typically 4V), and all electrically float on top of this. Lens L4 is referenced to the end cap (designated C2) voltage, and all voltages to the source-end of the instrument are referenced to the main drift cell body (C1). Figure 2.4 is a schematic of the voltages showing their relationship to one another.

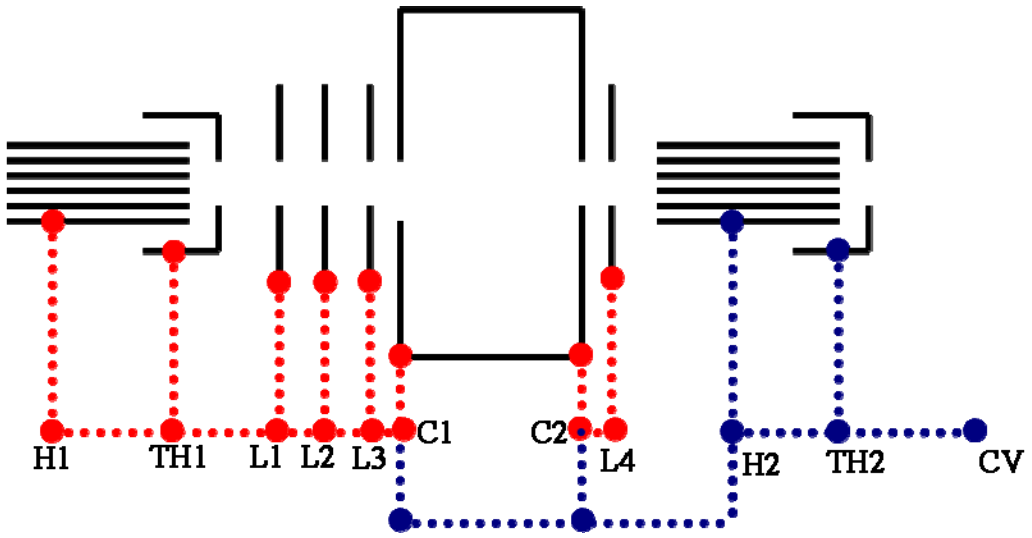



*Figure 2.4 MoQTOF drift cell and hexapole voltages schematic. The horizontal lines indicate ground for each voltage e.g. for L2, the ground voltage is the collision voltage (CV). Voltages indicated by a red dotted line are manually set by the user and supplied externally; those in blue are set by the instrument software and are supplied by the instrument power supply.*

The sample cone floats on top of the extractor cone voltage, which must be increased by the user as this is software controlled. Drift voltages for ion mobility experiments are described as the potential applied between C1 and C2. Table 2.2 contains typical voltages applied for studying denatured cytochrome *c* protein.

*Table 2.2. Typical drift cell and source electronic settings for an ion mobility experiment on the MoQTOF. Voltages for H1, TH1, L1, L2 and L3 all float above voltage C1 and this is shown in the Table as the applied voltage, and the voltage + C1 voltage (indicated as C1 = voltage). Increasing C1 by 5V will increase all this voltages by 5V. L4 floats on top of C2.C1 and C2 float on top of the collision voltage (CV).*

| Tuning Parameter | Voltage (V) |
|---|---|
| Sample Cone | 80 (on C1 = 150) |
| Extractor Cone | 50 (on C1 = 120) |
| H1 | 35 (on C1 = 105) |
| TH1 | 20 (on C1 = 90) |
| L1 | -40 (on C1 = 30) |
| L2 | -60 (on C1 = 10) |
| L3 | 20 (on C1 = 90) |
| C1 | 70 (ref. = CV) |
| C2 | 10 (ref. = CV) |
| L4 | 1 (-9 on C2) |

## 2.6 Pulser Unit

To provide discrete ion packets for analysis by IM-MS, the voltage on the top hat (TH1) lens was pulsed using a pulser unit. The pulser unit required the TH1 voltage as an input, as well as a TTL (transistor-transistor logic) pulse from a signal generator (Stanford Research Systems, Sunnyvale, California, U.S.A.). Another signal (with a period 1 μs longer than the pusher period) was sent to the embedded computer time-to-digital converter (TDC) card within the instrument so as to trigger the start of the experiment. Pulse widths were fixed at 40 μs for all experiments, with pulse frequencies varying depending upon the time-of-flight pusher period. Ion mobility experiments are composed of 200 "pushes" of the time-of-flight pusher (the high voltage plate that orthogonally propels ions into the flight tube). Therefore, the pulse frequency applied to the top hat must be 200 times the pusher period. This frequency could be increased by determining the lowest possible value of the pusher period for the *m/z* acquired over. When set too low, a pusher "spike" was observed in the mass spectrum. Increasing the pusher period by 1 to 2 μs caused this spike to disappear from the *m/z* range being interrogated and allowed the most efficient pulser frequency to be selected. Typically, 100 μs pusher periods were used to cover a range of 2000 *m/z*. This period would require a pulser frequency of 50 Hz. Larger mass ranges would require larger pusher periods and smaller pulser frequencies. It was therefore always a major consideration to have the smallest *m/z* range selected for the ions of interest to ensure that the pulser frequency was as high as possible to allow gas-phase ions to not accumulate for longer than necessary time periods in the hexapole. If trapped for extended periods of time, ions could potentially undergo gas-phase dissociation due to an excess of collisions with neighbouring ions.

## 2.7 Mass Spectrometer Calibration

The MoQTOF was calibrated weekly with a 2 mg $mL^{-1}$ solution of caesium iodide in 70% by volume isopropanol, providing a series of peaks from 132.91 ($Cs^+$) to ~4500 *m/z*.

## 2.8 Data Acquisition

Experimentally, a mass spectrum was initially obtained to determine the *m/z* range to interrogate, thereby also determining the pusher period and pulser frequency required to perform optimum ion mobility experiments. Tuning of source pressure, capillary voltage, sample and extractor cone voltages and x-y steering were also performed. Switching to ion mobility mode required selecting an option in the software (to receive the signal from the pulse generator to trigger the IM-MS detection) and determining the stopping voltage required (typically 120V;) to trap ions within the pre-cell hexapole. The first experiment is performed at a drift voltage of 60V. Subsequent experiments are performed at decreased drift voltages by decreasing the C1 voltage in defined increments to give drift voltages of, for example, 60V, 50V, 40V, 35V, 30V, 25V, 20V, 15V and 10V. The extractor cone was reduced by the same amount as this is not controlled from the drift cell electronics source. Cell

pressure and temperature is maintained constant and measured regularly throughout ion mobility experiments.

Within an ion mobility experiment, a total arrival time distribution (tATD) is produced once every 200 scans by action of pulsed ions being released from the pre-cell hexapole; transiting the drift cell and being mobility separated; being guided through the quadrupole region; being pushed into the ToF analyser; and finally detected. These 200 scans amount to 200 "pushes" of the ToF pusher plate and produce a total ion chromatogram (TIC). At least ten tATDs (2000 scans) were acquired for all samples and were summed to provide greater intensity of low abundance ions with a timescale of around five minutes for each drift voltage. Ion mobility experiments of nine drift voltages require approximately one hour to perform due to software waiting times and the time taken to modify the drift voltage.

Using modified MassLynx v4.1 software, the tATD (Figure 2.5.A) can be interrogated to produce a mass spectrum (Figure 2.5.B) of all ions that arrived within it. A single *m/z* or *m/z* range can then be selected to produce an arrival time distribution of that species as a function of scan number (Figure 2.5.C). Multiplying the scan number of the midpoint of the arrival time distribution by the pusher period will produce the arrival time ($t_a$) for that ion – the time taken for an ion to travel from TH1 to the pusher plate. This time must be corrected as it is the drift time ($t_d$) of the ion within the drift cell that one is interested in, thereby a subtraction of the time spent outside the drift cell ($t_{dead}$) must be performed within the data analysis procedure.

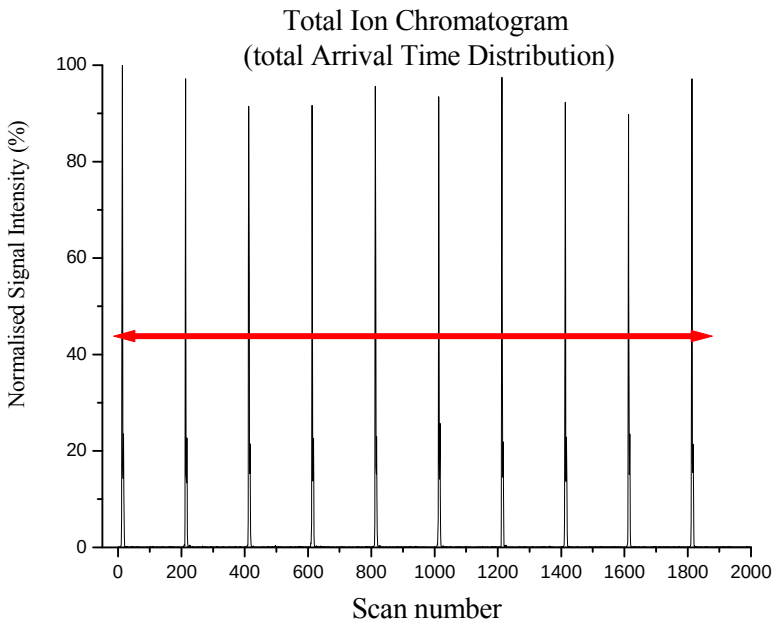



*Figure 2.5.A. Total ion chromatogram (total Arrival Time Distribution) for the peptide aptamer RRMKWWKKSGSGPTTIYY taken with a pusher period of 65 μs (15.4 kHz) at a drift voltage of 60V.*

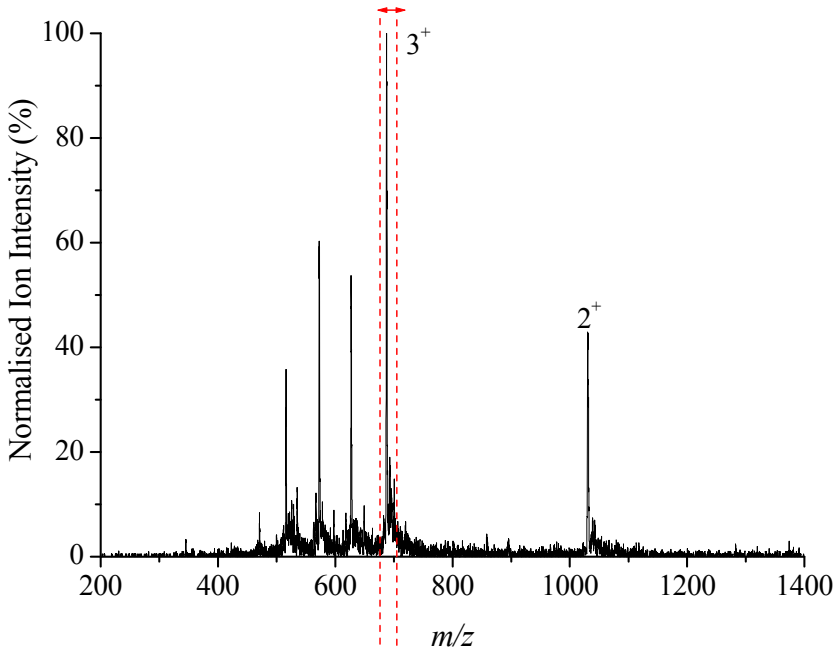



*Figure 2.5.B. Mass spectrum produced upon summing the total arrival time distribution peaks in Figure 2.4.A for the peptide aptamer RRMKWWKKSGSGPTTIYY.*

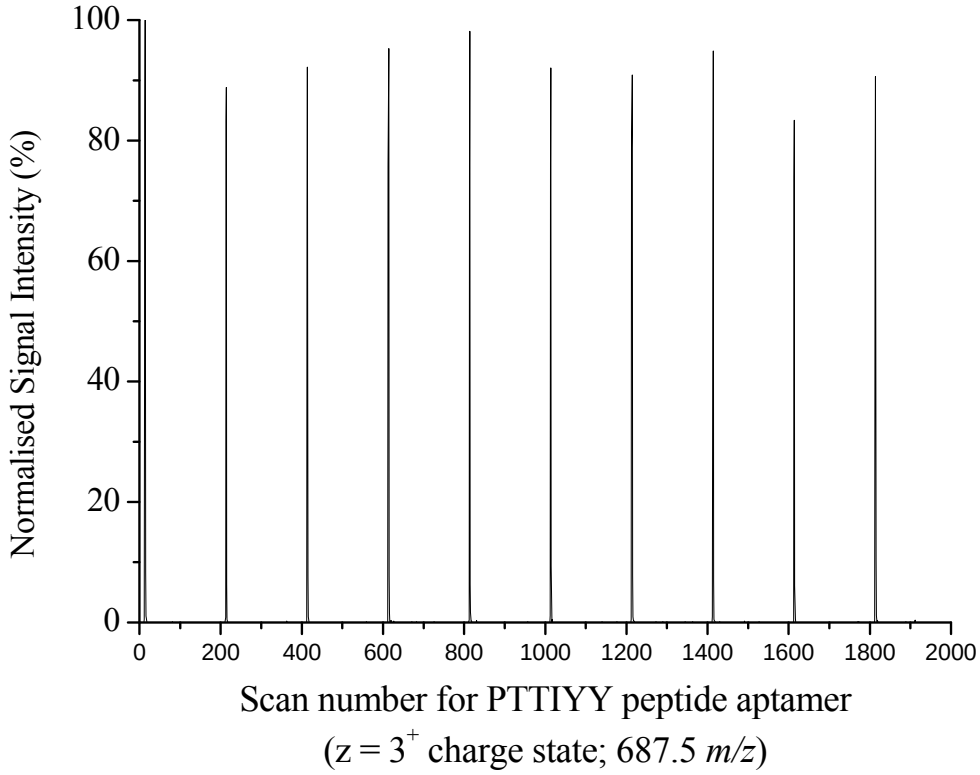



*Figure 2.5.C. Arrival time distribution produced upon summing the desired ion species from the mass spectrum in Figure 2.4.B. This figure shows the z = $3^+$ charge state for the peptide aptamer RRMKWWKKSGSGPTTIYY.*

## 2.9 Data Analysis

Data analysis for ion mobility measurements is a time consuming process due to the number of manipulations required from MassLynx raw data to Microsoft Excel (Microsoft Office 2003) or Microcal Origin 7.5 (Microcal Software Inc., Northampton, Massachusttes, U.S.A.) spreadsheets. This is reduced through use of a spreadsheet that allowed a number of parameters (arrival time midpoint in microseconds at each drift voltage measured; drift voltage; pusher period; temperature; pressure; protein mass; and charge state) to be entered and simple calculations to produce collision cross-section ($Å^2$) and mobility ($cm^2\ V^{-1}\ s^{-1}$) values. Arrival times are calculated by fitting a Gaussian distribution to the arrival time distribution peak and determining the midpoint. This midpoint value is a scan number, and simple multiplication by the pusher period yields the arrival time.

For many species, distinct peaks or peak shoulders were identified as potential second conformers. Two Gaussian peaks are then fitted to these peaks using the peak fitter function in Origin 7.5 (Figure 2.6). The peak shape (shared width but variable height) can be iteratively shaped so as to emulate both peaks as closely as possible, thereby producing two scan numbers and two arrival times. For each

peak fitted, a goodness-of-fit ($R^2$) value could be produced for the separate conformers identified. Values of $R^2 > 0.90$ were deemed acceptable for these curved fits. Values less than this were determined insignificant and a separate conformer was therefore not identified.

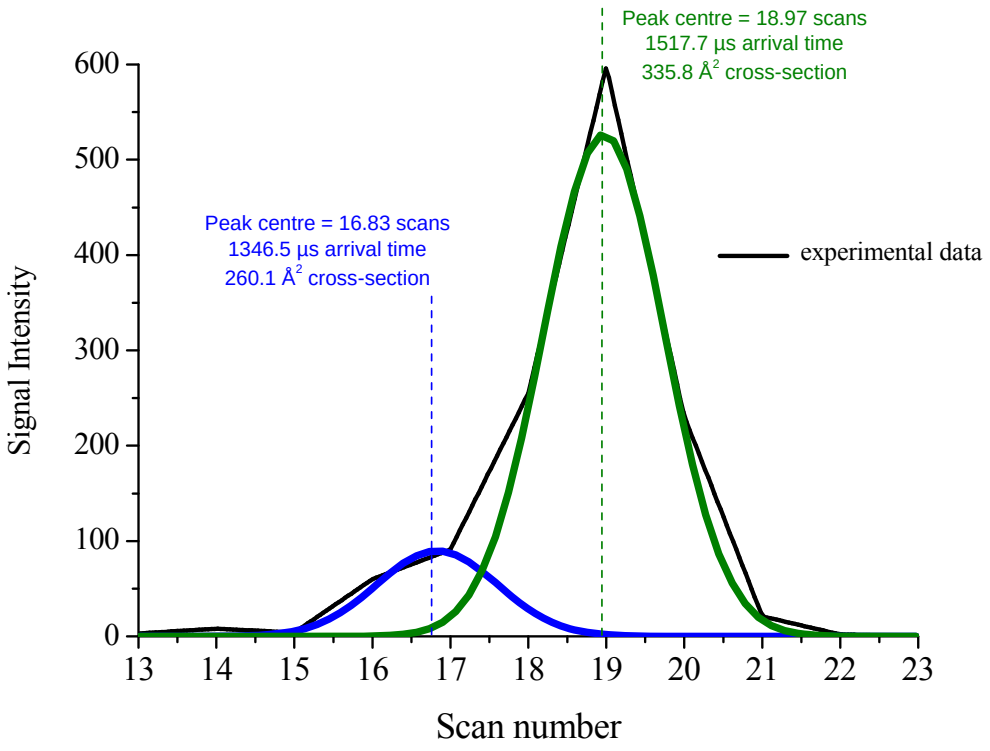



*Figure 2.6. Arrival time distribution for peptide aptamer RRMKWWKKSGSGPTTI $2^+$ charge state at 40V drift voltage. The two peaks represent two distinct gas-phase conformers with arrival times of 1.35 and 1.52 milliseconds. Deduction of the dead time produces a drift time of 604 and 775 microseconds which equates to collision cross-sections of 260.1 Å$^2$ and 335.8 Å$^2$ respectively.*

Upon extracting arrival times for a given *m/z* species at all drift voltages used, arrival times are plotted as a function of cell pressure divided by drift voltage. This arrival time includes both drift and dead times, therefore a straight line plot of the form in equation 2.1 will yield both drift and dead times, as well as the reduced mobility ($K_0$). In equation 2.1: $t_{stated}$ = time (seconds); $L$ (drift cell length) = 0.051m; $T_0$ = 273.15 K; $K_0$ = reduced mobility ($m^2\ V^{-1}\ s^{-1}$); $P_0$ = 760 Torr; $P$ = cell pressure (Torr); and $V$ = drift voltage (V).

$$t_{\text{arrival}} = t_{\text{drift}} + t_{\text{dead}} \qquad \text{where } t_{drift} = \frac{L^2 T_0}{K_0 P_0 T}\frac{P}{V} \qquad \text{Equation 2.1}$$

A plot of arrival time against $P/V$ (Figure 2.7) has a gradient inversely proportional to the reduced mobility of the species. The plot also provides the dead time as the intercept with the y-axis.

Typically, dead times are < 1 milliseconds. Linear regression analysis provided an $R^2$ value (coefficient of determination; 0 to 1.0) with an excellent linear fit tending towards a value of 1.0. In all instances, $R^2$ values of 0.9990 or higher were used to provide collision cross-section data. If the $R^2$ value for a species was lower than this, data were not used. In rare cases, removal of a single poor data point increased the $R^2$ value to become acceptable. All data were therefore a product of at least six drift voltages, where the intensity was high enough, and therefore six data points.

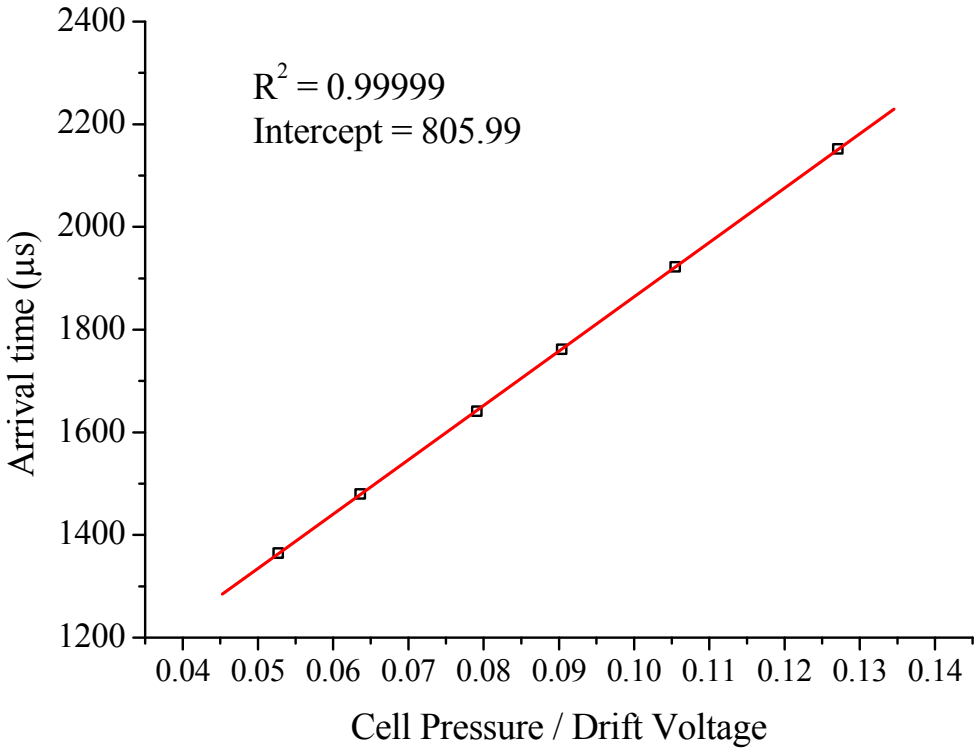



*Figure 2.7. Arrival time plot of aptamer peptide PTTIYY showing six data points collected over a drift voltage range of 60V to 25V. An $R^2$ value of 0.99999 indicates excellent agreement along the linear fit, and the dead time ($t_{dead}$) for this experiment is ~0.806 ms.*

Within the spreadsheet, the reduced mobility value obtained could be further manipulated to elucidate the collision cross-section through equation 2.2[8]: $\Omega$ = collisional cross-section (reported in Å$^2$); $ze$ = ionic charge (C); $N$ = neutral number density; $\mu$ = reduced mass of the ion-neutral pair (kg); $k_B$ = Boltzmann constant; $T$ = drift gas temperature (K); and $K_0$ = reduced mobility (cm$^2$ V$^{-1}$ s$^{-1}$).

$$\Omega = \frac{3ze}{16N}\left[\frac{2\pi}{\mu k_B T}\right]^{0.5}\frac{1}{K_0} \qquad \text{Equation 2.2}$$

## 2.10 Synapt HDMS Instrument

Data for two biological systems presented, Factor H and p53, were also collected on a Synapt HDMS instrument at Waters Mass Spectrometry Technologies Centre (Atlas Park, Manchester, UK) as described in Chapter 1.[9, 10]

### 2.10.1 Synapt Calibration

Collision cross-sections from Synapt data were calculated by using equine heart cytochrome *c* and myoglobin as calibrants, and following the Nature protocol produced by Ruotolo *et al*.[11]

### 2.10.2 Factor H & p53 Experimental Conditions

Data were acquired in positive ionisation mode using a Triversa Nanomate with a capillary voltage of 2.0 to 2.5 kV over a mass range of 300 to 4500 *m/z*. Acquisition parameters were obtained to provide optimal ion mobility separation: cone voltage 40V, trap region collision energy 5eV, and source temperature 120 $^{o}$C. Within the ion mobility separator region, nitrogen gas at 0.51 mbar was used. Velocity and wave height of the travelling wave was varied to obtain optimum signal (200 to 400 $ms^{-1}$). Discrete wave heights of 6, 8 and 10V as well as an 8 to 10V wave height gradient were used. Equine heart cytochrome *c* and equine myoglobin were used to produce a calibration curve for collision cross-section measurements (0.2 mg $mL^{-1}$ dissolved in 50/50 by volume water/acetonitrile containing 0.1% formic acid).

## 2.11 Buffers & Solvents

Water for all experiments was obtained from an Arium 111 Ultrapure system, fitted with a 0.2 µm filter, supplied from a Sartorius Arium 61316 Reverse Osmosis system. Ammonium acetate (77.08 g $mol^{-1}$; Sigma Aldrich, St. Louis, Missouri, U.S.A.) stock solutions were prepared in aqueous solution at pH 6.9. Addition of formic acid (46.03 g $mol^{-1}$; GPR Rectapur, VWR International, Fountenay sous Bois, France) was used to decrease pH, or ammonium hydroxide (35.05 g $mol^{-1}$; Sigma Aldrich) to increase pH of ammonium acetate solutions as required. Methanol (32.05 g $mol^{-1}$), acetonitrile (41.05 g $mol^{-1}$) and isopropanol (60.10 g $mol^{-1}$) were all obtained from VWR Scientific (VWR International, Fountenay sous Bois, France).

## 2.12 Protein Samples

### 2.12.1 Cytochrome *c* Protein

Cytochrome *c* from equine heart was purchased from Sigma Aldrich (product number C7752; ≥95% salt-free; $M_R$ 12,359.9 Da;) as lyophilised solid and stored at -20$^{o}$C until required. Stock solutions of 1 mg $mL^{-1}$ (~80 µM) were made up in purified water or 50 mM ammonium acetate solution (pH 6.85).

Protein was used without further purification. Cytochrome *c* solutions for multimeric investigations were prepared at 2.5 mg $mL^{-1}$ (~200 μM) and left to stand for a minimum of 24 hours at room temperature to induce aggregation. For some experiments these "high concentration" samples were left to aggregate for a period of up to three months. Native-like cytochrome *c* experiments were sprayed from aqueous conditions at 50 to 100 μM concentration. "Denatured" samples were sprayed at a concentration of 20 μM in 50/50 by volume methanol/water with 2% formic acid added.

### 2.12.2 Myoglobin Protein

Myoglobin from equine heart was purchased from Sigma Aldrich (product number M1882; ≥90% essentially salt-free; $M_R$ 17,567.3 Da;) as lyophilised solid and stored at -20°C until required. Protein was made up in deionised water and passed down a PD-10 desalting column (GE Healthcare Bio-sciences AB, Uppsala, Sweden) and eluted with 50 mM ammonium acetate (pH 6.85) allowing aliquots of stock solution, 1 mg $mL^{-1}$ (~60 μM) to be produced. Myoglobin was used predominantly as a standard for checking signal on the instrument after maintenance or as a calibrant in the Synapt HDMS work.

### 2.12.3 Haemoglobin Protein

Haemoglobin was purchased from Sigma Aldrich (product number H7379; human; $M_R$ 64,454.2 Da) as lyophilised solid and stored at -20°C until required. Desalting was achieved using PD-10 columns (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) by dissolving 10 mg haemoglobin powder in 1 mL 50 mM ammonium acetate (pH 6.85), passing it through the column, and eluting with the same buffer. Stock solutions were stored at 4°C for no more than five days. Protein concentration was measured by a UV/vis spectrometer (Cecil 1000 Series, Progen Scientific Ltd., Nuneaton, UK) at $\lambda$ = 274 nm with $\varepsilon_{oxy}$ = 138,048 $M^{-1}$ $cm^{-1}$ and established to be 120 μM. Aliquots of stock solution were diluted as appropriate and solution pH was adjusted to the desired value by addition of formic acid (pH lowered) or ammonium hydroxide solution (pH raised) for experiments that required non-native solution conditions. 2% by volume formic acid was used to acidify sample solutions and to induce protein unfolding. 2% by volume formic acid and 50% by volume methanol were added to buffered samples to induce further denaturation.

### 2.12.4 Factor H Protein Modules fH 10-15 and fH 19-20

Factor H fragment 10-15 (fH 10-15; $M_R$ 40,909.4 Da) and 19-20 (fH 19-20; $M_R$ 14,743.8 Da) were provided by the Barlow Group (University of Edinburgh) as ~1 mL solutions in 20 mM ammonium acetate at a concentration of ~200 μM. Samples were diluted in ammonium acetate to concentrations of ~25 μM and 10% isopropanol by volume was added prior to mass spectrometry analysis as this

assisted in desolvation. Due to a recombinant impurity, N-acetylglucosamine sugar molecules were covalently attached to the protein.

### 2.12.5 p53 Protein

p53 ($M_R$ 26,615.5 Da) samples were provided by the Hupp Group (CRUK, University of Edinburgh) as flash-frozen 25 μL aliquots. Samples were thawed and dialysed for one hour against 20 mM ammonium acetate (pH 6.85) at room temperature. 10% by volume isopropanol was added prior to mass spectrometry analysis giving a protein concentration of ~20 μM. For zinc chelation studies, 1 mM aqueous phenanthroline (180.21 g $mol^{-1}$) was added one hour prior to analysis.

### 2.12.6 AGR2 Protein

AGR2 (his-tagged mature AGR2; 183 residues; ~21,041 Da) was produced, prepared and supplied for MS and IM-MS characterisation by the Hupp Group. 1 mL sample volumes of > 5 mg $mL^{-1}$ concentration were provided in 20 mM Tris, 150 mM NaCl, 10 mM $MgCl_2$ and 20 mM imidazole. Dialyses of 50 μL samples into 1.5 L of 20 mM ammonium acetate (pH 6.84) overnight at 4°C, followed by desalting on a PD-10 column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) with a 20 mM ammonium acetate (pH 6.85) elution buffer was performed to achieve a solution that could be analysed successfully by mass spectrometry. This stock solution of ~1 mg $mL^{-1}$ in 20 mM ammonium acetate (~50 μM concentration) was used for all subsequent analyses. Samples were stored at 4°C and used within three days.

### 2.12.7 Peptide Aptamers

Aptamer peptides were supplied by the Hupp Group as lyophilised powder (~1 mg) and dissolved in deionised water to provide a concentration of 1 mg $mL^{-1}$. For mass spectrometry and IM-MS studies, aptamers were sprayed from a solution of 50/50 by volume methanol/water containing 2% formic acid at a concentration of ~100 μM. This provided the strongest signal intensity required for analysis. Peptides were diluted as required for complexing work with AGR2. Reading from the N-terminus, each peptide has a seven residue penetratin sequence (RRMKWKK) followed by a four residue spacer (SGSG). The remaining residues differ between the six aptamers; five of them being hexpeptides, and the sixth a tetrapeptide.

## 2.13 Chapter 2. References


1. McCullough, B. J.; Kalapothakis, J.; Eastwood, H.; Kemper, P.; MacMillan, D.; Taylor, K.; Dorin, J.; Barran, P. E., Development of an Ion Mobility Quadrupole Time of Flight Mass Spectrometer. *Anal. Chem.* **2008,** 80, (16), 6336-6344.

2. McCullough, B. J.; Eastwood, H.; Clark, D. J.; Polfer, N. C.; Campopiano, D. J.; Dorin, J. A.; Maxwell, A.; Langley, R. J.; Govan, J. R. W.; Bernstein, S. L.; Bowers, M. T.; Barran, P. E., Characterisation of DEFB107 by mass spectrometry: Lessons from an anti-antimicrobial defensin. *Int. J. Mass Spectrom.* **2006,** 252, (2), 180-188.

3. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** 246, (4926), 64-71.

4. Wilm, M. S.; Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Processes* **1994,** 136, (2-3), 167-180.

5. Wilm, M.; Mann, M., Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996,** 68, (1), 1-8.

6. Heck, A. J. R.; van den Heuvel, R. H. H., Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **2004,** 23, (5), 368-389.

7. Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V., Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chem. Rev.* **2007,** 107, (8), 3544-3567.

8. Mason, E. A.; McDaniel, E. W., *Transport Properties of Ions in Gases*. Wiley: New York, 1988.

9. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H., An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **2007,** 261, (1), 1-12.

10. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H., Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **2008,** 22, (20), 3297-304.

11. Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V., Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **2008,** 3, (7), 1139-52.

# Chapter 3. Mass Spectrometry and Ion Mobility-Mass Spectrometry Studies of Cytochrome *c*

*Chapter 3 introduces new data on a biological system well studied by mass spectrometry and ion mobility-mass spectrometry. Cytochrome c has been investigated under native-like and denaturing conditions, revealing a plethora of gas-phase multimer species. Monomer, dimer and trimer collision cross-sections are presented. Gas-phase dissociation of higher order species is discussed. Fragment ions associated with a novel fragmentation technique, native electron capture dissociation, are also presented.*

## 3.1 Cytochrome *c* Introduction

Cytochrome *c* is intensively studied in many areas of science because of its central role in electron transfer in living organisms.[1] Cell death may be triggered by a cytochrome *c* pathway, providing interest into the protein as an anticancer agent.[2, 3] It has also been a model system in protein sequence and structure evolution studies. The primary sequence is highly conserved, with residue identity ranging from 45% to 100% across all eukaryotes[4] and the most highly conserved region encompassing residues 67 to 87 (which contain the distal Met80 ligand). Cytochrome *c* contains a covalently-bound haem (haem c; protoporphyrin IX) prosthetic ligand to the protein backbone *via* two thioether linkages with Cys14 and Cys17. The CXX´CH motif is extremely highly conserved with His18 (within the motif) proximally coordinating the iron atom within the haem ligand. The iron ion is hexacoordinated by four pyrrole nitrogens in the haem ligand, the Nε1 atom of the His18 side chain, and the sulfur atom of the Met80 side chain.[4] Additionally, propionate 6 is hydrogen bonded to the backbone amide protons of Lys79 and the hydroxyl proton of Thr78; propionate 7 is hydrogen bonded to the side chains of Trp59 and Asn52. The secondary structure of equine heart cytochrome *c* has an N-terminal helix (residues 1 to 15; N-terminal is acetylated), followed by a loop that contains His18. A stretch of ~30 residues having little regular secondary structure (including the Ω-loop; residues 18 to 34), bar small hydrogen bonded glycine-containing loops, lead into the second helix (residues 50 to 56). This is followed by a short loop and the third helix (residues 61 to 69). The fourth helix (residues 71 to 75) precedes the loop containing the distal Met80 ligand. Finally, the C-terminal helix (residues 90 to 104) completes the protein structure.[5] An x-ray crystal structure of equine heart cytochrome *c* is shown in Figure 3.6 later in this Chapter.

Mass spectrometry and ion mobility-mass spectrometry studies of cytochrome *c* (as detailed in the Introduction) have provided structural information on the gas-phase conformations adopted upon ion generation by electrospray ionisation from a number of solvent compositions. Evidence for folding and refolding of gas-phase cytochrome *c* ions was produced by Badman and Clemmer.[6, 7] Structural transitions that occur in the timescale of seconds to minutes can be studied by FT-ICR mass

spectrometry using hydrogen/deuterium exchange whereby exposed amide protons on the protein backbone exchange with solvent or gas-phase deuterons, thereby increasing the observed protein mass. Transitions that occur over the time period ~10 milliseconds to 10,000 milliseconds were probed by storing the ions in an ion trap, and subsequently ejecting them into a ion mobility-mass spectrometer. Results showed that low charge state ($z = 7^+$ and $9^+$) cytochrome *c* monomers adopt initially compact structures upon transfer to the gas phase which unfold to considerably larger and more extended structures after ~30 milliseconds. After a longer time period (~250 milliseconds to 10,000 milliseconds) these extended structures have adopted more compact structures, indicating gas-phase refolding; however these structures are not as compact as those initially observed. Folding and unfolding transitions occur on the microsecond to millisecond timescale in solution, whereas in the gas phase, they occur on the millisecond to second timescale. Laser-induced unfolding studied by electron capture dissociation (ECD)[8] observed refolding of a $9^+$ cytochrome *c* ion over 120 seconds (120,000 milliseconds). Their data show that the ECD spectra obtained 120 seconds after laser excitation are similar to spectra recorded without laser excitation. This result confirms that the ions have refolded within 120 seconds of being denatured. Structural studies of cytochrome *c* have also revealed that it is composed of five foldon units that continually fold and unfold, even under native conditions.[9] Equilibrium and kinetic hydrogen exchange experiments allowed unfolded higher energy structures to be detected as protons that are protected in the native state can exchange with deuterons from $D_2O$ solvent only when their protecting H-bonds are transiently broken as the protein "searches" thermodynamically through its higher-energy forms.

## 3.2 Cytochrome *c* Experimental

As detailed in Chapter 2, cytochrome *c* stock solutions were produced in ammonium acetate (pH 6.85) or water and diluted as required. To collect mass spectra in traditional mass spectrometry mode, the MoQTOF drift cell was filled with ~3.0 Torr helium at a temperature of ~300 K. The drift voltage was fixed at 60V. Mass spectrometry and ion-mobility mass spectrometry source conditions were optimised for the species of interest: nano-ESI capillary voltage range was ~1.5 to 2.5 kV; sample cone range of 150 to 160V; extraction cone 110 to 120; source temperature 80°C; source pressure ~4.0 $\times 10^{-2}$ mbar for non-multimeric mass spectra and $4.0 \times 10^{-1}$ mbar for mass spectra with multimers present. Ion mobility settings were as above but with incremental decreases on the C1 voltage of 60V, 50V, 40V, 35V, 30V, 25V, 20V, 15V and 10V with the same change to the extractor cone voltage; pulse width 40 μs; stopping voltage (TH1) 125 to 140V; pusher period: 80 μs for acquisitions up to 2200 *m/z*; 112 μs for multimer experiments.

## 3.3 Mass Spectrometry Studies of Cytochrome *c*

Cytochrome *c* was investigated under a range of solvent conditions to compare with data obtained by ion mobility-mass spectrometry data. Charge state distribution can be used to infer structural information from MS data, as well as providing information regarding the protein in solution.

### 3.3.1 Mass Spectrometry of Cytochrome *c* Sprayed from Native-like & Aqueous Conditions

20 µM cytochrome *c* from 50 mM ammonium acetate (native-like) solution displayed a narrow charge state distribution ($z = 6^+$ to $8^+$) indicating a well folded protein (Figure 3.1). Protein from ~200 µM aqueous stock solution was diluted three fold and sprayed. Figure 3.2 is a mass spectrum of cytochrome *c* under aqueous conditions at ~67 µM with an increased source pressure ($4.0 \times 10^{-1}$ mbar).

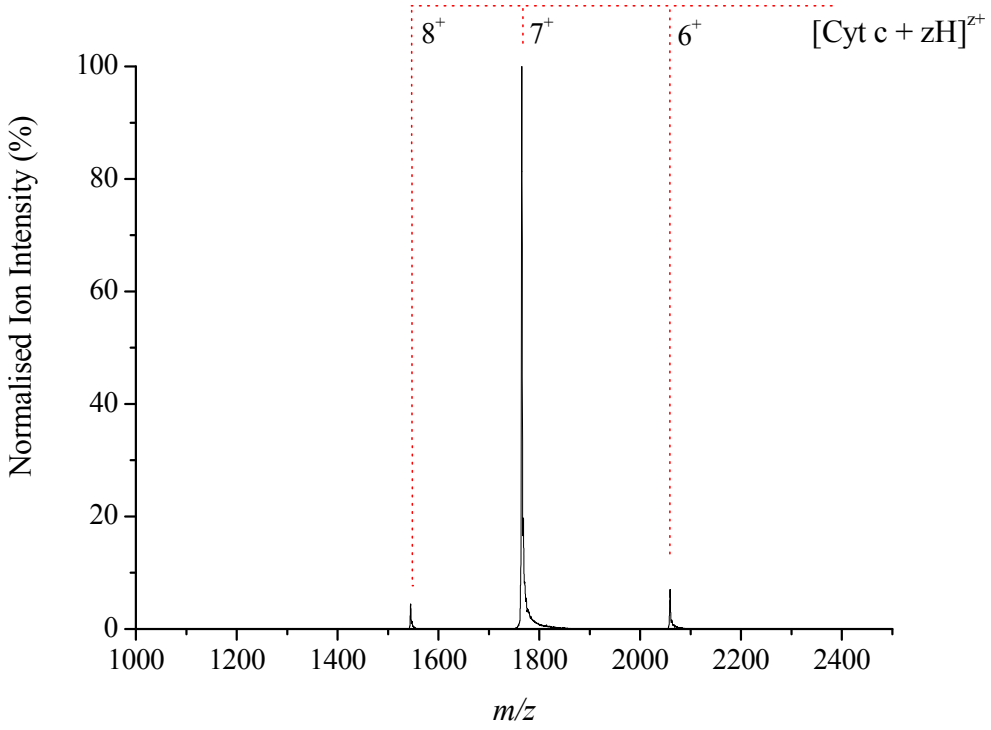



*Figure 3.1 Mass spectrum of 20 µM cytochrome c sprayed from 50 mM ammonium acetate (pH 6.85).*

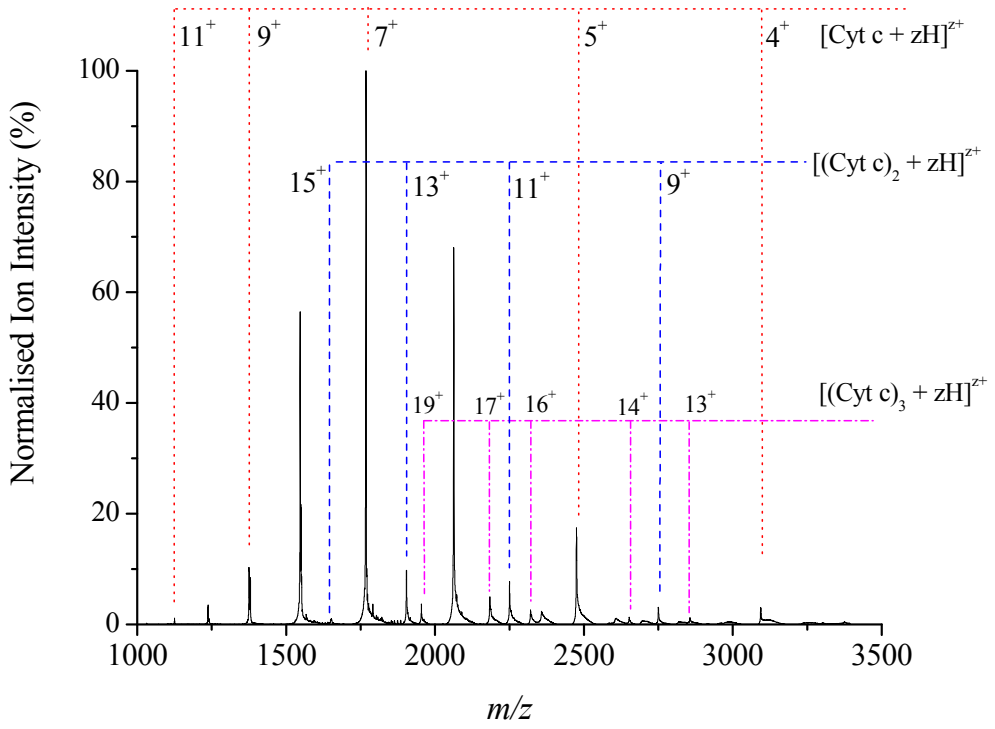



*Figure 3.2. Cytochrome c sprayed from aqueous solvent at a concentration of ~67 μM, pH 5.65. A wide range of gas-phase multimer species are observed. Signal is also present for a single tetramer (z = 17$^{+}$; 2909.2 m/z) and pentamer (z = 19$^{+}$; 3253.6 m/z), but at very low abundance (not marked on spectrum).*

Increased source pressure is required to observe the higher *m/z* species; at pressures less than ~9.0 × $10^{-2}$ mbar, only monomer species are observed within a relatively narrow charge state. This transfer of high *m/z* species is a result of solution-phase monomer association due to the increased concentration. Upon transfer to the gas phase, bulk solvent loss causes electrostatic interactions to dominate over hydrophobic interactions, and a population of dimer subunits are stabilised. From Figure 3.2, dimeric and trimer species have been assigned from calculated *m/z* values. Evenly-charged dimers have concomitant *m/z* values with monomer species of half the charge *i.e.* the peak at 1766.7 *m/z* is predominantly composed of the z = 7$^{+}$ monomer; however some of the peak intensity will be made up by the z = 14$^{+}$ dimer (cytochrome *c* dimer $M_R$ = 24,718.4 Da). The charge state distribution is wider than would be expected for a native-like protein (*c.f.* Figure 3.1) and higher charge states may be the result of charge partitioning and dissociation from a higher order multimer such as a dimer.[10] This will be discussed later in the Chapter.

### 3.3.2 Mass Spectrometry of Cytochrome *c* at pH 3.0

Cytochrome *c* from ammonium acetate was made up in 50/50 by volume methanol/water solution containing 2% formic acid with a pH of 3.0, at a concentration of 20 μM. The mass spectrum obtained under these conditions is shown in Figure 3.3.

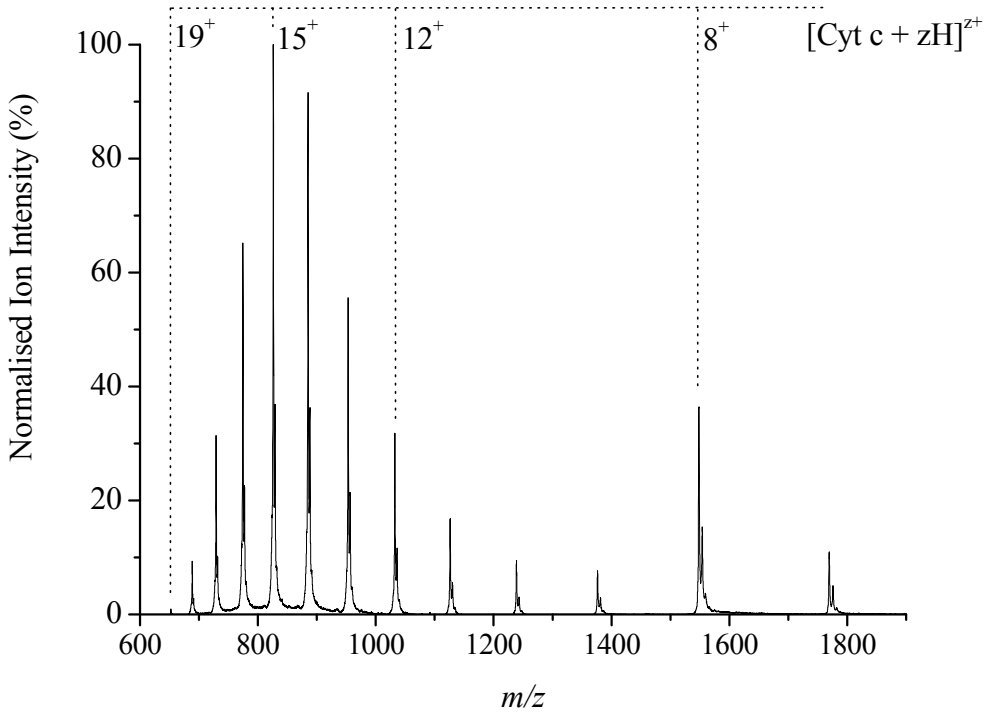



*Figure 3.3 Cytochrome c (20 μM) sprayed at pH 3.0 containing 50/50 by volume methanol/water and 2% formic acid.*

The mass spectrum of cytochrome *c* (pH 3.0) shows a wide charge state distribution over the mass range 600 to 1900 *m/z*. Thirteen charge states, $z = 7^+$ to $19^+$, are observed indicating solvent accessible basic residues on the protein backbone have been protonated, causing unfolding to overcome Coulombic repulsion. Equine heart cytochrome *c* contains 24 basic (2 R, 19 K and 3 H) and 12 acidic (3 D and 9 E) amino acid residues, allowing denaturing mass spectrometry to be used as a tool to infer structural information. The mass spectrum displays a bimodal charge distribution with centres at charge states $8^+$ and $15^+$. This is an indicator that two conformations, a folded and unfolded state, are present in solution under these conditions and transferred to the gas-phase.[11] Further addition of acid (or denaturation with a strong denaturant like urea) would cause a shift in the most intense peak to higher charge (lower *m/z*), and a decrease in intensity of lower charged (higher *m/z*) as the folded conformation is lost and the unfolded state is predominantly populated.[12] These observations are explained in that tightly folded molecules are expected to have significantly smaller projected areas compared to less-structured protein molecules as their side chains are involved in intramolecular

interactions. Therefore, fewer charges can be accommodated on the protein surface during the ion evaporation stage. This results in two protein conformations whose geometries are different from one another and also have different charge densities.

## 3.4 Ion Mobility-Mass Spectrometry of Cytochrome *c*

IM-MS data has been collected for cytochrome *c* under the same solvent conditions as shown for mass spectrometry experiments. Collision cross-sections are plotted as a function of charge state to observe trends upon protonation.

### 3.4.1 Ion Mobility-Mass Spectrometry of Cytochrome *c* from Native-like & Aqueous Conditions

Cross-sections for three charge states ($z = 6^+$ to $8^+$) obtained from cytochrome *c* are presented in Table 3.1 alongside cross-sections taken from the Clemmer database[13] for the same charge states. Multiple cross-sections from the Clemmer database arise from conformations obtained under differing injection energy parameters. MoQTOF values were obtained at a single injection energy (35V). Similar charge states between MoQTOF and the database values are obtained for the $z = 6^+$ and $8^+$ species.

*Table 3.1. Collision cross-sections elucidated under native-like conditions for cytochrome c tabulated against cross-sections for protonated cytochrome c obtained from the Clemmer database. MoQTOF values are an average of three replicates & stated with twice the standard deviation of the mean ($2\sigma$).*

| Charge state ($z^+$) | Collision cross-section obtained on the MoQTOF (Å$^2$) | Collision cross-sections obtained from Clemmer database (Å$^2$) |
|---|---|---|
| 6 | 1240 ± 11 | 1244<br>1393<br>1602 |
| 7 | 1530 ± 15 | 1247<br>1620<br>1785<br>2007 |
| 8 | 1840 ± 14 | 1250<br>1702<br>1845<br>2061 |

Cross-sections for eight monomer charge states ($z = 5^+$ to $8^+$), five dimer charge states ($z = 11^+$ to $19^+$; odd numbers only) and three trimer species ($z = 16^+$, $17^+$ and $19^+$) were elucidated for cytochrome *c* sprayed from aqueous conditions. These are presented in Table 3.2.

*Table 3.2. Collision cross-sections of aqueous cytochrome c (67 μM concentration) monomer, dimer and trimer species over a fifteen charge state range. Values are an average of three replicates & stated with twice the standard deviation of the mean (2σ). In table, n.o. represents not observed.*

| Charge state | Monomer cross-section ($Å^2$) | Dimer cross-section ($Å^2$) | Trimer cross-section ($Å^2$) |
|---|---|---|---|
| 5 | 1217 ± 18 | n.o. | n.o. |
| 6 | 1243 ± 12 | n.o. | n.o. |
| 7 | 1546 ± 14 | n.o. | n.o. |
| 8 | 1865 ± 18 | n.o. | n.o. |
| 9 | 2066 ± 18 | n.o. | n.o. |
| 10 | 2228 ± 21 | n.o. | n.o. |
| 11 | 2255 ± 15 | 2586 ± 24 | n.o. |
| 12 | 2407 ± 27 | n.o. | n.o. |
| 13 | n.o. | 3248 ± 26 | n.o. |
| 14 | n.o. | n.o. | n.o. |
| 15 | n.o. | 3389 ± 18 | n.o. |
| 16 | n.o. | n.o. | 3812 ± 31 |
| 17 | n.o. | 3953 ± 20 | 4014 ± 34 |
| 18 | n.o. | n.o. | n.o. |
| 19 | n.o. | 4157 ± 24 | 4430 ± 30 |

Cross-sections for monomer, dimer and trimer species from aqueous cytochrome *c* are plotted as a function of charge state in Figure 3.4. Monomer cross-sections range from 1217$Å^2$ for z = $5^+$ to 2407$Å^2$ for z = $12^+$ consistent with previous work,[14] in agreement with research by Clemmer[15] and consistent with coulombically-driven unfolding. Five dimer cross-sections ranging from 2586$Å^2$ to 4157$Å^2$ (z = $11^+$ to $19^+$ respectively) were elucidated which were reported for the first time[16] from this research. Three trimer species with cross-sections of 3812$Å^2$, 4014$Å^2$ and 4430$Å^2$ for z = $16^+$, $17^+$ and $19^+$ respectively were obtained also.

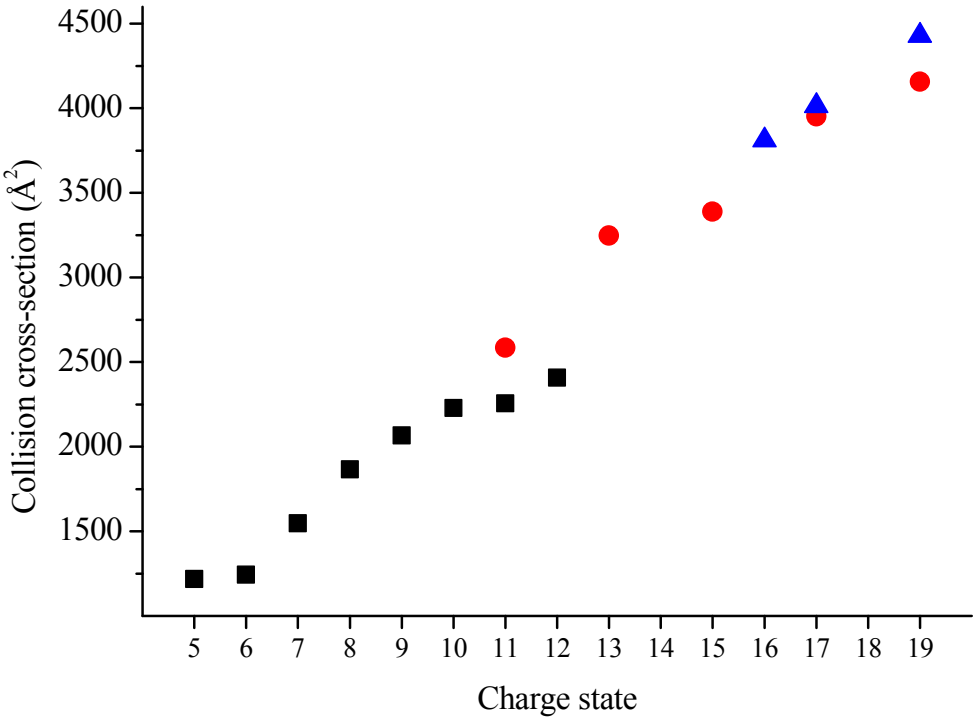



*Figure 3.4. Cytochrome c collision cross-sections for monomer (black squares), dimer (red circles) and trimer (blue triangles) species over a range of charge states.*

It is observed that increasing charge causes unfolding of protein species to overcome Coulombic repulsion. This trend is obvious with the monomer species. Cross-sections for dimer are, as expected, larger than monomer values, but their cross-sections are not a sum of the two monomer constituents, thus indicating an amount of gas-phase structural rearrangement to stabilise the dimeric species. As electrostatic interactions are of greater influence in the gas phase, it is the interaction between charged amino acid residues on regions of the molecule that are the most likely the stabilising structures in the dimer. Dimer and monomer cross-sections with charge state $z = 11^+$ differ by 12.8%. This comparison illustrates that upon dimer formation, protonation events can occur at a greater number of exposed basic residues (but presumably excluding those at the dimer interface) when compared with the monomer alone. This also contributes to the lower cross-section as more remote locations on the exposed surface will be protonated and the onset of unfolding events will occur at higher charge states *i.e.* more charge is acquired before unfolding occurs. This has implications in dimer gas-phase dissociation which is discussed later in this Chapter. Protons within the trimer species will be distributed throughout the three monomer constituent subunits. A smaller number of trimer charge states were observed, and a much smaller change in cross-section with charge. Unfolding to larger cross-sections will possibly not occur until a large amount ($z > 21^+$ perhaps) of charge is sequestered on the assembly or, more likely is that dissociation to: i) monomer subunits; or ii) a dimer and a

monomer subunit, will occur before unfolding. This may explain the small observed population of trimer species.

## 3.5 Monomer Cross-section Comparison with Calculated Structure

Using the x-ray crystal structure of cytochrome *c* (PDB ID: 1CRC; Figure 3.6)[17] and the mobility calculator program MOBCAL,[18, 19] a theoretical collision cross-section for the monomer species using the exact hard-spheres scattering (EHSS) method was calculated to be 1339Å$^2$, consistent with previous work.[15] This is shown as a dashed line on Figure 3.5 along with the monomer cross-sections taken from Table 3.2

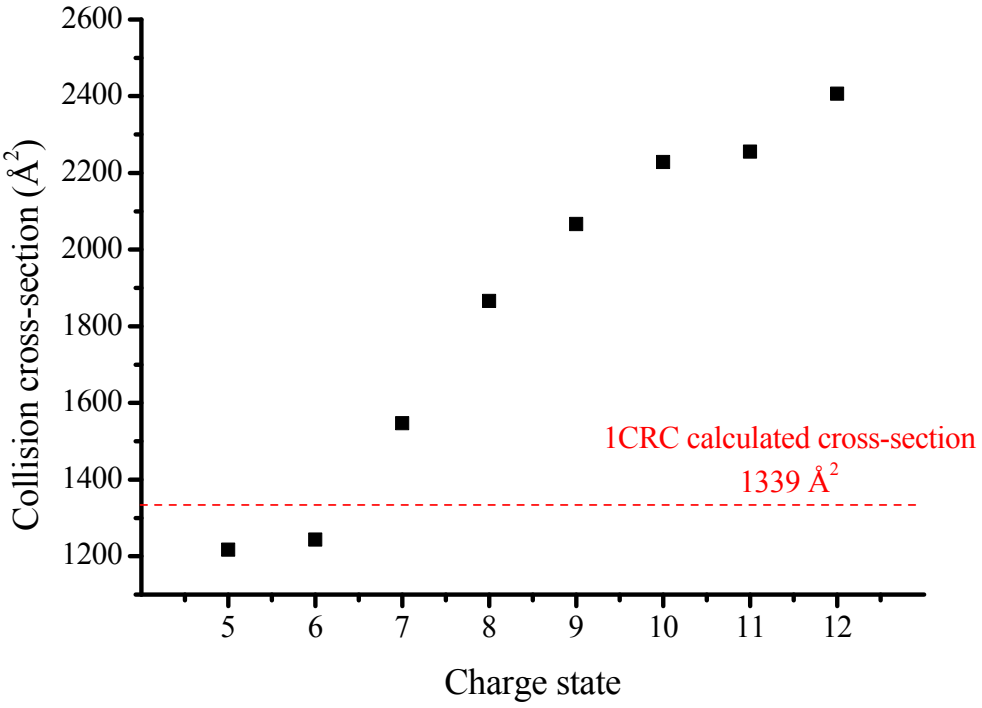



*Figure 3.5 Cytochrome c monomer cross-section plot with the x-ray crystal calculated cross-section plotted as the red dotted line at 1339 Å$^2$.*

From Figure 3.5, it can be seen that the calculated cross-section falls between the $6^+$ and $7^+$ charge states of the IM-MS data, indicating that mass spectrometry can provide gas-phase structures comparable to condensed-phase structures at low charge state. Condensed-phase structures are produced under conditions where salts and other small molecules may be trapped within their structure, giving a larger cross-section than compared with gas-phase structures (where they are lost). Similarly, gas-phase structures are consistently smaller than condensed-phase structures due to gas-

phase collapse; amino acid side chains folding back and interacting with the protein backbone in the absence of bulk solvent.[20]

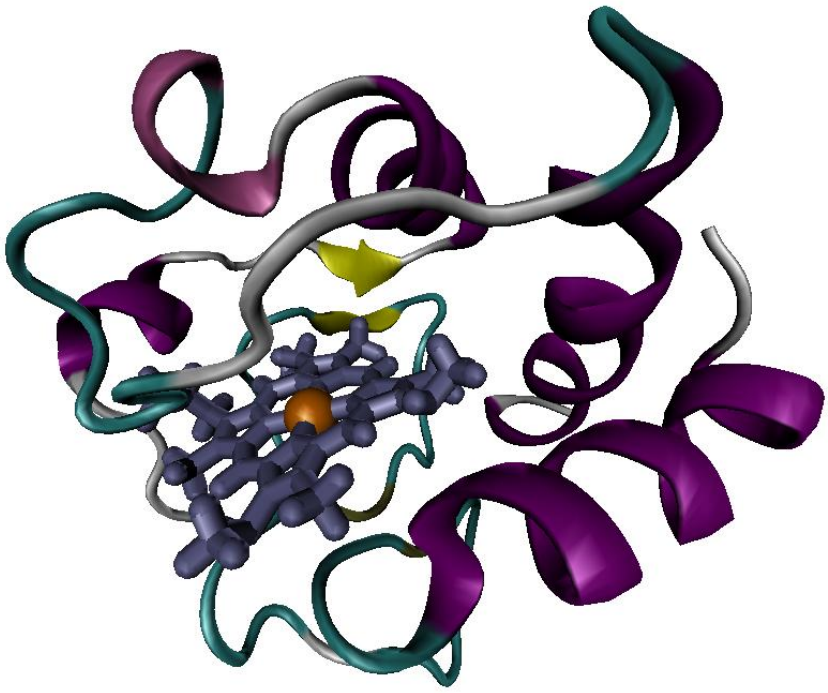


*Figure 3.6. Cytochrome c structure (PDB ID: 1CRC)[17] modified using Visual Molecular Dynamics software.[21] Alpha-helical regions are shown in purple; beta-sheet regions in yellow; turns in green; and the haem ring in blue with an orange iron ion coordinated at the centre.*

## 3.6 Dimer Cross-section Comparison with Calculated Structure

Using the x-ray crystal structure of cytochrome *c* (PDB ID: 1CRC)[17] and the mobility calculator program MOBCAL,[18, 19] two cytochrome *c* monomers were coarsely docked together to give an approximate dimer arrangement. This structure was used to calculate an EHSS approximate dimer collision cross-section of 2374Å$^2$ and is plotted in Figure 3.7 (dashed red line) along with the dimeric cytochrome *c* cross-sections from Table 3.2.

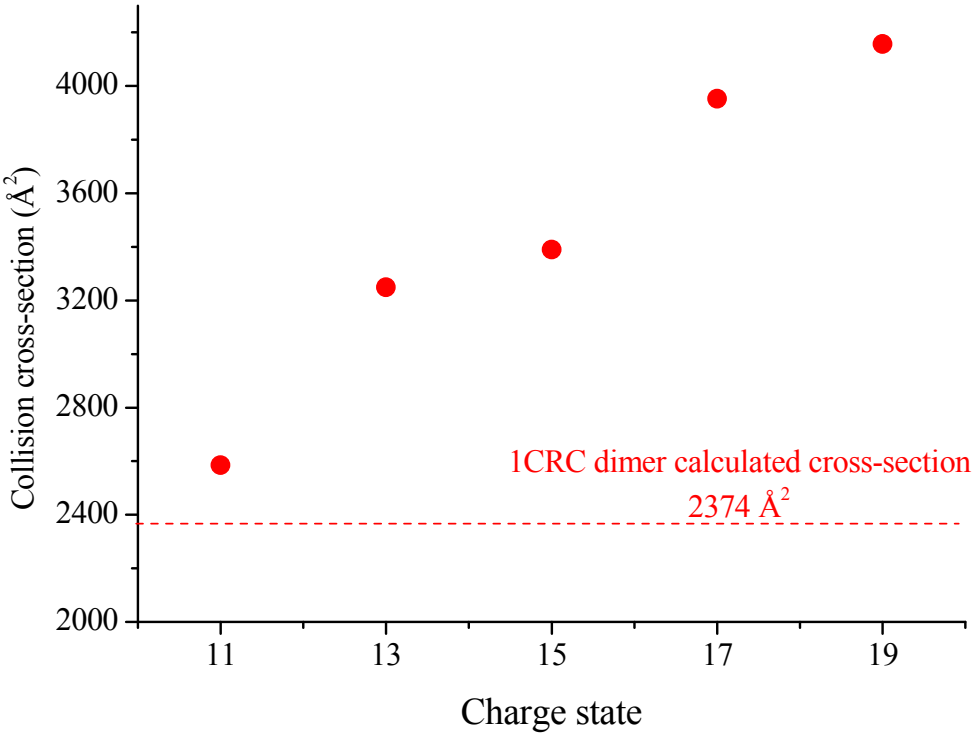



*Figure 3.7 Cytochrome c dimer cross-section plot with the coarse docked dimer x-ray crystal calculated cross-section plotted as the red dotted line at 2374Å².*

The calculated dimer cross-section value is smaller (8.2%) than the lowest experimental cross-section, but provides an approximate value for comparison. The coarse docking procedure merely brought the two monomeric subunits into close contact with one another, but did account for any structural rearrangements that would most certainly occur *in vivo*. Figure 3.8 shows the asymmetric dimer structure found in crystals of equine ($Fe^{III}$) cytochrome *c* at low ionic strength,[17, 22] and highlights the ionisable amino acid residues (acidic and basic) in light gray. It shows that all ionisable groups of either monomer are fully solvated and none of them are in contact with the other monomer. It is therefore possible to assume that the solution charge distribution for both monomers as part of the dimer structure is the same or similar for free monomeric ($Fe^{III}$) cytochrome *c*.

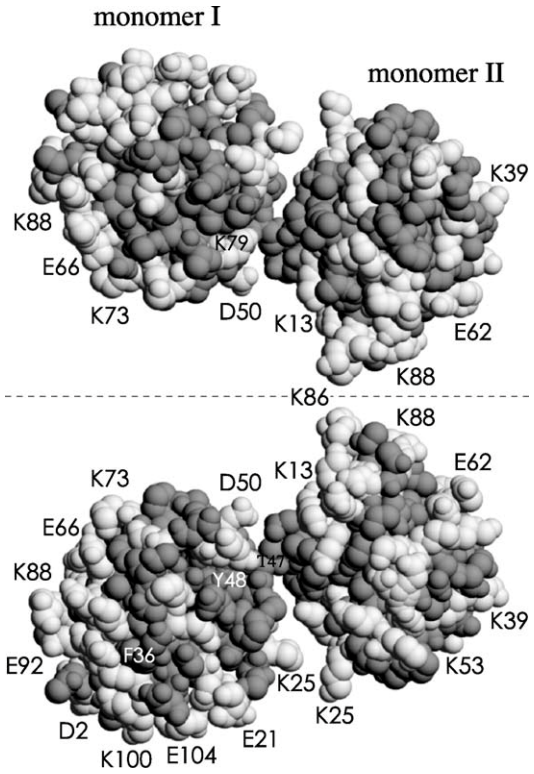



*Figure 3.8.[22] Asymmetric cytochrome c dimer structure showing ionisable residues in light gray. Views differ by a 180° rotation around the dotted axis line.*

## 3.7 Gas-phase Dimer Dissociation

Studying multimeric protein structures in the gas phase allows intrinsic properties of the protein to be studied in the absence of bulk solvent. Of interest are the pathways that are taken by gas-phase structures upon dissociation to smaller molecules, with the goal of obtaining structural information about the large intact assembly and the smaller molecule that has been lost. Interrogation of the intact *E.coli* 50S ribosomal subunit (33 proteins and 2 RNA strands in a non-covalent complex) by gas-phase activation revealed proteins containing a disproportionately high degree of charge for their mass compared to the intact complex.[23] Other groups have reported this dissociation behaviour for multimeric protein complexes with, most commonly, a monomer subunit being ejected with a disproportionate amount of charge for its mass compared to the mass of the remaining complex.[24-26] Jurchen and Williams[10, 27] have suggested charge partitioning behaviour is a function of several parameters: complex charge state; the dissociation energy (using SORI excitation); the solution composition the complexes are formed in; and the protein's conformational flexibility within the complex. They reported that, with respect to charge state, dimer ions with z = $19^+$ and $17^+$ charge dissociate predominantly symmetrically, producing monomer ions carrying similar amounts of charge. As the dimer charge state decreases, this dissociation changes to an asymmetric pattern with one monomer carrying more charge than the other. This result is complemented by the cross-sections observed using IM-MS of aqueous solutions of cytochrome *c* in this research. Table 3.3 correlates ion mobility data with the observations from Jurchen's work.

*Table 3.3. Correlation of dimer charge state with cross-section. The dissociation behaviour description relates the findings to the work of Jurchen and Williams.*[10]

| Dimer charge state ($z^+$) | Cross-section ($Å^2$) | Dissociation behaviour |
|---|---|---|
| 11 | 2586 ± 24 | All asymmetric |
| 13 | 3248 ± 26 | Predominantly asymmetric |
| 15 | 3389 ± 18 | Predominantly asymmetric |
| 17 | 3953 ± 20 | Predominantly pseudo-symmetric |
| 19 | 4157 ± 24 | Pseudo-symmetric |

From Table 3.3 it can be seen that increasing dimer charge state increases cross-section. This is accompanied by a gradual shift in dissociation behaviour from asymmetric behaviour only ($z = 11^+$) to pseudo-symmetric behaviour alone at higher charge state ($z = 19^+$). From the IM-MS data, this can be explained by the unravelling of one monomer within the dimer complex to a greater extent than the other. This unfolding monomer exposes a greater number of basic residues available for protonation, and so creates a charge asymmetry between the molecules. A point is reached whereby Coulombic interactions force dissociation of the dimer into the highly charged unfolded monomer, and a lower charged more compact monomer. On these experimental timescales, it must be noted that the more compact monomer may also unravel after dissociation. At high charge states, the monomer subunits within the dimer unfold and are protonated to similar extents, resulting in pseudo-symmetric dissociation.

## 3.8 Ion Mobility-Mass Spectrometry of Cytochrome *c* at pH 3.0

A large number of monomer cross-sections were elucidated by IM-MS for cytochrome *c* at pH 3.0. Cross-sections are plotted (black squares) as a function of charge state in Figure 3.9 alongside cross-sections obtained from the Clemmer database (red circles) for protonated cytochrome *c* and show excellent agreement for the majority of cross-sections.

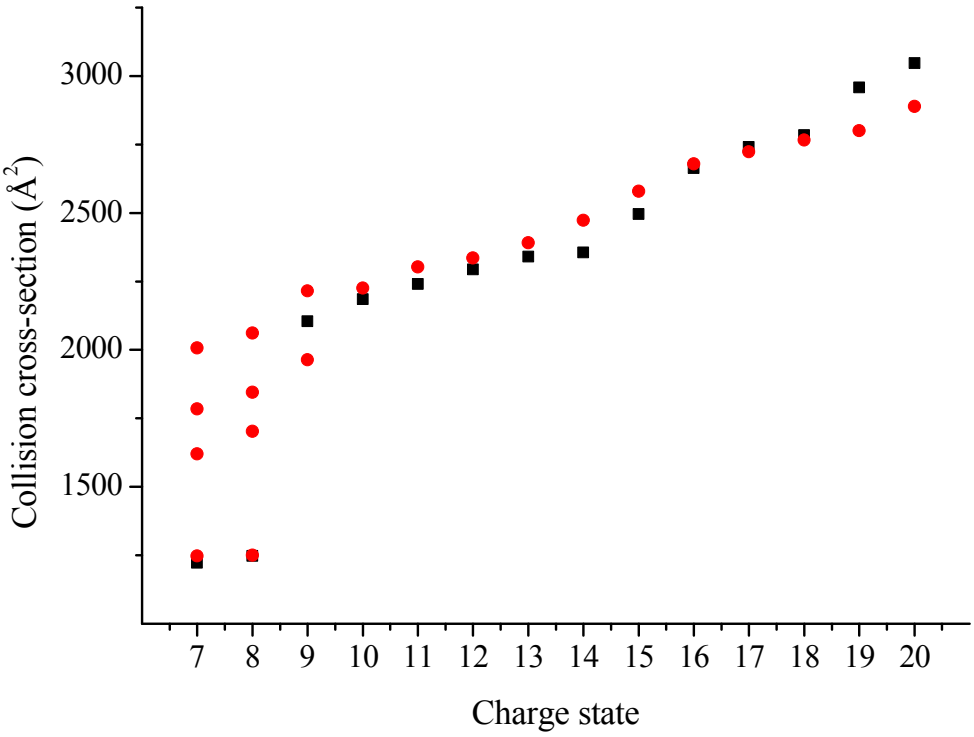



*Figure 3.9. Cytochrome c cross-sections obtained at pH 3.0 on the MoQTOF instrument (black squares) compared with cross-sections from the Clemmer database (red circles; multiple gas-phase conformers at low charge state).*

Cross-sections obtained on the MoQTOF are within ~5% of the database values for all charge states from z = $7^+$ to $18^+$ (*n.b. i) if the two z = $9^+$ database values are averaged, this is within ~1%; and ii) z = $19^+$ and $20^+$ are within ~6%)* which shows excellent agreement. Low charge states are shown to retain compact structures under pH 3.0. Further addition of acid or increasing the organic solvent composition would increase these cross-sections if detected in the gas phase. Cross-section values are presented in Table 3.4.

*Table 3.4. Collision cross-section values for monomeric cytochrome c obtained at pH 3.0. Values are an average of three replicates & stated with twice the standard deviation of the mean (2σ).*

| Charge state | Monomer (pH 3.0) collision cross-section ($Å^2$) |
|---|---|
| 7 | 1223 ± 13 |
| 8 | 1248 ± 16 |
| 9 | 2104 ± 15 |
| 10 | 2185 ± 21 |
| 11 | 2240 ± 27 |
| 12 | 2294 ± 22 |
| 13 | 2340 ± 28 |
| 14 | 2355 ± 28 |
| 15 | 2496 ± 32 |
| 16 | 2663 ± 36 |
| 17 | 2741 ± 24 |
| 18 | 2784 ± 22 |
| 19 | 2958 ± 29 |
| 20 | 3046 ± 28 |

It is clear that cross-section increases with charge to overcome Coulombic repulsion. The small population of compact conformers at low charge states ($z = 7^+$ and $8^+$) are interesting. By combining all data for monomer cross-sections (monomer from native-like conditions; monomer from aqueous solution; and monomer from denaturing conditions), some further information is gleamed. A plot of all monomer cross-sections against charge state is presented in Figure 3.10.

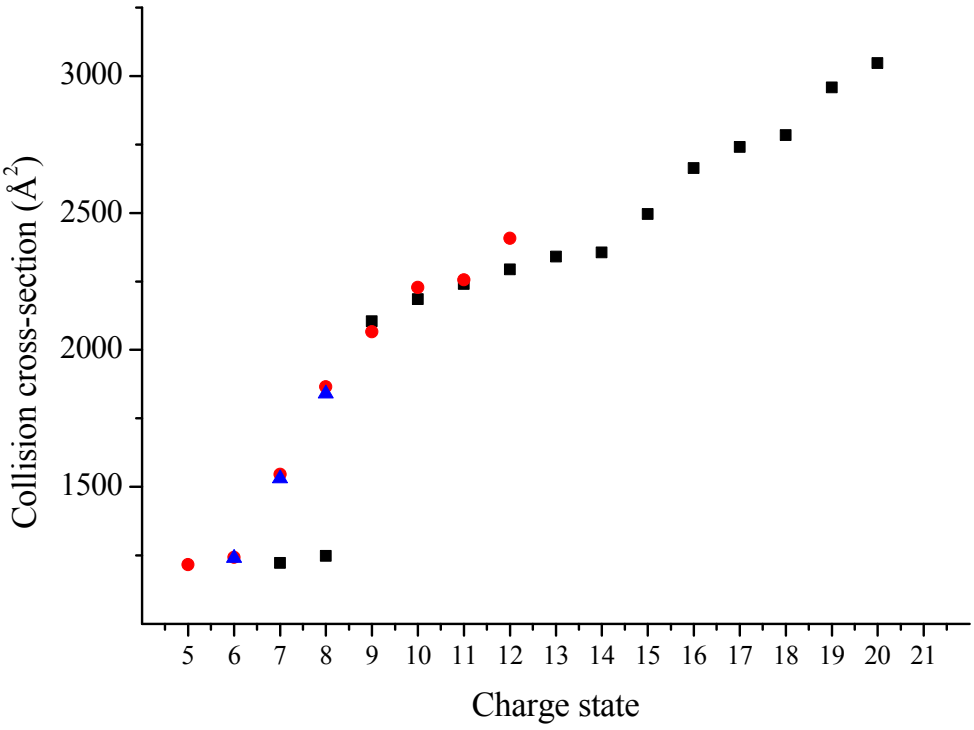



*Figure 3.10. Cytochrome c monomer cross-sections from native-like conditions (blue triangles); from aqueous solution (red circles); and from denaturing conditions (black squares).*

This plot shows that monomer cross-sections obtained from protein sprayed from ammonium acetate or water have similar cross-sections over the charge state range z = $6^+$ to $8^+$. Other charge states from aqueous conditions, z = $5^+$ and $9^+$ to $12^+$, may have come from gas-phase dimer dissociation that would not be found under native-like conditions. Additional evidence for this suggestion is from comparing pH 3.0 and aqueous protein cross-sections in the range z = $9^+$ to $12^+$; pH 3.0 monomer under the influence of organic solvent and acid has been protonated and unfolded to adopt extended gas-phase conformations. Aqueous conditions are not denaturing, and should not lead to extended conformations; therefore these more extended conformations must arise from dimer dissociation with monomer unfolding prior to dissociation from the dimer assembly. This may also be assumed for the z = $5^+$ species formed under aqueous solutions being compact, but absence of data from either denaturing or native-like conditions cannot provide solid evidence. Of interest are the pH 3.0 cross-sections of z = $7^+$ and $8^+$. These species are significantly more compact (~21% and 33% respectively) than either species of the same charge from aqueous or native-like conditions. This may be attributed to initial unfolding of the z = $7^+$ and $8^+$ in solution under denaturing conditions to form a semi-extended structure (perhaps domain, C- or N-terminal unfolding, or foldon unfolding) and upon subsequent transfer to the gas phase, the structure has been able to adopt a more compact conformer due to gas-phase collapse of the semi-extended structure.

## 3.9 Native Electron Capture Dissociation - Evidence from Mass Spectrometry

Native electron capture dissociation (NECD)[22, 28, 29] was reported by Breuker and McLafferty in 2003 when they observed what appeared to be ECD mass spectra of cytochrome *c* that instead reflected the native protein structure formed without the experimental addition of electrons. ECD is a dissociation technique whereby electrons cleave the protein backbone without rupture of the non-covalent bonding of the tertiary structure.[8, 30] Characteristic fragments are produced due to cleavage at the N-$C_\alpha$ bond (c and $z^\bullet$ ions) and $C_\alpha$-N and $C_{carbonyl}$-N amide bond ($a^\bullet$ and y ions). In dissociation techniques such as collision-induced dissociation, b and y fragment ions are observed. Thus the presence of c ions without b ions was of note in the absence of electrons. NECD was observed[28] in cytochrome *c* solution concentrations above 33 μM, but above 100 μM the fragment yield decreased (possibly due to competitive formation of higher order aggregates). Temperature and solution pH also influenced NECD fragmentation, with decreased temperature and pH decreasing fragment abundance. As c and y ions were only identified from native cytochrome *c* ESI spectra, the authors suggested that the structure of the compact native structure may provide a mechanistic rationalisation of the unexpected ion formation. One electron must be transferred to the haem group to promote $Fe^{II}$ formation, and the second electron causes fragmentation of the peptide backbone on amino acid residues that are adjacent to the haem ring. The most prominent fragments arose from cleavage on the N-terminal side of Met80, whose sulfur atom distally coordinates the haem iron centre. Other abundant products originate from cleavages next to Thr40, Thr49, and Asn52 which are all hydrogen bonded to the carboxylate units of haem propionates 7, 6, and 7 respectively in $Fe^{III}$ cytochrome *c*.[5]

The proposed NECD mechanism[28] involves one monomer subunit within the dimer unfolding in the source region. The high-energy transition state in solution unfolding involves unfolding of the N- and C-terminal (residues 3 to 14, and 90 to 101 respectively) by cleavage of the hydrophobic Phe10-Leu94 interhelix bond. This helical regions contain eight basic residues that will be exposed upon unfolding and can be protonated. The native structure is assumed to have ~8 protons on the monomer originally, and three of them have been proposed to be found on these helical regions. Partial denaturation attracts five protons from the other monomer, sufficient to protonate the newly exposed basic residues. Eight protons now reside on the terminal helices. This causes charge asymmetry in the dimer, and electrons from the unfolded monomer are attracted to the partially unfolded monomer. One electron reduces the $Fe^{III}$ centre, and another electron is transferred to the haem ligand to affect the ECD process. This mechanism is shown in Figure 3.11.

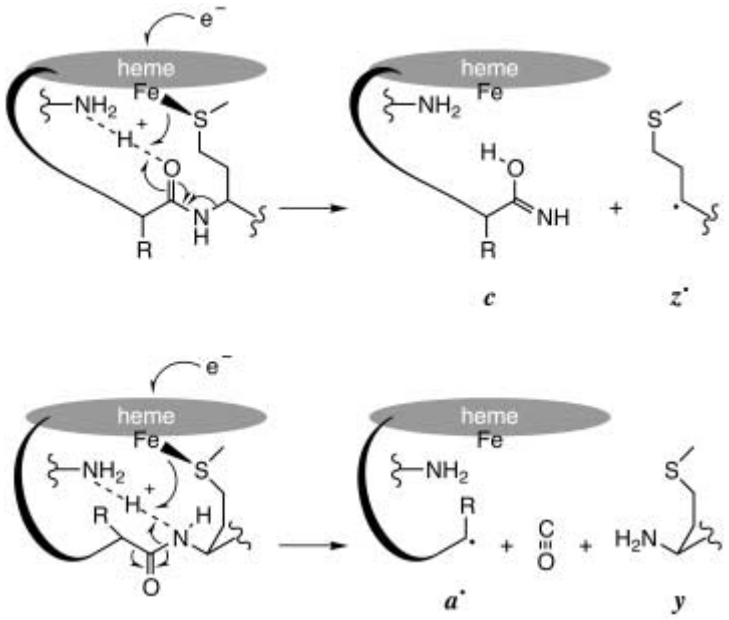



*Figure 3.11.[28] Proposed mechanism of NECD fragmentation near Met80.*

Further work on NECD and cytochrome *c* dimers revealed that additional cleavages could be observed if the protein solution was left for a period of three months,[29] indicating dimers can adopt multiple different structures. Dimer structure from stored solutions show greater stabilisation of protein-haem bonds in the monomer that unfolds upon the NECD process (monomer I) so that they are still intact when NECD occurs. Cleavage products from a solution of freshly made up 75 μM cytochrome *c* in water (pH 5.0) are consistent with monomer II (the compact monomer that gives up electrons in the NECD mechanism) binding near K79/M80 of monomer I (the unfolding monomer). NECD spectra do not show fragments from cleavage in the terminal regions (next to F10, L94, L98 or in the 18-34 segment) complementing evidence that initial unfolding causing charge asymmetry involves separation of the terminal helices, rupture of the H18 to haem iron coordinate bond (which will increase the electron affinity of the haem group) and unfolding of the Ω-loop (His18 to Gly34).

## 3.10 Ion Mobility-Mass Spectrometry of Fragments Derived from Native Electron Capture Dissociation of Cytochrome *c*

Mass spectrometry and ion mobility-mass spectrometry analyses were performed on an aqueous solution of 62 μM cytochrome *c* at pH 4.5. A number of fragments could be observed in the mass range 900 – 1350 *m/z* by MS as shown in the mass spectrum (Figure 3.12); however due to low signal intensity only a small number of fragments produced could be analysed by IM-MS and cross-sections elucidated. Table 3.5 presents the eight fragment cross-sections that could be produced.

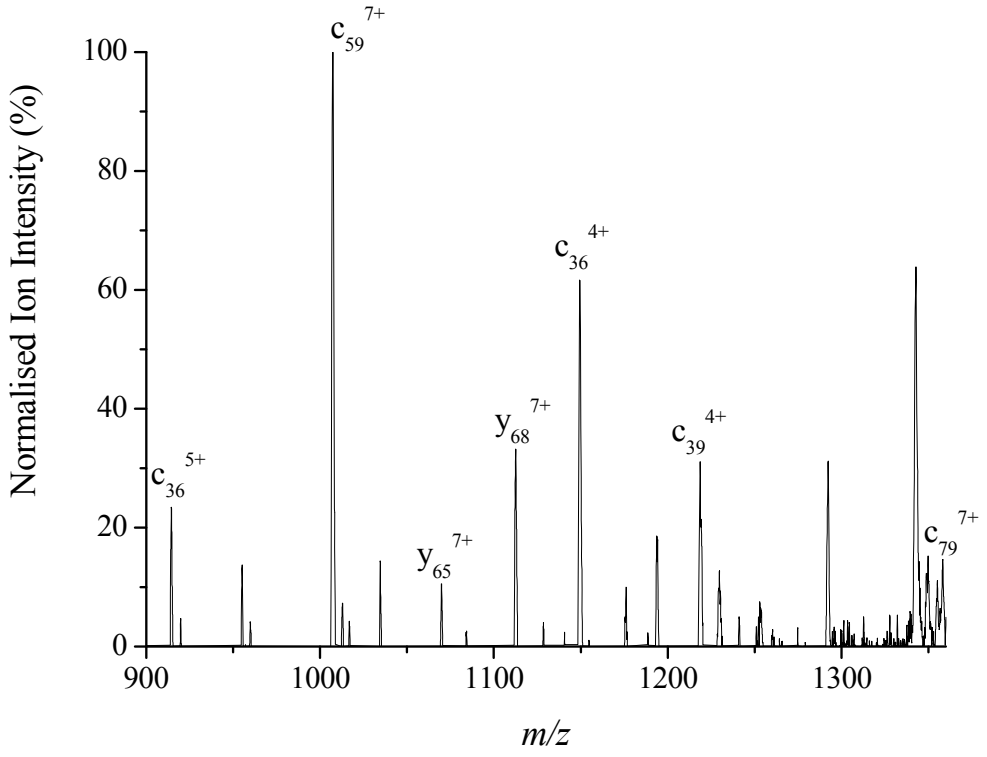



*Figure 3.12. Mass spectrum of aqueous 62 μM cytochrome c at pH 4.5 showing a number of fragments produced by the NECD process.*

*Table 3.5. Collision cross-sections of cytochrome c fragments produced from NECD.*

| Fragment ion | Charge (+ve) | *m/z* value observed (*m/z* calculated) | Collision cross-section ($Å^2$) |
|---|---|---|---|
| $c_{36}$ | 4 | 1144.4 (1144.8) | 1141 |
| $c_{36}$ | 6 | 763.1 (763.5) | 1342 |
| $c_{39}$ | 4 | 1229.4 (1230.1) | 957 |
| $c_{39}$ | 5 | 983.8 (984.3) | 1410 |
| $c_{59}$ | 7 | 1010.1 (1010.6) | 1396 |
| $c_{79}$ | 7 | 1351.9 (1352.4) | 1063 |
| $y_{68}$ | 6 | 1299.7 (1300.0) | 1372 |
| $y_{68}$ | 7 | 1113.9 (1114.4) | 1607 |

In table 3.5, $c_n$ denotes ion fragments that are on the N-terminal side, and $y_n$ denotes fragments on the C-terminal side. The subscript n is the number of amino acids on the fragment from the N-terminal (c ions) or C-terminal (y ions). Five different fragments were elucidated by IM-MS with three of these displaying two different charge states. Differing charge states were consecutive (*i.e.* z = n and n + 1)

for two of them, but for $c_{36}$ the charge states were $z = 4^+$ and $6^+$ ($z = 5^+$ could not be elucidated but was observed in the mass spectrum).

Two charge states for the quasi-complement fragments $c_{36}$ and $y_{68}$ were detected which are outside the Ω-loop, and have been identified in previous work at the same charge states.[22] All fragments show charge states of $z \geq 4$ due to the available number of basic residues for protonation. Fragment $c_{36}$ has seven lysine and two histidine residues, with four lysines on the N-terminal helix (residues 3 to 14) possibly accounting for the observed $c_{36}^{4+}$ fragment observed from IM-MS data. This assumes that fragment separation after backbone cleavage is faster than any further proton transfer reactions (ECD has been shown to be faster than intramolecular proton scrambling). This cross-section (1141 $Å^2$) increases upon further protonation to $z = 6^+$ to 1342 $Å^2$. Fragment $y_{68}$ shows higher charge states than its quasi-complement fragment due to the presence of twelve lysine and two arginine residues. Significant unfolding is observed between the two charge states, $z = 6^+$ and $7^+$ perhaps indicating domain unfolding.

Fragment $c_{39}$ contains eight lysines, two histidines and one arginine; and $c_{59}$ contains ten lysines, two histidines and one arginine. Fragment ion $c_{39}^{5+}$ and $c_{59}^{7+}$ cross-sections differ by only 14 $Å^2$ but by two charges. Only two further lysine residues are available in the $c_{59}$ fragment in segment V (see Figure 3.13) which are most likely the protonation sites for the additional two protons and do not cause significant unfolding. This may also imply that segment IV (residues 40 to 48 – no acidic or basic residues) has unfolded to the same extent in both fragments and adopted similar gas-phase conformations. Fragment $c_{79}$ contains fourteen lysines, two histidines and one arginine. Lack of the C-terminal helix (which is known to unfold in the initial stages) would provide a compact cross-section due to intact intramolecular hydrogen bonding, but comparison with $c_{59}^{7+}$ (which is 333 $Å^2$ more extended) implies that the additional twenty amino acid sequence (in segments VI and VII) assist in stabilising the structure so as to adopt a more ordered and compact gas-phase structure. Figure 3.13[22] shows the segments (I to VIII) of the cytochrome *c* sequence with the basic residues highlighted. Cleavages displayed relate to the most abundant fragments observed in previous work. Cross-sections have obtained for five of these fragments with the smallest cross-section recorded was that of the $c_{39}^{4+}$ (957 $Å^2$) fragment which is smaller than the similarly related $c_{36}$ ($z = 4^+$, 1141 $Å^2$). This is possibly due to fragmentation of two different dimer species. The largest cross-section, 1607 $Å^2$ for $y_{68}^{7+}$ shows a large amount of protein unfolding and complements the charge observed in previous work.[22] This fragment contains a large number of basic residues (thirteen in total) that could be protonated upon unfolding, and its proximity to the haem ring (Met80 adjacent to it is coordinated to the iron centre) adds support to the electron capture mechanism (Figure 3.11).

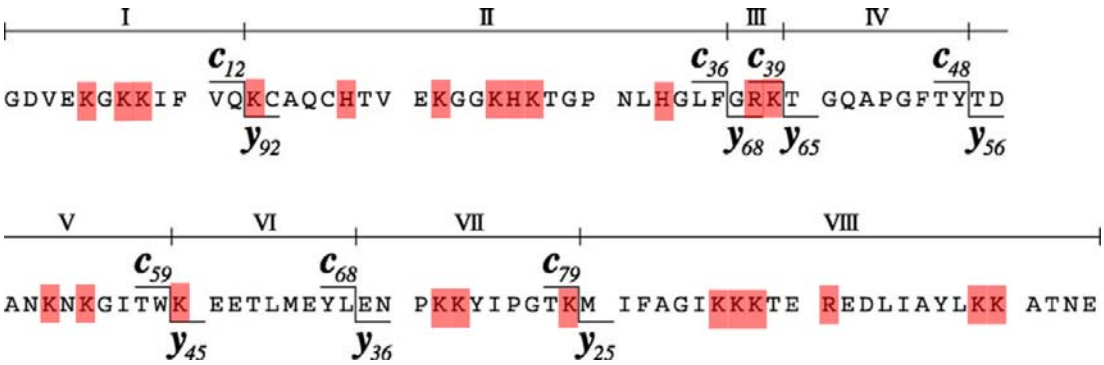



*Figure 3.13.[22] Cytochrome c sequence showing the eight segments (I to VIII) the protein is divided into and the most abundant fragments from previous work on an FT-ICR instrument.*

## 3.11 Cytochrome *c* Conclusions

Cytochrome *c* is able to adopt a large number of gas-phase conformations that result from the solvent composition and concentration from which they are obtained. Monomers unfold upon sequential protonation from compact structures to extended structures under native-like, aqueous and denaturing conditions. Spraying from increased concentrations provides interesting results: solution aggregation to form higher order multimers such as dimer and trimer species can be stabilised and studied by mass spectrometry. Subsequent dissociation of dimer species to monomer species is rarely a simple halving of the dimer charge: complex gas-phase dissociation processes must occur depending upon the dimer charge and the conformational flexibility of the constituent monomer subunits. Frequently, one monomer subunit will unfold to a greater degree than its counterpart and will be protonated to a greater extent. Dissociation can occur at this point to form monomers species with disproportionate charge and cross-sections similar to those obtained under denaturing (unfolded) conditions. At concentrations that favour dimer formation but not aggregation to higher order multimers, an interesting gas-phase dissociation mechanism is revealed that has only been studied using FT-ICR-MS up until now. Native electron capture dissociation of dimeric cytochrome *c* to form fragment ions from the backbone adjacent to the haem ligand show a range of charge states and cross-sections linked to the unfolding of one monomer subunit whilst bound to a second folded monomer in the dimer. Again, one monomer unfolds to a greater degree and causes a proton charge asymmetry. Transfer of two electrons from the folded monomer causes reduction of the $Fe^{III}$ centre to $Fe^{II}$ and the other causing protein backbone cleavage, but retaining a degree of tertiary structure.

Further research is required to understand the interesting gas-phase dissociation mechanisms underlying the dimer species.

## 3.12 Chapter 3. References

1. Pascher, T.; Chesick, J. P.; Winkler, J. R.; Gray, H. B., Protein Folding Triggered by Electron Transfer. *Science* **1996,** 271, (5255), 1558-1560.

2. Jiang, X.; Wang, X., Cytochrome c-mediated Apoptosis. *Annu. Rev. Biochem.* **2004,** 73, (1), 87-106.

3. Zhao, T.; King, F. L., Direct Determination of the Primary Binding Site of Cisplatin on Cytochrome c by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (6), 1141-1147.

4. Banci, L.; Bertini, I.; Rosato, A.; Varani, G., Mitochondrial cytochromes c : a comparative analysis. *J. Biol. Inorg. Chem.* **1999,** 4, (6), 824-837.

5. Banci, L.; Bertini, I.; Gray, H. B.; Luchinat, C.; Reddig, T.; Rosato, A.; Turano, P., Solution Structure of Oxidized Horse Heart Cytochrome *c*. *Biochemistry* **1997,** 36, (32), 9867-9877.

6. Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E., Monitoring structural changes of proteins in an ion trap over approximately 10-200 ms: unfolding transitions in cytochrome c ions. *Anal. Chem.* **2001,** 73, (24), 6000-7.

7. Badman, E. R.; Myung, S.; Clemmer, D. E., Evidence for Unfolding and Refolding of Gas-Phase Cytochrome c Ions in a Paul Trap. *J. Am. Soc. Mass Spectrom.* **2005,** 16, (9), 1493-1497.

8. Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W., Kinetic Intermediates in the Folding of Gaseous Protein Ions Characterized by Electron Capture Dissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2001,** 123, (40), 9792-9799.

9. Maity, H.; Maity, M.; Krishna, M. M. G.; Mayne, L.; Englander, S. W., Protein folding: The stepwise assembly of foldon units. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** 102, (13), 4741-4746.

10. Jurchen, J. C.; Williams, E. R., Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. *J. Am. Chem. Soc.* **2003,** 125, (9), 2817-2826.

11. Loo, J. A.; Loo, R. R. O.; Udseth, H. R.; Edmonds, C. G.; Smith, R. D., Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1991,** 5, (3), 101-105.

12. Kaltashov, I. A.; Eyles, S. J., Studies of biomolecular conformations and conformational dynamics by mass spectrometry. *Mass Spectrom. Rev.* **2002,** 21, (1), 37-71.

13. Clemmer, D. E. http://www.indiana.edu/~clemmer.

14. McCullough, B. J.; Kalapothakis, J.; Eastwood, H.; Kemper, P.; MacMillan, D.; Taylor, K.; Dorin, J.; Barran, P. E., Development of an Ion Mobility Quadrupole Time of Flight Mass Spectrometer. *Anal. Chem.* **2008,** 80, (16), 6336-6344.

15. Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **1997,** 119, (9), 2240-2248.

16. Faull, P. A.; Korkeila, K. E.; Kalapothakis, J. M.; Gray, A.; McCullough, B. J.; Barran, P. E., Gas-phase metalloprotein complexes interrogated by ion mobility-mass spectrometry. *Int. J. Mass Spectrom.* **2009,** 283, (1-3), 140-148.

17. Sanishvili, R.; Volz, K. W.; M., W. E.; Margoliash, E., The low ionic strength crystal structure of horse cytochrome *c* at 2.1 Å resolution and comparison with it's high ionic strength counterpart. *Structure* **1995,** 3, (7), 707-716.

18. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996,** 100, (40), 16082-16086.

19. Shvartsburg, A. A.; Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996,** 261, (1-2), 86-91.

20. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H., Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **2008,** 22, (20), 3297-304.

21. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Molec. Graphics* **1996,** 14, (1), 33-38.

22. Breuker, K., Segmental charge distributions of Cytochrome c on transfer into the gas phase. *Int. J. Mass Spectrom.* **2006,** 253, (3), 249-255.

23. Rostom, A. A.; Fucini, P.; Benjamin, D. R.; Juenemann, R.; Nierhaus, K. H.; Hartl, F. U.; Dobson, C. M.; Robinson, C. V., Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, (10), 5185-5190.

24. Felitsyn, N.; Kitova, E. N.; Klassen, J. S., Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior. *Anal. Chem.* **2001,** 73, (19), 4647-61.

25. Light-Wahl, K. J.; Schwartz, B. L.; Smith, R. D., Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **1994,** 116, (12), 5271-5278.

26. Versluis, C.; van der Staaij, A.; Stokvis, E.; Heck, A. J. R.; de Craene, B., Metastable ion formation and disparate charge separation in the gas-phase dissection of protein assemblies studied by orthogonal time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2001,** 12, (3), 329-336.

27. Jurchen, J. C.; Garcia, D. E.; Williams, E. R., Further studies on the origins of asymmetric charge partitioning in protein homodimers. *J. Am. Soc. Mass Spectrom.* **2004,** 15, (10), 1408-1415.

28. Breuker, K.; McLafferty, F. W., Native electron capture dissociation for the structural characterization of noncovalent interactions in native cytochrome c. *Angew. Chem. Int. Ed. Engl.* **2003,** 42, (40), 4900-4.

29. Breuker, K.; McLafferty, F. W., The thermal unfolding of native cytochrome c in the transition from solution to gas phase probed by native electron capture dissociation. *Angew. Chem. Int. Ed. Engl.* **2005,** 44, (31), 4911-4.

30. Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., Detailed Unfolding and Folding of Gaseous Ubiquitin Ions Characterized by Electron Capture Dissociation. *J. Am. Chem. Soc.* **2002,** 124, (22), 6407-6420.

# Chapter 4. Mass Spectrometry & Ion Mobility-Mass Spectrometry Studies of Haemoglobin

*Chapter 4 presents the largest metalloprotein studied in this Thesis – Haemoglobin. With a molecular weight in excess of 64 kDa, haemoglobin is the largest biomolecule to be investigated to date using the MoQTOF and required extensive tuning and parameter optimisation to obtain reproducible signal. Signal attributed to monomer, dimer and tetramer species is characterised by mass spectrometry, and collision cross-sections for a range of gas-phase species is calculated by ion mobility-mass spectrometry. Comparisons with computationally-derived cross-sections have also been made.*

## 4.1 Haemoglobin Introduction

Haemoglobin (abbreviated Hb) is an iron-containing haemoprotein composed of two alpha ($\alpha_{apo}$; 15,126.2 Da) and two beta ($\beta_{apo}$; 15,867.2 Da) monomer subunits.[1] Each subunit is capable of binding a single haem ligand (616.5 Da) in a hydrophobic pocket. These monomer subunits arrange into a $\alpha\beta$ dimer which further associate with a second dimer to form a $\alpha_2\beta_2$ tetramer containing four haem ligands (Figure 4.1).

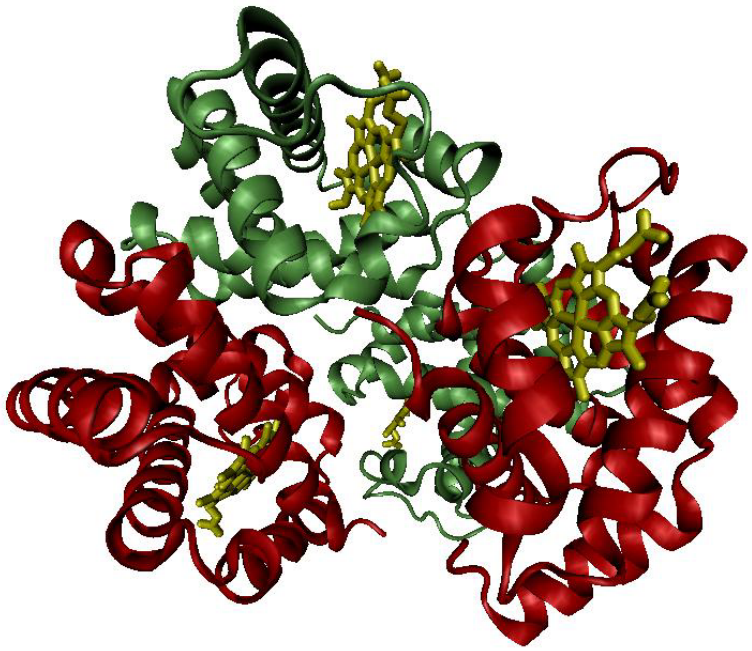


*Figure 4.1. Haemoglobin tetramer molecule showing alpha (green) and beta (red) subunits that make up the tetramer. Haem ligands are shown in yellow. PDB ID: 1GZX.[1]*

This tetramer structure is the biologically active protein form and transports molecular oxygen ($O_2$; molecular weight 32.0 g $mol^{-1}$) *via* binding with a non-covalently bound prosthetic haem group within

each monomer subunit. In human haemoglobin, the alpha subunit contains 141 amino acid residues (10 histidines, 11 lysines and 3 arginines); the beta subunit contains 146 (9 histidines, 11 lysines, and 3 arginines) and there is a sequence homology of ~43% between the two monomer chains. This homology provides a number of similarities, but more importantly differences, between the subunits that will be discussed later in this Chapter. The four subunits are positioned approximately at the vertex of a tetrahedron with each subunit facing one another across a central cavity, thereby creating a two fold rotation of symmetry within the tetramer molecule. Typical haemoglobin concentrations within a mammalian erythrocyte have been determined on the order of 5 mM.[2]

The haemoglobin active site is a protoporphyrin IX (haem B) system that contains a central iron ion bound to four pyrrole nitrogen atoms in a heterocyclic organic ring system.[3] The iron ion can exist in either the $Fe^{II}$ (ferro-) or $Fe^{III}$ (ferri-) oxidation state but to perform oxygen binding, iron must be in the $Fe^{II}$ state. Upon binding molecular oxygen in an, "end-on bent" geometry, iron oxidation to the $Fe^{III}$ state occurs, returning to $Fe^{II}$ upon oxygen dissociation. Should the iron be oxidised to $Fe^{III}$ in the absence of oxygen, methaemoglobin is produced which is unable to bind oxygen. Methaemoglobin reductase enzyme is responsible for converting methaemoglobin to biologically-viable haemoglobin.[4] The haem group is stabilised within a hydrophobic pocket of the globular protein. Electrostatic interactions between haem propionate groups and the protein, as well as hydrophobic interactions, support the main interaction of a proximal histidine residue (His87 on the $\alpha$ subunit and His92 on the $\beta$ subunit) to the iron centre. When oxygen is not bound to the iron centre, a distal histidine residue (His58 on the $\alpha$ subunit, His63 on the $\beta$ subunit) aids in preventing iron oxidation and also impedes carbon monoxide association, thereby reducing the risk of oxygen starvation.

Biologically-viable Hb tetramer can exist in two forms: *oxy*- and *deoxy*haemoglobin. These two states affect the three dimensional protein quaternary structure and its ability to bind oxygen at available haem sites. The T (tense) state is populated when Hb does not have any oxygen bound to its monomer subunits, the name arising from the presence of additional stabilising interactions between the subunits. In the presence of oxygen, a conformational change to the R (relaxed) state occurs. This is achieved by the cooperative nature of the protein. Deoxyhaemoglobin has a low affinity for oxygen, but upon binding the first oxygen molecule, movement in the protein backbone enable the second oxygen molecule to bind more easily, until facile binding of the third and fourth oxygen molecules are achieved.[5] This allostery is mirrored in the dissociation of oxygen at respiring tissues *i.e.* oxyhaemoglobin has a high affinity for oxygen therefore loss of the first oxygen molecule is difficult, but upon dissociation allows the loss of the remaining oxygen molecules to occur with increasing ease.

## 4.2 Mass Spectrometry Research on Haemoglobin

Much research has been focused on haemoglobin and myoglobin.[6-15] Myoglobin (Mb) is a monomeric protein (153 amino acids) that shares a similar helical structure and composition to that of the haemoglobin monomeric subunits. Work on studying haemoproteins in the gas phase by mass spectrometry has been directed at the interaction between protein and prosthetic haem ligand. Blackbody infrared radiative dissociation (BIRD) was utilised by Gross *et al.* to show that myoglobin is thermally more stable than the haemoglobin α-monomer with respect to haem loss.[12] Schmidt *et al.* investigated the influence of electrostatic interactions between haem ligand and haemoproteins.[16] Under positive electrospray ionisation conditions, loss of haem from *holo*-myoglobin and haemoglobin ions was achieved. Signal indicating the presence of ferri-haem (singly charged species) was detected, providing experimental evidence for a gas-phase complex between an $Fe^{III}$ ion and a $z = 2^-$ charged porphyrin ring system. Under negative ESI conditions, they observed a shift in the charge state distribution towards less charged (higher *m/z* ratio) ions for tetrameric haemoglobin and a considerably enhanced gas-phase stability for both the holo-forms of α-monomer haemoglobin and myoglobin.

Griffith and Kaltashov[17] utilised mass spectrometry and circular dichroism (CD) techniques in an attempt to estimate relative fractions of bovine haemoglobin species in solution as a function of pH. ESI mass spectrometry was used to monitor haemoglobin association/dissociation equilibria over a pH range of 3.0 to 10.0. At low pH, monomeric species dominated the spectra. $\alpha_{holo}$ monomers were capable of retaining haem-binding capacity until the pH was lowered to 4.0, with haem dissociation complete at pH 3.0. Evidence for a significant fraction of highly structured α chains under low pH solution conditions was observed due to the presence of a bimodal charge state distribution. $\beta_{holo}$ monomer chains were not observed at any pH, however $\beta_{apo}$ was observed maintaining an unfolded structure as determined by the broad charge state distribution and higher average charge density compared to the α chains. At pH values greater than 4.0, it has been proposed that the protein is expected to exist in the natively-structured state[18] and the mass spectra showed an abundance of oligomeric haemoglobin species. To obtain a better understanding of the monomer conformers present, the authors used a chemometric technique (Singular Value Decomposition)[19] and assigned four conformers for the haemoglobin species observed based upon previous results obtained with myoglobin: N, the most compact state; I, an intermediate state (pH 4.0 intermediate); E, an extended conformation; and U, an unfolded state (random coil). Results published showed that at pH 8.0, the $\alpha_{holo}$ monomer displayed a high proportion (~70%) of N state with a smaller contribution from I and E states, whereas $\beta_{apo}$ displayed a range of conformations from the most compact to the unfolded state (U). Decreasing pH to 4.0 significantly changed the profile of the $\alpha_{holo}$ monomer with a reduction of ~50% in the N state, and the presence of $\alpha_{apo}$ dominating in the extended conformation indicated a loss of haem-binding ability and structured folding. Interestingly, the $\beta_{apo}$ monomer profile at pH 4.0 did not significantly change with only a small decrease in N state and small increase in E state

observed. CD spectroscopy allowed direct interrogation of the haem environment by monitoring the Soret band ($\lambda$ = 410 – 430 nm) where absorption arises due to coupling of the $\pi \rightarrow \pi^*$ transitions of the haem ligand and neighbouring protein aromatic amino acid residues.[20] Upon lowering the pH, the band diminished as the haem environment was disturbed and the ligand lost. At intermediate pH values (5.0 to 7.0), the intensity is strong and fairly constant indicating a high percentage of haem-bound protein. Unfortunately at pH values above neutral, the Soret band cannot be used as a reliable indicator of protein-haem binding due to deprotonation of non-coordinating aromatic residues.[21]

It has been proposed[22] that three distinct pathways of haemoglobin tetramer assembly may exist:

1. *via* a haem-containing heterodimer ($\alpha_{holo}\beta_{holo}$);
2. *via* a pseudo-holo haemoglobin dimer ($\alpha_{holo}\beta_{apo}$);
3. or *via* a dimer containing no haem ligands ($\alpha_{apo}\beta_{apo}$).

Evidence[23] has been provided for the pathway that includes an $\alpha_{holo}\beta_{apo}$ pseudo-holo heterodimer intermediate that ensures correct formation of the active haemoglobin tetramer. Griffith and Kaltashov proposed[17] an *in vitro* tetramer association pathway beginning with binding of a natively folded $\alpha_{holo}$ monomer and partially unfolded $\beta_{apo}$ monomer to form a highly structured haem-deficient dimer $\alpha_{holo}\beta_{apo}$. This is followed by haem acquisition to form the $\alpha_{holo}\beta_{holo}$ dimer and then subsequent association of two $\alpha_{holo}\beta_{holo}$ dimers to form the tetramer. Recent haemoglobin mass spectrometry and ion mobility-mass spectrometry research by Scarff *et al.*[24] show that the use of commercially available lyophilised haemoglobin in experiments may not provide a true picture of the protein interactions occurring *in vivo*. Scarff utilised fresh whole blood samples and a Synapt HDMS instrument to produce tetramer, dimer and monomer species as reported in this Thesis. Variations from Scarff's work to work up until now include:

i) the presence of the $\beta_{holo}$ monomer;
ii) absence of the $\alpha_{holo}\beta_{apo}$ pseudo-holo heterodimer;

Scarff's work suggests the haem-deficient dimer ($\alpha_{holo}\beta_{apo}$) is not an intermediate in tetramer formation (*via* association of $\alpha_{holo}$ and $\beta_{apo}$) and that $\beta_{apo}$ is potentially not as extensively disordered as believed. Ongoing research will further clarify if upon removal of haemoglobin from the *in vivo* environment and subsequent lyophilisation affects gas-phase dissociation.

Ion mobility-mass spectrometry studies of haemoglobin to obtain cross-sections have been performed using travelling wave[24] technology with favourable comparisons made to solution and condensed-phase structural data. Within the research group of Don Douglas (University of British Columbia, Canada), haemoprotein cross-section research[7, 11, 25-27] using energy loss experiments in a triple quadrupole mass spectrometer provide an alternate pathway to ion mobility to calculate cross-

sections. In energy loss experiments, ions are injected into an RF quadrupole and the loss of kinetic energy from neutral gas collisions is measured and related to the cross-section through kinetic theory.[11] Cross-sections obtained by this method are ~13 to 20% smaller than cross-sections calculated from hard-sphere scattering.[11]

## 4.3 Haemoglobin Experimental

As detailed in Chapter 2, haemoglobin stock solutions of 120 μM were produced in 50 mM ammonium acetate (pH 6.85) and diluted as required. pH adjustments were required for haemoglobin for acidic and basic conditions. To collect mass spectra in traditional mass spectrometry mode, the MoQTOF drift cell was filled with ~3.0 Torr helium and had a temperature of ~300 K. The drift voltage was fixed at 60V. Mass spectrometry and ion-mobility mass spectrometry source conditions were optimised for the species of interest: nano-ESI capillary voltage range was typically ~1.7 – 2.5 kV; sample cone range of 145 to 165V; extraction cone 105 to 125V; source temperature 80$^{o}$C; source pressure ~$4.0 \times 10^{-2}$ mbar for monomer mass spectra and $4.0 \times 10^{-1}$ mbar for mass spectra with tetramers present. Ion mobility settings were as above but with incremental decreases on the C1 voltage of 60V, 50V, 40V, 35V, 30V, 25V, 20V, 15V and 10V with the same change to the extractor cone voltage; pulse width 40 μs; stopping voltage (TH1) 125 to 140V; pusher period 100 to 120 μs (mass range dependent).

## 4.4 Haemoglobin Results

Haemoglobin experiments were divided into three areas: native-like conditions; denaturing conditions with acid; and elevated pH conditions. Mass spectrometry and ion mobility-mass spectrometry results are presented. For clarity, haemoglobin species will be denoted as set out in Table 4.1.

*Table 4.1. Haemoglobin notation used throughout this Chapter. The final column specifies whether the species contains the haem ligand or not.*

| Species | Notation | Contains haem ligand? |
|---|---|---|
| α globin monomer – apo | $\alpha_{apo}$ | No |
| α globin monomer – holo | $\alpha_{holo}$ | Yes |
| | | |
| β globin monomer - apo | $\beta_{apo}$ | No |
| β globin monomer - holo | $\beta_{holo}$ | Yes |
| | | |
| αβ dimer – apo | $\alpha\beta$ or $\alpha_{apo}\beta_{apo}$ | No |
| αβ dimer – pseudo-holo heterodimer | $\alpha_{holo}\beta_{apo}$, $\alpha_{apo}\beta_{holo}$ | Yes, within one subunit only |
| αβ dimer – holo or holo heterodimer | $\alpha_{holo}\beta_{holo}$ | Yes, within both subunits |
| | | |
| $\alpha_2\beta_2$ tetramer | $(\alpha_{holo}\beta_{holo})_2$ | Yes, within all four subunits |

### 4.4.1 Native-Like Conditions Studied by Mass Spectrometry

A solution of 30 μM haemoglobin in 50 mM ammonium acetate at pH 6.85 was sprayed in positive ion mode by nano-electrospray ionisation. Figure 4.2 is the resultant mass spectrum displaying a plethora of gas-phase species observed over a wide *m/z* range (500 to 3900 *m/z*).

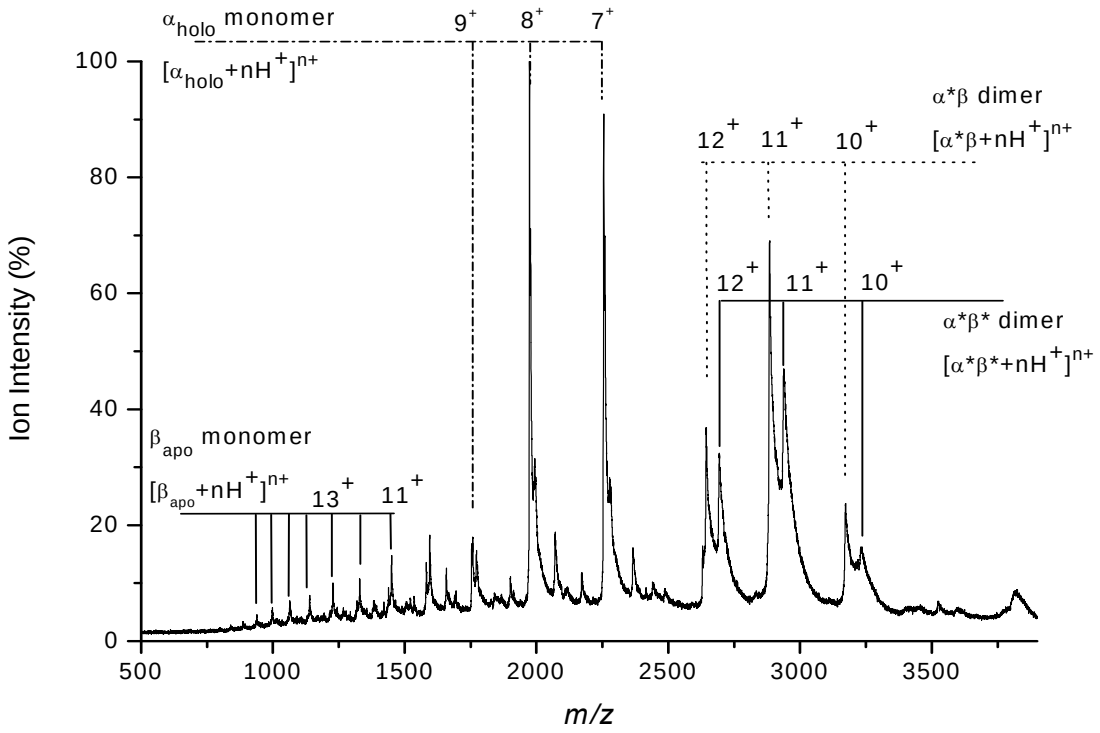



*Figure 4.2. Mass spectrum over the range 500 to 3900 m/z of 30 μM haemoglobin in 50 mM ammonium acetate at pH 6.85 displaying a number of gas-phase multimeric species.*

The most intense peaks correspond to $\alpha_{holo}$ with a narrow charge state distribution ($z = 7^+$ to $9^+$) and a low charge density. This is perhaps expected for a holo protein species that has sufficient preservation of the native fold to retain the haem ligand. There are no peaks detected for the $\alpha_{apo}$ species implying that $\alpha_{holo}$ is a well-organised structure that does not undergo haem ligand loss or sufficient structural rearrangement in the gas phase, under these experimental conditions, so as to lose the non-covalent interactions necessary for the haem ligand to be maintained in the hydrophobic pocket.

At lower *m/z* values (< 1500 *m/z*), a series of peaks have been assigned to $\beta_{apo}$ globin monomer ions. A wide charge state distribution is observed with higher charge densities than that of the $\alpha_{holo}$ monomer indicating the $\beta_{apo}$ monomer has partially unfolded and lost the haem ligand. Gas-phase dissociation of the tetramer multimer to the constituent monomer species has destabilised the β-monomer structure to a greater extent than that of the α-monomer.

At higher *m/z* values (> 2500 *m/z)*, multimeric species indicate experimental conditions were sufficient to preserve non-covalent interactions and quaternary structure into the gas phase. Two dimeric species are observed:

i) the $\alpha_{holo}\beta_{holo}$ (holo heterodimer) species with both monomer chains retaining haem ligands;
ii) the $\alpha_{holo}\beta_{apo}$ (pseudo-holo heterodimer) species.[28]

Both dimeric species have a narrow charge state distribution centred at $z = 11^+$ which may indicate compact tertiary structure.[29] Unfolding of these species may produce the monomeric species observed at lower *m/z* values. The pseudo-holo heterodimer can be assigned as $\alpha_{holo}\beta_{apo}$ because dissociation of this species to $\alpha_{holo}$ and $\beta_{apo}$ is evidenced in the mass spectrum, and further asserted by the absence of $\alpha_{apo}$ and $\beta_{holo}$ peaks. The final species assigned in the mass spectrum are the $[(\alpha_{holo}\beta_{holo})_2 + 16H]^{16+}$ and $[(\alpha_{holo}\beta_{holo})_2 + 17H]^{17+}$ tetramers. The narrow charge state distribution reflects the compact structure and relative structural homogeneity of the tetramer species.[17, 29] Under these conditions, the tetramer signal was weak; however upon deconvolution of the mass spectrum a small peak for tetramer was present.

### 4.4.2 Haemoglobin Studied by Mass Spectrometry at pH 4.5

Addition of 2% by volume formic acid (solution pH 4.5) to haemoglobin induces structural movement and dissociation of the tetramer to monomer subunits. Figure 4.3 is a mass spectrum (range 600 to 2500 *m/z*) showing a wide charge state distribution and high charge density characteristic of denatured and unfolded proteins. All peaks can be attributed to $\alpha_{apo}$ or $\beta_{apo}$ species and both monomer species exhibit a significant degree of unfolding and disordered structure, having lost their haem-binding interactions. Unbound haem ligand is not soluble under aqueous conditions[30] and is not observed in the mass spectrum. A trimodal charge state distribution indicates three possible $\alpha_{apo}$ conformations in solution equilibrium[31] observed *in vacuo*:

i) Unfolded conformer centred at $[\alpha_{apo} + 17H]^{17+}$.
ii) Partially unfolded conformer centred at $[\alpha_{apo} + 12H]^{12+}$.
iii) Semi-folded conformer centred at $[\alpha_{apo} + 8H]^{8+}$.

Unfolded conformers dominate the distribution. A similar $\beta_{apo}$ distribution is observed:

i) Unfolded conformer centred at $[\beta_{apo} + 16H]^{16+}$.
ii) Partially unfolded conformer centred at $[\beta_{apo} + 10H]^{10+}$.
iii) Semi-folded conformer centred at $[\beta_{apo} + 8H]^{8+}$.

The presence of folded conformers at low charge states, and the absence of dimeric species, confirms that inter-subunit forces are weaker than the intramolecular forces within the globin chains.

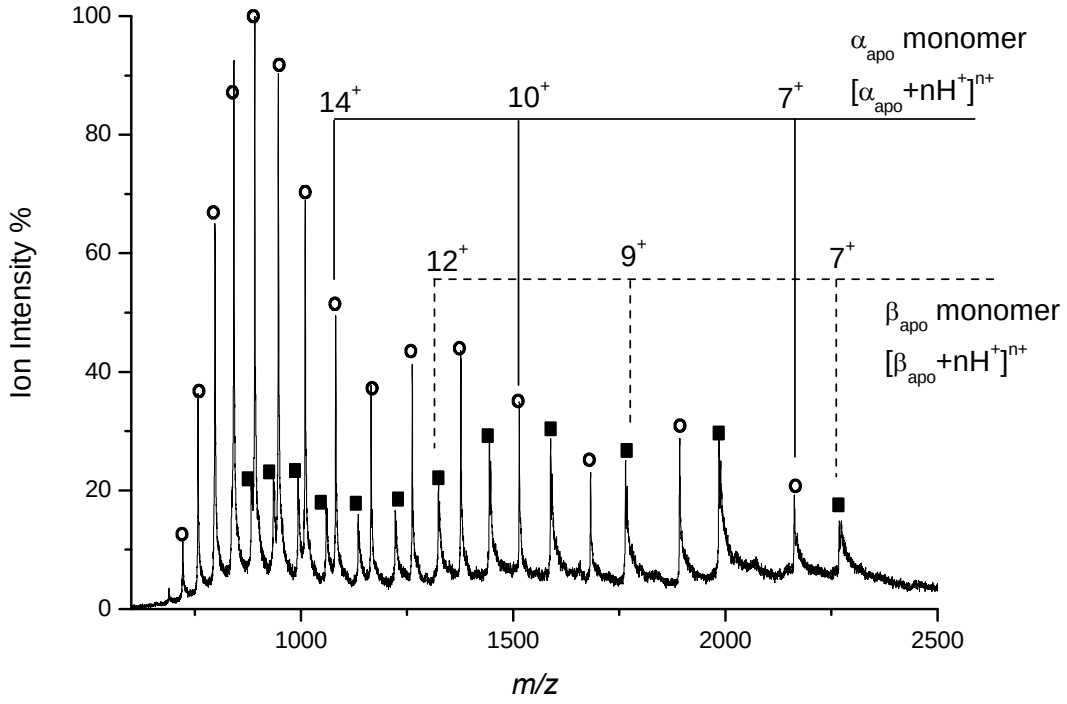



*Figure 4.3. Mass spectrum over the range 500 to 2500 m/z of haemoglobin at pH 4.5 in 2% by volume formic acid. $\alpha_{apo}$ peaks are denoted with unfilled circles; $\beta_{apo}$ peaks with filled squares.*

### 4.4.3 Haemoglobin Studied by Mass Spectrometry at pH 9.5

Dropwise addition of ammonium hydroxide to raise the pH of the protein solution to pH 9.5 promoted tetramer signal, consistent with previous work.[28] Figure 4.4 shows the accompanying mass spectrum with three tetramer (denoted "Q" for quaternary structure) signals of the form, $[Q + nH]^{z+}$, with z = $16^+$, $17^+$ and $18^+$. The mass spectrum also contains one dimer ($\alpha_{holo}\beta_{holo}$; z = $11^+$ and $12^+$) and one monomer ($\alpha_{holo}$; z = $6^+$ to $8^+$) species, both of which have likely dissociated from the parent tetramer molecule in the gas phase due to their low charge density and in retaining the haem ligand.

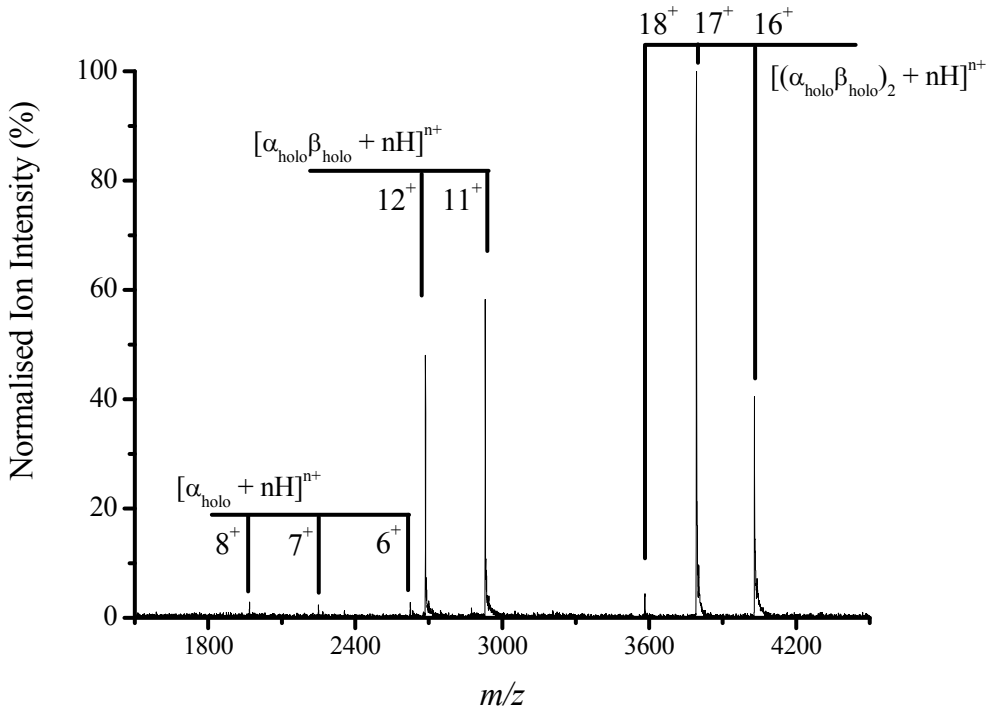



*Figure 4.4. Mass spectrum of haemoglobin in ammonium acetate at pH 9.5. The elevated pH stabilises tetramer signal (z = 17+ dominating spectrum; tetramer denoted $(\alpha^*\beta^*)_2$ in spectrum). $\alpha_{holo}\beta_{holo}$ dimer (denoted $\alpha^*\beta^*$ in spectrum) and $\alpha_{holo}$ ($\alpha^*$ in spectrum) also present.*

Absence of $\beta_{holo}$ is perhaps unexpected, but previous work[17, 28] using commercially available lyophilised haemoglobin (as here) reports the same and attributes it to the α monomer ionising preferentially over the β monomer because of its greater non-polar character, thereby competing more effectively for charge.[32] Verluis and Heck reported[28] the pI of the $\alpha_{apo}$ monomer to be between 8 and 9, and the pI of the $\beta_{apo}$ monomer between 6.5 and 7. This suggests that the α monomer could more readily accommodate positive charge.

### 4.4.4 Native-Like Conditions Studied by Ion Mobility-Mass Spectrometry

For all species analysed by ion mobility-mass spectrometry under native-like conditions, a single arrival time peak was observed, attributed to a compact structure. This may not indicate a singular gas-phase conformation but perhaps conformations with similar mobilities and collision cross-sections. Four $\alpha_{holo}$ monomer species were distinguished and are displayed in Table 4.2.

*Table 4.2. Cross-sections for haemoglobin $\alpha_{holo}$ monomer under native-like conditions. Values are an average of three replicates and stated with twice the standard deviation of the mean ($2\sigma$).*

| Charge state of $\alpha_{holo}$ (native-like) | Collision cross-section (Å$^2$) |
|---|---|
| 6 | 1114.9 ± 5.6 |
| 7 | 1420.4 ± 4.5 |
| 8 | 1525.3 ± 6.7 |
| 9 | 1608.3 ± 7.1 |

Collision cross-sections for $\alpha_{holo}$ species are observed to increase with increasing charge with a near linear increase from $z = 7^+$ to $9^+$ which is attributed to Coulombic repulsion. A larger increase in cross-section (21.5%) is observed between the initial charge states from $z = 6^+$ to $7^+$ which may indicate a protein unfolding event where a region of the monomer structure has changed quite significantly. This unfolding event has not, however, affected the region stabilising the haem ligand.

Ten $\beta_{apo}$ monomer cross-sections were determined with charge states $z = 8^+$ to $17^+$ and are reported in Table 4.3. Once again, there is a near-linear increase in cross-section at low charge densities ($z = 8^+$ to $11^+$) which results in a total increase of 410 Å$^2$ due to protein unfolding to accommodate the additional positive charge. Within the charge region $z = 11^+$ to $14^+$, the cross-section total increase is less than 75 Å$^2$, indicating positive charge is being sequestered at sites potentially more remote, and a large conformational change is not necessary to reduce the repulsive force. As further charge is acquired, the protein undergoes a conformational change accompanied by a large increase in cross-section with an increase in 596 Å$^2$ from $z = 14^+$ to $16^+$. This could be the loss of a secondary fold in a significant portion of the monomeric protein.

*Table 4.3. Haemoglobin $\beta_{apo}$ collision cross-sections under native-like conditions. Values are an average of three replicates and stated with twice the standard deviation of the mean ($2\sigma$).*

| $\beta_{apo}$ (native-like) charge state | Collision cross-section (Å$^2$) |
|---|---|
| 8 | 1461.2 ± 6.3 |
| 9 | 1626.2 ± 3.1 |
| 10 | 1733.3 ± 4.6 |
| 11 | 1871.2 ± 8.6 |
| 12 | 1886.1 ± 3.2 |
| 13 | 1898.9 ± 5.2 |
| 14 | 1945.1 ± 1.4 |
| 15 | 2079.9 ± 4.1 |
| 16 | 2541.3 ± 5.1 |
| 17 | 2622.4 ± 2.3 |

These larger cross-sections support the suggestion[17] that the β monomer has an intrinsically unstructured nature. Small regions of the protein are potentially structured but protonation causes large structural changes. This may indicate few conserved interactions within the molecule.

Under physiological conditions, $\beta_{apo}$ monomers utilise $\alpha_{holo}$ monomers as scaffolding templates to order themselves into a structure capable of binding a haem ligand. This initially forms the $\alpha_{holo}\beta_{apo}$ pseudo-holo heterodimer. This species has been proposed as a crucial intermediate of the tetramer assembly pathway. Three of these species were characterised by ion mobility-mass spectrometry with charge states z = $10^+$ to $12^+$. Upon formation of this dimer, the $\beta_{apo}$ chain is then able to adopt the conformation necessary to bind a haem ligand to form the $\alpha_{holo}\beta_{holo}$ heterodimer. The same three charge states (z = $10^+$ to $12^+$) were observed and characterised, producing cross-sections for both dimer species. All values are presented in Table 4.4.

*Table 4.4. Collision cross-sections of haemoglobin $\alpha_{holo}\beta_{apo}$ and $\alpha_{holo}\beta_{holo}$ dimers. Values are an average of three replicates & stated with twice the standard deviation of the mean ($2\sigma$).*

| Dimer charge state | $\alpha_{holo}\beta_{apo}$ dimer collision cross-section (Å$^2$) | $\alpha_{holo}\beta_{holo}$ dimer collision cross-section (Å$^2$) |
|---|---|---|
| 10 | 2225.4 ± 10.1 | 2173.8 ± 9.4 |
| 11 | 2249.2 ± 8.2 | 2355.2 ± 8.6 |
| 12 | 2468.8 ± 8.7 | 2483.9 ± 8.8 |

Collision cross-sections for all dimers imply minimal protein unfolding with increasing protonation, and are within 5% of one another at the same charge state *i.e.* the z = $10^+$ $\alpha_{holo}\beta_{holo}$ dimer cross-section

is 2.3% smaller than the $\alpha_{holo}\beta_{apo}$ dimer; the $z = 11^+$ $\alpha_{holo}\beta_{apo}$ dimer cross-section is 4.5% smaller than the $\alpha_{holo}\beta_{holo}$ dimer; and the $z = 12^+$ $\alpha_{holo}\beta_{apo}$ dimer cross-section is 0.6% smaller than the $\alpha_{holo}\beta_{holo}$ dimer. This indicates that for these gas-phase conformations, haem ligand loss through structural rearrangement and loss of non-covalent bonding has altered collision cross-sections to a small extent only, and that the dimer structure contains strong inter-subunit interactions.

The native-like conditions used here produced signal for two tetramer species with $z = 16^+$ and $17^+$. This narrow charge state distribution indicates well-folded gas-phase tetramer conformers and sustained preservation of non-covalent interactions. Collision cross-sections of 3394.2 $Å^2$ (± 11.4 $Å^2$) and 3520.4 $Å^2$ (± 10.6 $Å^2$) were calculated for the $z = 16^+$ and $17^+$ tetramer species respectively. This small (3.6%) increase in cross-section suggests charge sequestering at sites that do not induce a large global conformational change to overcome unfavourable interactions. Higher charge states were not observed.

### 4.4.5 Haemoglobin Studied by by Ion Mobility-Mass Spectrometry at pH 4.5

Monomer globin chains of both $\alpha_{apo}$ and $\beta_{apo}$ were observed in the low pH ion mobility-mass spectrum over a wide range of charge states. Sixteen $\alpha_{apo}$ cross-sections and thirteen $\beta_{apo}$ cross-sections were calculated and are presented in Table 4.5. At low charge states ($z \leq 12^+$), $\alpha_{apo}$ cross-sections are more compact than $\beta_{apo}$ cross-sections indicating a more folded structure. This could be due to a lack of solvent-accessible protonation sites as they are involved in intramolecular bonding and/or are within the protein core. Cross-sections increase in a near-linear fashion with charge state due to unfolding to overcome Coulombic repulsion. At higher charge states, it is apparent that either a more complicated unfolding mechanism is occurring or that the unfolding is not attributed to Coulombic repulsion. Possible exposure of a basic residue-rich section would produce a significant increase in cross-sections with additional charge.

*Table 4.5. Collision cross-sections of haemoglobin $\alpha_{apo}$ and $\beta_{apo}$ monomers produced at pH 4.5 with acid. Values are an average of three replicates & stated with twice the standard deviation of the mean ($2\sigma$). Species not observed marked n.o.*

| $\alpha_{apo}$ collision cross-section (Å$^2$) | Monomer charge state | $\beta_{apo}$ collision cross-section (Å$^2$) |
|---|---|---|
| 1381.2 ± 3.4 | 7 | 1421.2 ± 3.2 |
| 1514.8 ± 4.5 | 8 | 1532.1 ± 3.1 |
| 1625.3 ± 5.8 | 9 | 1678.9 ± 5.3 |
| 1699.6 ± 1.2 | 10 | 1864.3 ± 6.3 |
| 1863.3 ± 8.6 | 11 | 1941.4 ± 4.3 |
| 2040.7 ± 7.6 | 12 | 2067.6 ± 8.4 |
| 2252.0 ± 7.5 | 13 | 2155.9 ± 8.4 |
| 2441.4 ± 8.3 | 14 | 2267.2 ± 9.6 |
| 2590.1 ± 2.1 | 15 | 2398.4 ± 9.6 |
| 2614.1 ± 2.4 | 16 | 2595.2 ± 10.4 |
| 2772.2 ± 3.4 | 17 | 2679.1 ± 2.4 |
| 3200.0 ± 3.9 | 18 | 3057.3 ± 2.6 |
| 3227.4 ± 5.3 | 19 | 3146.8 ± 5.4 |
| 3429.9 ± 6.4 | 20 | n.o. |
| 3465.4 ± 8.7 | 21 | n.o. |
| 3498.2 ± 9.8 | 22 | n.o. |

At charge states greater than z = 12$^+$, $\alpha_{apo}$ monomer cross-sections are less compact than equivalently-charged $\beta_{apo}$ monomers, indicating that the $\alpha_{apo}$ monomer has lost its structural integrity and begun to unfold to a greater degree. Figure 4.5 is a plot showing $\alpha_{apo}$ and $\beta_{apo}$ cross-sections over a range of charge states under denaturing conditions with acid. At these higher charge states, cross-sections for both species are similar and show loss of structure.

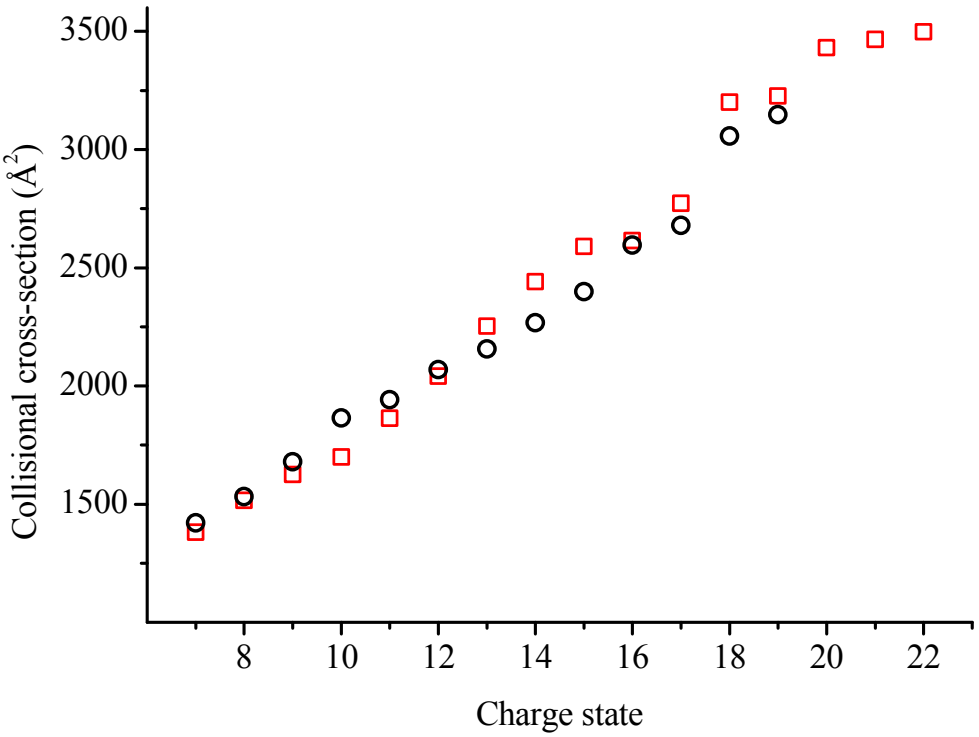



*Figure 4.5. Plot of haemoglobin $\alpha_{apo}$ (red squares) and $\beta_{apo}$ (black circles) cross-sections as a function of charge state under denaturing conditions with acid.*

From the plot, it can be more easily seen that the two monomers have cross-sections which mirror one another closely in the absence of the haem ligand. There is a slight "jump" in cross-section between z = $17^+$ and $18^+$ perhaps due to a larger unfolding event. Charge states for z = $20^+$ to $22^+$ were not observed for $\beta_{apo}$.

### 4.4.6 Elevated pH Studies of Tetrameric Haemoglobin by Ion Mobility-Mass Spectrometry

Increased pH (9.5) strengthened tetramer signal and allowed the complex to be interrogated by ion mobility-mass spectrometry. Five tetramer species could be elucidated. The z = $18^+$ species obtained under traditional mass spectrometry conditions (shown in Figure 4.4) could not be reproduced for ion mobility investigation; however three new species (z = $13^+$, $14^+$ and $15^+$) were produced perhaps due to slight variation in source tuning and pressure parameters, leading to increased desolvation. These tetrameric species were successfully transferred to the gas phase by reducing the source region hexapole pressure to $5 \times 10^{-1}$ mbar (and increasing solution pH). Collision cross-sections are displayed in Table 4.6.

*Table 4.6. Collision cross-sections of haemoglobin tetramers produced under elevated solution pH and decreased source hexapole pressure conditions. Values are an average of two replicates and stated with twice the standard deviation of the mean (2σ).*

| Tetramer charge state | Collision cross-section ($Å^2$) |
|---|---|
| 13 | 3051.2 ± 12.6 |
| 14 | 3215.4 ± 14.6 |
| 15 | 3407.9 ± 10.5 |
| 16 | 3460.3 ± 16.5 |
| 17 | 3649.0 ± 11.1 |

Collision cross-section increases with increasing charge, however there is only an increase of 16.4% in cross-section from $z = 13^+$ to $17^+$. This is a relatively small increase as would be expected for a large molecular complex that would have a number of solvent-accessible protonation sites present to sequester proton charge at. This increase does imply a small coulombically-driven change in conformation.

## 4.5 MOBCAL-Derived Collision Cross-Sections

Structures of condensed-phase haemoglobin have been obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank[33] (PDB). Collision cross-sections have been calculated using the exact hard-spheres scattering (EHSS) method by MOBCAL.[34, 35] Table 4.7 shows the haemoglobin species, the PDB ID file used to obtain the cross-section, the crystal structure resolution and the cross-section produced.

*Table 4.7. Collision cross-sections calculated in silico using the Exact Hard-spheres Scattering method by MOBCAL for five haemoglobin species.*

| Haemoglobin species | PDB code | Resolution (Å) | Collision cross-section ($Å^2$) |
|---|---|---|---|
| Tetramer (oxy) | 1GZX | 2.10 | 4272 |
| Tetramer (deoxy) | 2DN2 | 1.25 | 4207 |
| $\alpha_{holo}\beta_{holo}$ dimer | 2DN1 | 1.25 | 2699 |
| $\alpha_{holo}$ monomer | 2DN1 | 1.25 | 1587 |
| $\beta_{holo}$ monomer | 2DN1 | 1.25 | 1661 |

Computationally-derived cross-sections are larger than experimental cross-sections, regardless of charge state, for dimer and tetramer species. This may indicate a shrinking of the protein structure on transfer to a solvent-free environment for analysis. The calculated cross-section for the $\alpha_{holo}$ monomer lies between the experimental values of the $z = 8^+$ and $9^+$ charge states, as shown in Figure 4.6.

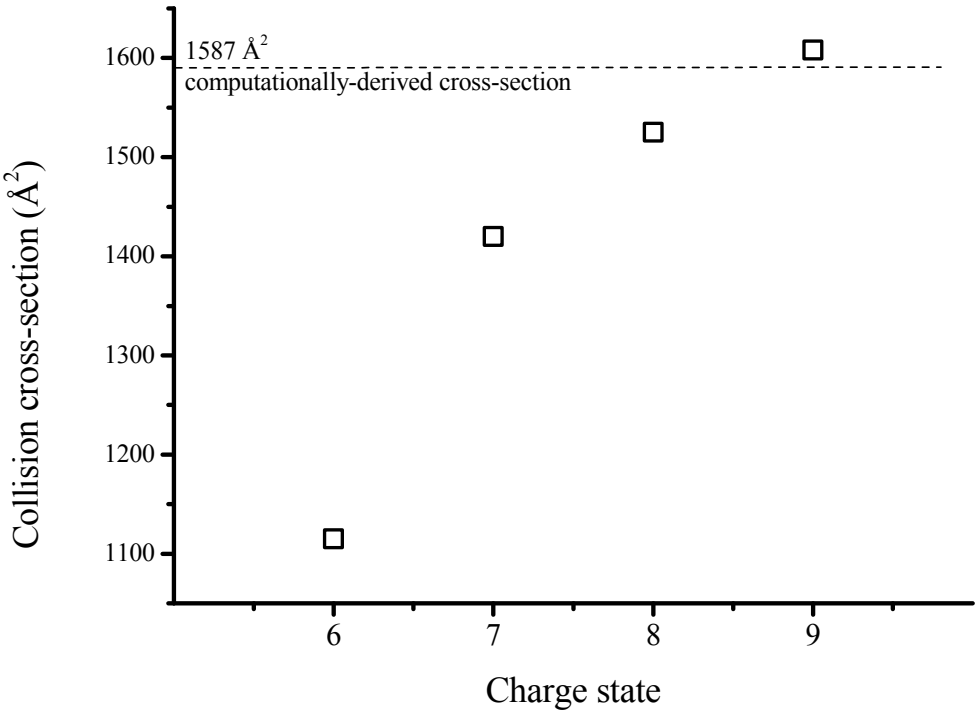



*Figure 4.6. Plot of $\alpha_{holo}$ collision cross-section as a function of charge state showing the computationally-derived cross-section determined using MOBCAL.*

Visual Molecular Dynamics (VMD) software[36] was used to create the structures of tetrameric and dimer haemoglobin from their x-ray crystal structure coordinates. Tetrameric haemoglobin is composed of a heterodimer made up of an alpha subunit and a beta subunit. This is shown in Figure 4.7.

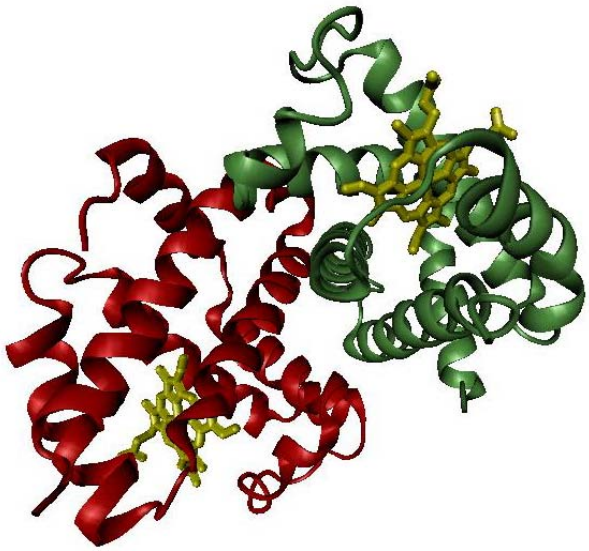


*Figure 4.7. Haemoglobin heterodimer ($\alpha_{holo}\beta_{holo}$) showing the alpha subunit (green), the beta subunit (red) and the haem ligand (yellow). PDB ID: 2DN1.*[37]

## 4.6 Haemoglobin Conclusions

Intact haemoglobin tetramer has been successfully transferred to the gas phase and characterised by mass spectrometry and ion mobility-mass spectrometry. Collision cross-sections have been calculated for a wide range of monomeric charge states with cross-sections for the $\alpha_{holo}$ and $\alpha_{apo}$ obtained by spraying the protein from native-like and denaturing conditions respectively. From the same solvent conditions, only $\beta_{apo}$ monomer could be produced indicating the greater conformational flexibility of this molecule. This has been attributed to its *in vivo* activity and the need for the $\alpha_{holo}$ to be present to act as a template to allow $\beta_{apo}$ to form a stable ordered structure so as to bind haem correctly.

A narrow charge state distribution of structured compact heterodimers containing one and two haem ligands was observed and cross-sections elucidated, showing that upon loss of one haem ligand, there is little significant loss in tertiary structure. Solution pH was increased above neutral to acquire cross-sections of the intact tetramer over five charge states. Cross-sections compared with those calculated from condensed-phase structures are smaller, indicating significant shrinkage in the gas phase.

The work presented herein may elucidate information into haemoglobin dissociation, more so than association, due to the experimental conditions and the gas-phase species observed.

Comparison of cytochrome *c* dimers (Chapter 3) with haemoglobin dimers indicates the importance for intermolecular interactions that occur *in vivo*. Cytochrome *c* dimers are not observed *in vivo* and the presence of gas-phase conformers over a wide charge states distribution ($z = 11^+$ to $19^+$) shows protein unfolding before dissociation. Haemoglobin dimers, crucial in haemoglobin formation, display a narrow charge state distribution ($z = 10^+$ to $12^+$) of lower charge density due to compact and structured folding. Higher charge states are not observed under native-like conditions, and few monomer species are detected. Cytochrome *c* monomers, however, are abundant showing dissociation from the non-specific aggregation molecule. These experiments show the relevance for studying biological molecules in the gas phase as information on their intact structure and the dissociated subunits can be obtained.

## 4.7 Chapter 4. References


1. Paoli, M.; Liddington, R.; Tame, J.; Wilkinson, A.; Dodson, G., Crystal Structure of T State Haemoglobin with Oxygen Bound At All Four Haems. *J. Mol. Biol.* **1996,** 256, (4), 775-792.

2. Riggs, A. F., Self-association, cooperativity and supercooperativity of oxygen binding by hemoglobins. *J. Exp. Biol.* **1998,** 201, (8), 1073-1084.

3. Voet, D.; Voet, J. G., *Biochemistry*. Second ed.; John Wiley & Sons, Inc.: Chichester, U.K., 1995.

4. Linberg, R.; Conover, C. D.; Shum, K. L., Hemoglobin Based Oxygen Carriers: How Much Methemoglobin is too Much? *Artif. Cells Blood Substit. Immobil. Biotechnol.* **1998,** 26, (2), 133 - 148.

5. Perutz, M. F., Nature of Haem-Haem Interaction. *Nature* **1972,** 237, (5357), 495-499.

6. Chen, Y. H.; Campbell, J. M.; Collings, B. A.; Konermann, L.; Douglas, D. J., Stability of a highly charged noncovalent complex in the gas phase: holomyoglobin. *Rapid Commun. Mass Spectrom.* **1998,** 12, (15), 1003-1010.

7. Covey, T.; Douglas, D. J., Collision cross sections for protein ions. *J. Am. Soc. Mass Spectrom.* **1993,** 4, (8), 616-623.

8. Feng, R.; Konishi, Y., Stepwise refolding of acid-denatured myoglobin: evidence from electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1993,** 4, (8), 638-645.

9. Konishi, Y.; Feng, R., Conformational Stability of Heme Proteins in vacuo. *Biochem.* **1994,** 33, (32), 9706-9711.

10. Katta, V.; Chait, B. T., Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Am. Chem. Soc.* **1991,** 113, (22), 8534-8535.

11. Chen, Y.-L.; Collings, B. A.; Douglas, D. J., Collision cross sections of myoglobin and cytochrome c ions with Ne, Ar, and Kr. *J. Am. Soc. Mass Spectrom.* **1997,** 8, (7), 681-687.

12. Gross, D. S.; Zhao, Y.; Williams, E. R., Dissociation of heme-globin complexes by blackbody infrared radiative dissociation: Molecular specificity in the gas phase? *J. Am. Soc. Mass Spectrom.* **1997,** 8, (5), 519-524.

13. Lee, V. W. S.; Chen, Y.-L.; Konermann, L., Reconstitution of Acid-Denatured Holomyoglobin Studied by Time-Resolved Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1999,** 71, (19), 4154-4159.

14. Hunter, C. L.; Mauk, A. G.; Douglas, D. J., Dissociation of Heme from Myoglobin and Cytochrome b5: Comparison of Behavior in Solution and the Gas Phase. *Biochem.* **1997,** 36, (5), 1018-1025.

15. Sogbein, O. O.; Simmons, D. A.; Konermann, L., Effects of pH on the kinetic reaction mechanism of myoglobin unfolding studied by time-resolved electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2000,** 11, (4), 312-319.

16. Schmidt, A.; Karas, M., The influence of electrostatic interactions on the detection of heme-globin complexes in ESI-MS. *J. Am. Soc. Mass Spectrom.* **2001,** 12, (10), 1092-1098.

17. Griffith, W. P.; Kaltashov, I. A., Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry. *Biochem.* **2003,** 42, (33), 10024-33.

18. Fronticelli, C.; Bucci, E., Conformational and functional characteristics of bovine hemoglobin. In *Methods Enzymol.*, Academic Press: 1994; Vol. 231, pp 150-163.

19. Dobo, A.; Kaltashov, I. A., Detection of Multiple Protein Conformational Ensembles in Solution via Deconvolution of Charge-State Distributions in ESI MS. *Anal. Chem.* **2001,** 73, (20), 4763-4773.

20. Woody, R. W.; Hsu, M., Origin of the heme Cotton effects in myoglobin and hemoglobin. *J. Am. Chem. Soc.* **1971,** 93, (14), 3515-3525.

21. Papadopoulos, P. G.; Walter, S. A.; Li, J.; Baker, G. M., Proton interactions in the resting form of cytochrome oxidase. *Biochem.* **1991,** 30, (3), 840-850.

22. Vasudevan, G.; McDonald, M. J., Spectral Demonstration of Semihemoglobin Formation during CN-Hemin Incorporation into Human Apohemoglobins. *J. Biol. Chem.* **1997,** 272, (1), 517-524.

23. Vasudevan, G.; McDonald, M. J., Ordered Heme Binding Ensures the Assembly of Fully Functional Hemoglobin: A Hypothesis. *Curr. Protein Pept. Sci.* **2002,** 3, (4), 461-466.

24. Scarff, C. A.; Patel, V. J.; Thalassinos, K.; Scrivens, J. H., Probing Hemoglobin Structure by Means of Traveling-Wave Ion Mobility Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (4), 625-631.

25. Wright, P. J.; Zhang, J.; Douglas, D. J., Conformations of Gas-Phase Ions of Ubiquitin, Cytochrome c, Apomyoglobin, and beta-Lactoglobulin Produced from Two Different Solution Conformations. *J. Am. Soc. Mass Spectrom.* **2008**.

26. Wright, P. J.; Douglas, D. J., Gas-Phase H/D Exchange and Collision Cross Sections of Hemoglobin Monomers, Dimers, and Tetramers. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (3), 484-495.

27. Douglas, D. J., An aerodynamic drag model for protein ions. *J. Am. Soc. Mass Spectrom.* **1994,** 5, (1), 17-18.

28. Versluis, C.; Heck, A. J. R., Gas-phase Dissociation of Hemoglobin. *Int. J. Mass Spectrom.* **2001,** 210-211, 637-649.

29. Light-Wahl, K. J.; Schwartz, B. L.; Smith, R. D., Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **1994,** 116, (12), 5271-5278.

30. Li, Y. T.; Hsieh, Y. L.; Henion, J. D.; Ganem, B., Studies on heme binding in myoglobin, hemoglobin, and cytochrome c by ion spray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1993,** 4, (8), 631-637.

31. Babu, K. R.; Douglas, D. J., Methanol-induced conformations of myoglobin at pH 4.0. *Biochem.* **2000,** 39, (47), 14702-10.

32. Kuprowski, M. C.; Boys, B. L.; Konermann, L., Analysis of protein mixtures by electrospray mass spectrometry: effects of conformation and desolvation behavior on the signal intensities of hemoglobin subunits. *J. Am. Soc. Mass Spectrom.* **2007,** 18, (7), 1279-85.

33. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucl. Acids Res.* **2000,** 28, (1), 235-242.

34. Shvartsburg, A. A.; Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996,** 261, (1-2), 86-91.

35. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996,** 100, (40), 16082-16086.

36. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Molec. Graphics* **1996,** 14, (1), 33-38.

37. Park, S.-Y.; Yokoyama, T.; Shibayama, N.; Shiro, Y.; Tame, J. R. H., 1.25 Å Resolution Crystal Structures of Human Haemoglobin in the Oxy, Deoxy and Carbonmonoxy Forms. *J. Mol. Biol.* **2006,** 360, (3), 690-701.

# Chapter 5. Mass Spectrometry & Ion Mobility-Mass Spectrometry Studies of p53 & Anterior Gradient-2 Proteins

*Chapter 5 describes work performed on two proteins, p53 and anterior gradient-2, involved in important cellular activities. Both are linked in vivo with p53 being a tumour suppressor protein and AGR2, a p53 inhibitor. Disruption of the interaction between p53 and AGR2 using peptide aptamers may lead to p53 upregulation and tumour suppression. Mass spectrometry and ion mobility-mass spectrometry characterisation of the DNA-binding core domain of p53 is presented in the presence and absence of a coordinated zinc ion. Characterisation of multimeric AGR2 protein and each aptamer is presented, with data for a complex between one aptamer and AGR2 also shown.*

## 5.1 p53 Protein Introduction

As one of the most intensively studied tumour suppressors of the past two decades,[1] p53 continues to provide encouraging evidence in the fight against cancer. Nicknamed "Guardian of the Genome",[2] p53 suppresses tumour formation by inducing cell-cycle arrest, apoptosis and senescence upon cellular stress, thereby destroying DNA-damaged cells before they can proliferate,[3] and it is thought that the p53 pathway will continue to be the subject of intense anti-cancer drug developments for the next 25 years.[4] As a highly sequence specific DNA-binding transcription factor, p53 concentration is maintained at low resting levels due to the interaction with the E3 protein ligase MDM2 (mouse double minute 2).[5, 6] MDM2 maintains p53 concentration at low levels *via* an autoinhibitory feedback loop which mediates ubiquination and degradation of p53;[7] however MDM2 is often over-produced in a range of cancers, resulting in attenuation of the p53 response.[8] MDM2-mediated inhibition of p53 protein results in degradation of the p53 protein by the proteasome.[1] Interaction between MDM2 and p53 has been shown to occur at the N-terminal domains of both molecules[9] with three buried amino acid residues of p53 binding in a deep hydrophobic binding cleft on MDM2.[3] p53 binds to DNA through a structurally complex domain, stabilised by a tetrahedral geometry zinc ion ($Zn^{2+}$) bound by three cysteine and one histidine residues.[10] This metal atom plays a regulatory role in the control of p53 folding and DNA-binding activity.

Méplan *et al.*[11] have investigated the role of metal binding on the stability of the p53:DNA complex, revealing zinc chelation disrupts the architecture of the p53 binding domain, resulting in rapid cysteine oxidation and disulphide-linked aggregate formation. Méplan prepared a recombinant form of p53 lacking zinc. Upon addition of zinc at physiological concentrations (~5 μM), renaturation and reactivation of p53 occurred, allowing DNA-binding to occur. They further investigated this with a range of divalent metals and found that $Co^{2+}$ at concentrations of ~125 μM had a similar effect, implying that the metal atom acts to hold the domain in a favourable conformation for DNA-binding to occur. Full length p53 (~43.7 kDa) is composed of 393 amino acids and is divided into three major regions: a central DNA-binding core domain, the C-terminal tetramerisation domain and the N-

terminal transactivation domain.[12, 13] The core domain (Figure 5.1; residues 94 to 312) has been characterised by NMR spectroscopy[14] and it is here that many p53 inactivation mutations occur in nearly 50% of human cancers.[15] Further targeted studies into this core binding domain and its interaction with DNA may provide new possibilities into destroying cancer tumours by drug molecule interactions with p53.

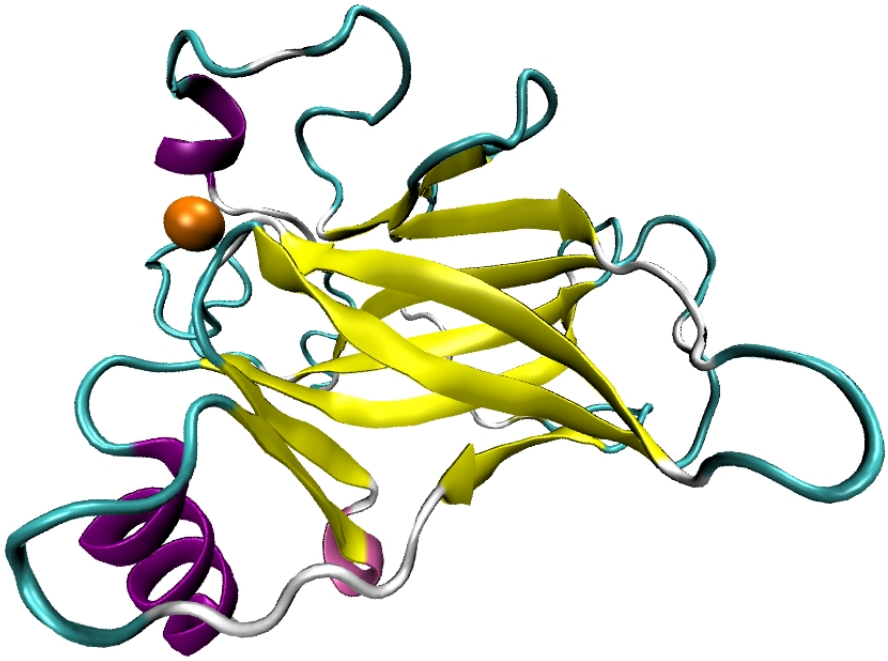


*Figure 5.1 The DNA-binding core domain of p53 (residues 94 to 312; PDB ID 2FEJ;)[14] showing a predominantly sheet-like structure (yellow); alpha-helices (purple); loop regions (green); and the orange coordinated zinc ion.*

## 5.2 p53 Experimental

p53 was sprayed from 20 mM ammonium acetate (pH 6.85) on the MoQTOF instrument, and from ammonium acetate containing 10% by volume isopropanol on the Synapt HDMS instrument to aid desolvation. MoQTOF instrument conditions were: nano-ESI capillary voltage 1.80 kV; source temperature 80$^{o}$C; sample cone 140V; extractor cone 108V; cell pressure ~3.1 Torr helium; cell temperature ~301 K; pulse width 40 µs; and pusher period 120 µs. Synapt HDMS instrument conditions are detailed in Chapter 2, section 10.

## 5.3 p53 Results

p53 protein results will be divided into sections to allow a comparison between the two instruments used with respect to the species observed by MS and collision cross-section by IM-MS.

### 5.3.1 Mass spectrometry of p53 Containing Zinc on the Synapt HDMS Instrument

Figure 5.2 presents a mass spectrum obtained on a Synapt HDMS instrument for p53 sprayed from native-like ammonium acetate solution containing 10% by volume isopropanol. The most abundant peak is the $z = 9^+$ and a bimodal distribution is observed (centres at $z = 9^+$ and $16^+$) with an observed monomer charge state range of $z = 7^+$ to $17^+$, indicative of two solution-phase conformers transferred to the gas phase. Peaks display phosphate adducts at low charge density. The most abundant peaks display low charge density as would be expected for a protein with a native-like compact structure; however higher charge density at lower abundance is evidence for a small population of more unfolded structures possibly attributed to isopropanol addition. Deconvolution of the mass spectrum produces a mass of 24,613.2 Da.

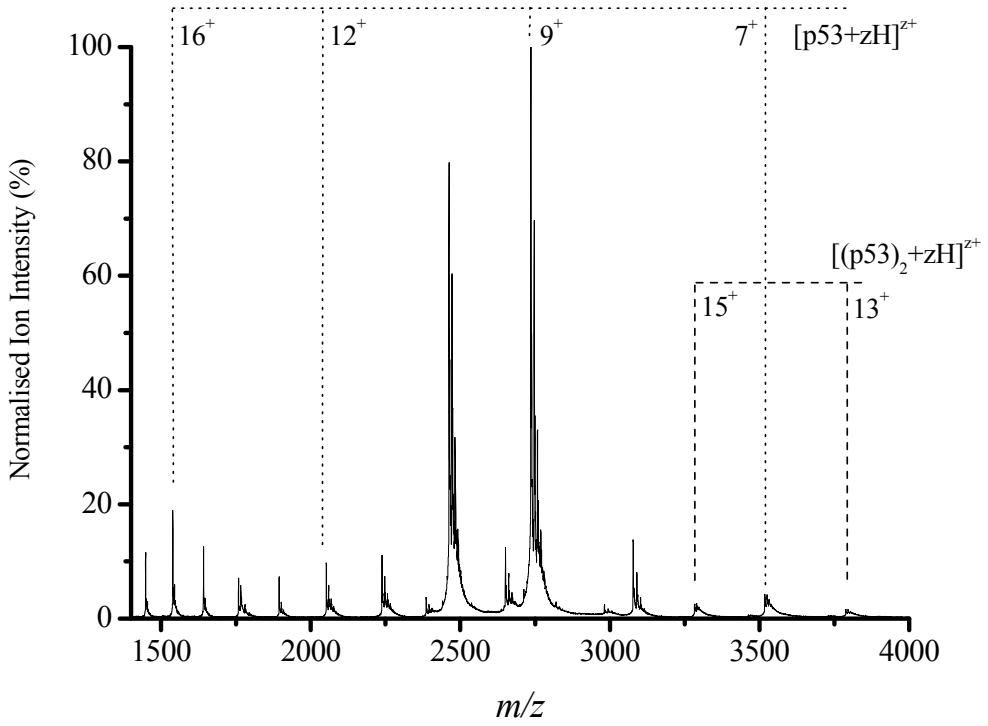



*Figure 5.2. Mass spectrum of p53 containing zinc sprayed from 20 mM ammonium acetate (pH 6.85) and 10% by volume isopropanol on the Synapt HDMS instrument.*

### 5.3.2 Mass spectrometry of p53 Containing Zinc on the MoQTOF Instrument

Figure 5.3 presents a mass spectrum obtained on the MoQTOF instrument for p53 sprayed from native-like buffer solution. Isopropanol was not required to assist in protein desolvation. The most abundant peak is the $z = 10^+$ and no monomer bimodal distribution is observed. Peaks have adducts of phosphate at low charge density. The monomer charge state range of $z = 7^+$ to $17^+$ is observed as with the Synapt mass spectrum, but the intensity of the higher charge density species are significantly

decreased indicating a more native-like protein than that observed on the Synapt. Deconvolution of the mass spectrum produces a mass of 24,617.2 Da.

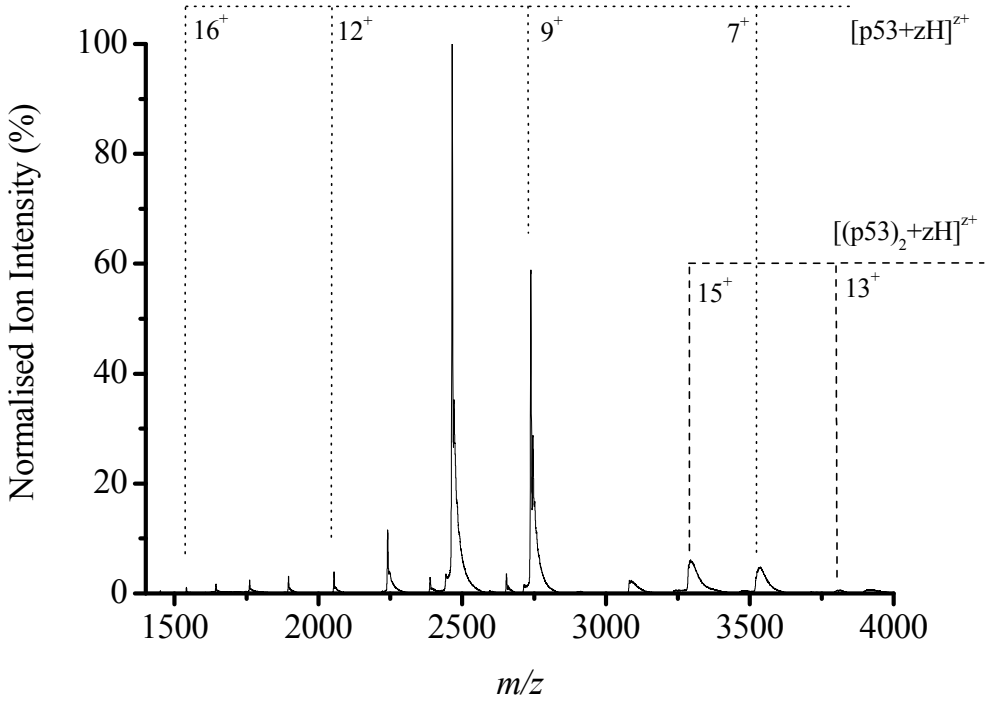



*Figure 5.3. Mass spectrum of p53 containing zinc sprayed from 20 mM ammonium acetate (pH 6.85) on the MoQTOF instrument.*

### 5.3.3 Comparison of Synapt & MoQTOF Mass Spectrometry of p53 Containing Zinc

Mass spectra for native p53 differ slightly with respect to the most abundant peaks between Synapt and MoQTOF data. Charge state distributions are similar, with the base peak varying by a single charge state ($9^+$ in Synapt; $10^+$ in MoQTOF). MoQTOF charge state intensity decreases at higher charge, whereas signal abundance decreases to charge state $13^+$ and increases again to a maximum at $16^+$ in Synapt data, indicating the presence of more denatured protein molecules. This may be attributed to the presence of the isopropanol, however at this concentration, native-like structures still dominate the mass spectrum. Dimer species signal intensity is more pronounced in MoQTOF data, however both instruments have been tuned appropriately so as to transfer higher *m/z* species.

### 5.3.4 Mass Spectrometry of p53 with Zinc Removed on the Synapt HDMS Instrument

p53 with zinc removed was analysed using the Synapt (calculated $M_R$ = 24,550.1; deconvoluted $M_R$ = 24,550.2 Da) and the mass spectrum is shown in Figure 5.4. Data was obtained between over a range of 1500 to 3000 *m/z*, but clearly this should have been reinvestigated over a wider charge state range.

From the data available, it is clear that complete chelation of the zinc was not achieved, however the zinc-free protein dominates at higher charge states. At charge states z = $9^+$ and $10^+$, the zinc-bound protein is dominant, and a number of auxiliary adducts of phosphate are also observed. Protein was dialysed before use to remove salt, however it is apparent that further dialysis was required. Loss of zinc has also contributed to protein unfolding, but the presence of zinc-bound at high charge states indicates that the binding conformation has not been lost.

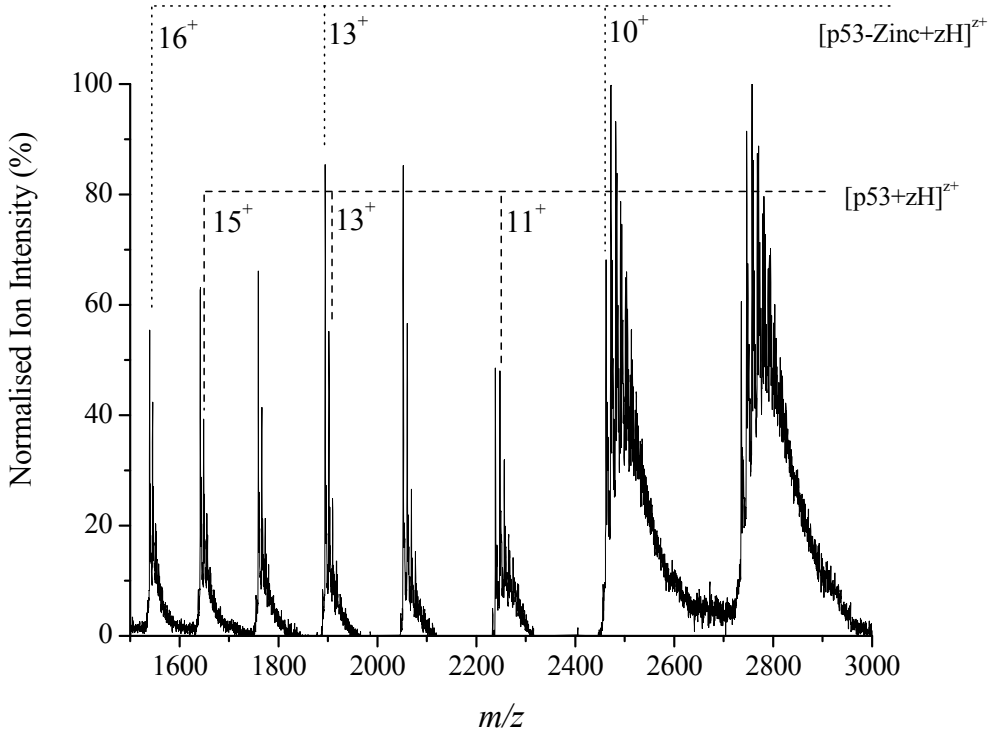



*Figure 5.4. Mass spectrum of p53 with chelating agent (phenanthroline) added showing both p53 without zinc, and p53 still retaining zinc sprayed from 20 mM ammonium acetate (pH 6.85) and 10% by volume isopropanol on the Synapt HDMS instrument.*

### 5.3.5 Mass spectrometry of p53 with Zinc Removed on the MoQTOF Instrument

Figure 5.5 is the mass spectrum of p53 with zinc removed (deconvoluted $M_R$ = 24,550.8 Da) analysed on the MoQTOF displays a narrow charge state distribution of peaks from z = $7^+$ to $13^+$ (3508.2 to 1889.5 *m/z* respectively). Charge states z = $7^+$ to $10^+$ have a second peak at higher *m/*z value that can be assigned to p53 still containing zinc, indicating full chelation of the metal was not achieved either because an insufficient concentration of phenanthroline was added or the incubation period was not long enough (however both parameters were the same in Synapt and MoQTOF experiments). At charge states greater than $10^+$, this peak is not observed indicating the protein can no longer stabilise the metal ion. The dominant charge state of z = $10^+$ indicates p53 has not denatured extensively.

Dimer peaks of $[(p53)_2 + 15H]^{15+}$ (3274.3 *m/*z) and $[(p53)_2 + 13H]^{13+}$ (3777.9 *m/z*; not shown on Figure 5.5 mass range) were also observed.

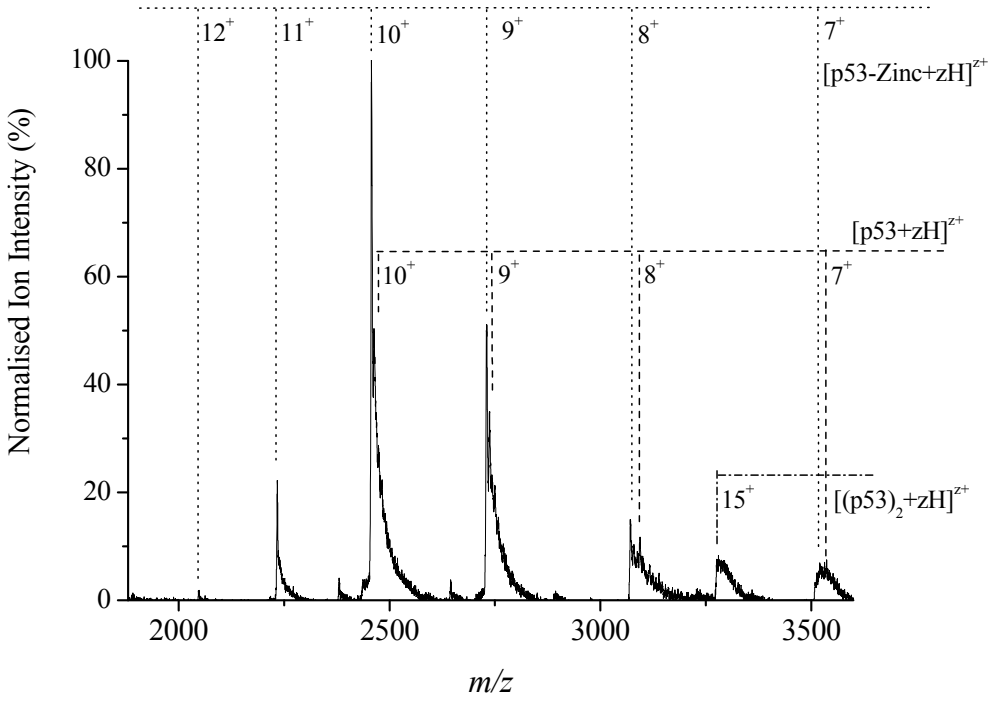



*Figure 5.5. Mass spectrum of p53 with chelating agent (phenanthroline) added showing both p53 without zinc, and p53 still retaining zinc sprayed from 20 mM ammonium acetate on the MoQTOF instrument.*

### 5.3.6 Comparison of Synapt & MoQTOF Mass Spectrometry of p53 with Zinc Removed

MoQTOF data displays a more "native-like" mass spectrum than Synapt when the protein has been subjected to zinc removal using phenanthroline. One would expect that removing a binding site stabilising entity (the zinc ion) would promote formation of a less structured arrangement, as displayed by the Synapt data. However, MoQTOF data shows that within the gas-phase, removal of the zinc has had a less detrimental effect on the overall structure. This is supported by the experimental work by Méplan (as detailed in the Introduction) as removal of zinc produces physiologically inactive p53, but addition renatures the protein. This suggests that the overall structure is not altered (denatured) to such an extent that the zinc-binding region cannot be restored.

### 5.3.7 Ion Mobility-Mass Spectrometry of p53 Containing Zinc on the Synapt HDMS Instrument

A wide charge state range for charge states z = $9^+$ to $16^+$ was obtained for monomeric p53 containing zinc, allowing a total of fifteen collision cross-sections to be determined at a wave height of 8V. These are plotted as a function of charge state in Figure 5.6, with collision cross-section data in Table 5.1. Cross-section is observed to increase with increasing charge and is attributed to protein unfolding upon proton charge sequestering by appropriate solvent accessible sites. These sites are proposed to be basic residues.

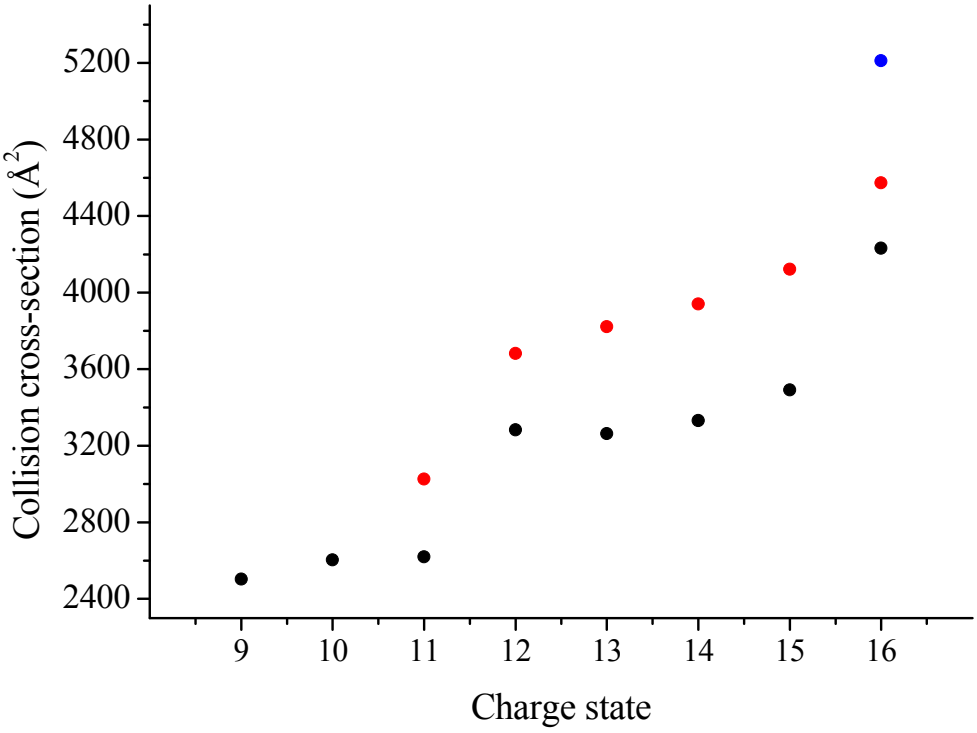



*Figure 5.6. Cross-section plot for p53 containing zinc data obtained on the Synapt HDMS instrument. Five charge states (z = $11^+$ to $15^+$) revealed two gas-phase conformations; one charge state, z = $16^+$, revealed three gas-phase conformations.*

There are 19 arginine (R), 8 lysine (K) and 9 histidine (H) residues in the p53 molecule, as well as the N-terminus, where proton addition could potentially occur. Unfolding occurs to overcome Coulombic repulsion. The lowest two charge states, z = $9^+$ and $10^+$ provided single resolvable conformers (2503Å$^2$ and 2604Å$^2$ respectively). At these low charge states the protein is believed to be in native-like, compact conformations. With increasing charge (z = $11^+$ to $15^+$) two conformers can be elucidated and indicate that multiple gas-phase structures are produced. At higher charge, z = $16^+$, protein collision cross-section continues to increase and three resolvable gas-phase conformations are

produced (4231Å$^2$, 4572Å$^2$ and 5211Å$^2$). p53 has unfolded to a more extended structure and consequently lost a number of the non-covalent interactions (hydrogen bonds and electrostatic interactions) that sustain the protein structure.

*Table 5.1. Collision cross-sections for p53 containing zinc obtained on the Synapt HDMS and MoQTOF instruments. Values are an average of three replicates and stated with twice the standard deviation of the mean (2σ). Species not observed indicated by n.o.*

| Charge state | Synapt p53 containing Zn collision cross-section (Å$^2$) | MoQTOF p53 containing Zn collision cross-section (Å$^2$) |
|---|---|---|
| 9 | 2503 ± 21 | 1689 ± 13<br>1703 ± 14 |
| 10 | 2604 ± 24 | 1920 ± 17<br>2076 ± 23 |
| 11 | 2619 ± 22<br>3026 ± 21 | 1969 ± 25<br>2154 ± 25 |
| 12 | 3282 ± 25<br>3682 ± 28 | 2324 ± 19<br>n.o. |
| 13 | 3263 ± 31<br>3822 ± 33 | 2327 ± 31<br>2577 ± 26 |
| 14 | 3331 ± 24<br>3941 ± 34 | 2748 ± 36<br>2848 ± 29 |
| 15 | 3491 ± 18<br>4121 ± 31 | 3195 ± 31<br>3276 ± 32 |
| 16 | 4231 ± 16<br>4572 ± 15<br>5211 ± 36 | 2820 ± 38<br>3269 ± 31<br>n.o. |
| 17 | n.o. | 3298 ± 41 |

### 5.3.8 Ion Mobility-Mass Spectrometry of p53 Containing Zinc on the MoQTOF Instrument

A slightly larger charge state range was obtained on the MoQTOF ($z = 9^+$ to $17^+$) when p53 was sprayed from 20 mM ammonium acetate (pH 6.85). Unlike Synapt analysis, addition of 10% isopropanol was not required to obtain strong signal. When added, isopropanol was attributed to detrimental signal intensity. Collision cross-sections for p53 are smaller for all charge states when compared with Synapt data, with cross-sections ranging from 1689Å$^2$ for $z = 9^+$ to 3298Å$^2$ for $z = 17^+$. For seven of the nine charge states, two conformations were elucidated (only one conformation for $z = 12^+$ and $17^+$). These are plotted as a function of charge state in Figure 5.7, with collision cross-section data in Table 5.1 (alongside Synapt data). Low charge states adopt compact conformations, with unfolding occurring as charge is increased. MoQTOF conformations of $z = 14^+$ and $15^+$ differ by 3.5 and 2.5% respectively, indicating similar gas-phase structures may be adopted at this charge state. Dimer signal intensity was insufficient under ion-mobility mass spectrometry conditions to calculate cross-sections from MoQTOF data.

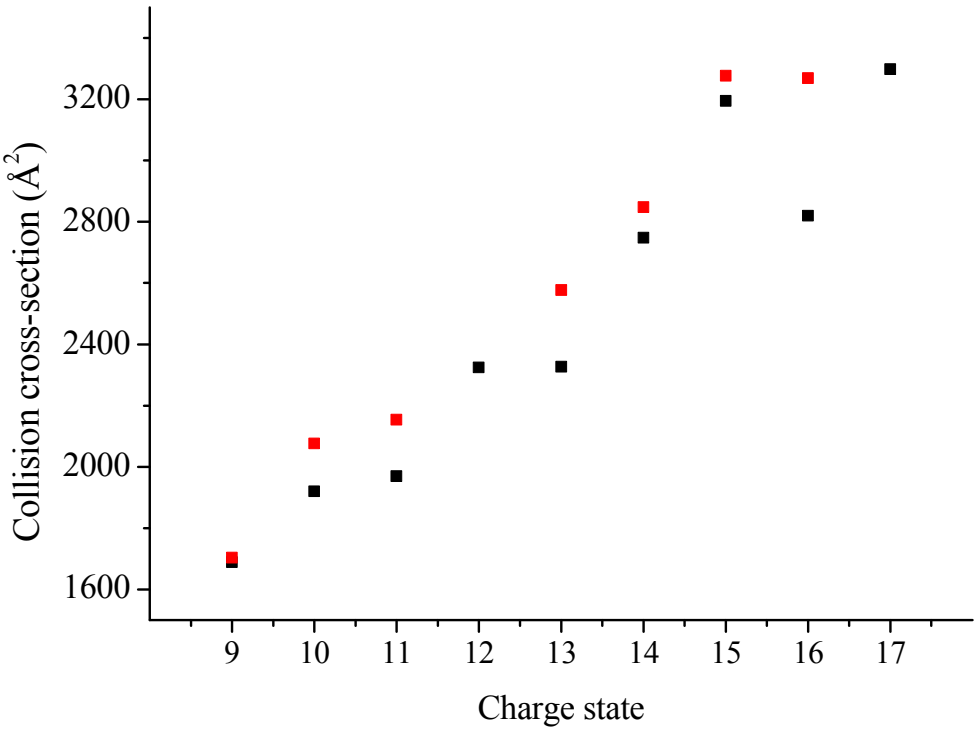



*Figure 5.7. Cross-section plot for p53 containing zinc data obtained on the MoQTOF instrument. Seven charge states (z = $9^+$ - $11^+$ and $13^+$ to $16^+$) revealed two gas-phase conformations.*

An example of the raw p53 MoQTOF experimental data is shown in Figure 5.8 for the $11^+$ charge state conformers. This Figure represents the species at a drift voltage of 35V.

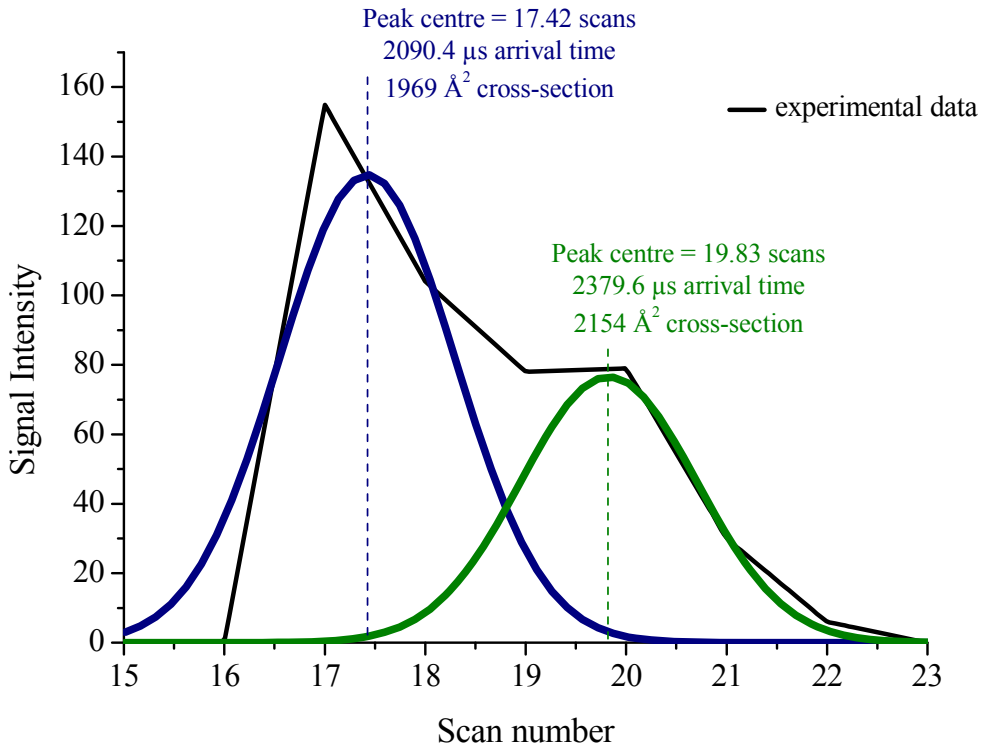



*Figure 5.8. Raw MoQTOF data for the $11^+$ charge state of p53 containing zinc showing two conformers taken at a drift voltage of 35V.*

### 5.3.9 Comparison of Synapt & MoQTOF Ion Mobility-Mass Spectrometry of p53 Containing Zinc

Figure 5.9 is a combined plot of cross-section for p53 containing zinc (Figures 5.6 and 5.7 combined). Variations in cross-section between instrument data can not be directly attributed to biologically different conformations. Compact conformations are observed from both data sets at low charge states, and increase with increasing charge. MoQTOF values are always more than 8.5% smaller than Synapt values for native p53 containing zinc ($z = 15^+$; 3195Å$^2$ MoQTOF vs. 3491Å$^2$ Synapt). This has been attributed predominantly to greater source pressure control in the MoQTOF, allowing better desolvation to occur and a more compact gas-phase structure compared with Synapt sampling. Potentially, dimer molecules of p53 will dissociate in the gas phase (as shown for other gas-phase dimer proteins)[16-19] and contribute to the monomeric species abundance. This could be another potential source of the higher charged monomers observed under mass spectrometry conditions.

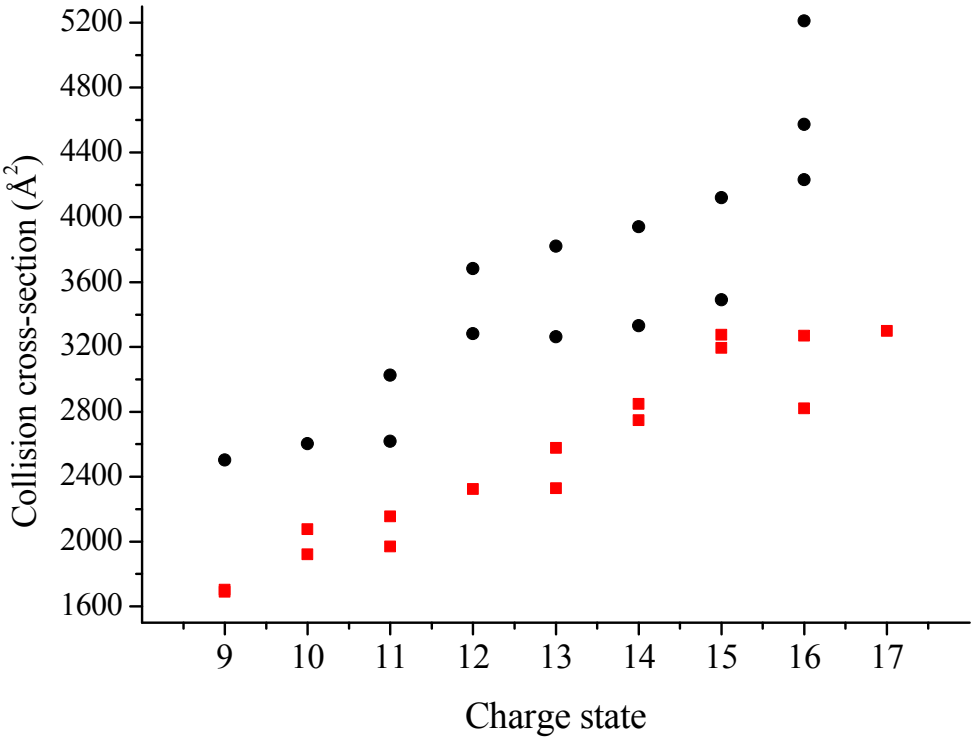



*Figure 5.9. Combined cross-section data over all nine charge states for Synapt HDMS (black circles) and MoQTOF (red squares) of p53 containing zinc. All cross-sections calculated from MoQTOF data are more compact than Synapt values.*

### 5.3.10 Ion Mobility-Mass Spectrometry of p53 with Zinc Removed on the Synapt HDMS Instrument

Collision cross-sections obtained for p53 with zinc removed are in Table 5.2 and plotted as a function of charge state in Figure 5.10. Seven charge states were visible in the Synapt IM-MS spectrum with z = $9^+$ to $15^+$ (wave height 8V). At the two lowest charge states, z = $9^+$ and $10^+$, a single conformation was resolved with cross-sections of 2464Å$^2$ and 2558Å$^2$ respectively.

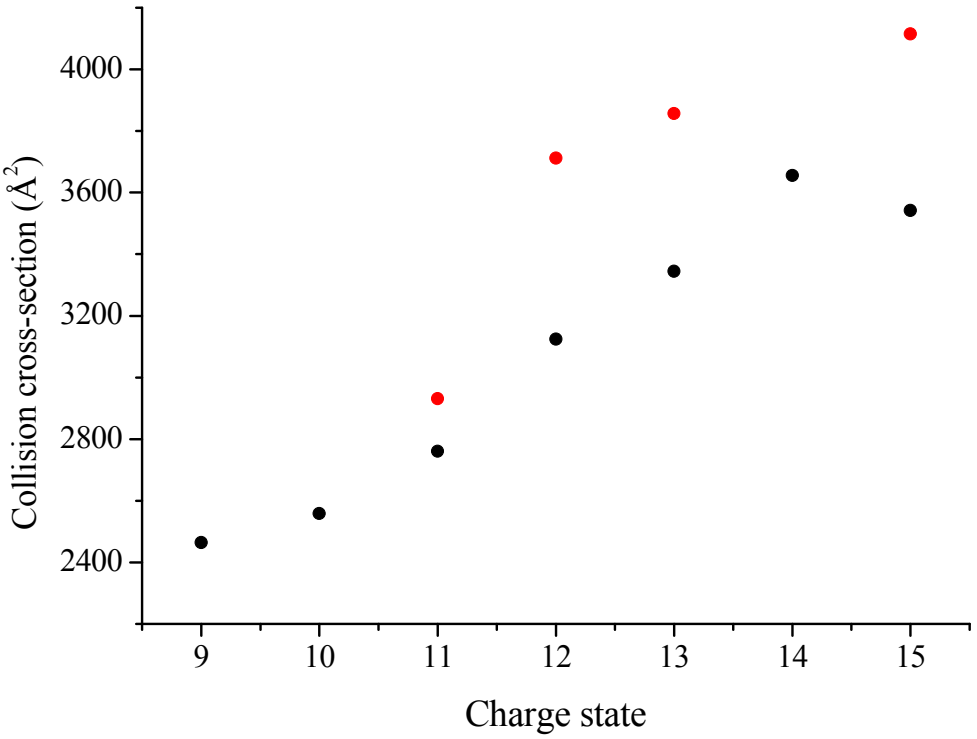



*Figure 5.10. Cross-section plot for p53 with zinc removed data obtained on the Synapt HDMS instrument. Four charge states (z = $11^+$, $12^+$, $13^+$ and $15^+$) revealed two gas-phase conformations.*

At low charge states, structure is most likely of a compact native-like form but loss of the zinc ion may cause structural rearrangement. For z = $11^+$ to $15^+$ (excluding $14^+$), two conformations for each charge state were successfully resolved. In comparison with the data obtained for p53 containing zinc, low charge states (z = $9^+$ and $10^+$) are more compact, having a smaller collision cross-section in the absence of zinc and may have undergone gas-phase collapse. Increasing charge appears to increase cross-section in a near-linear fashion for smaller Synapt conformers due to Coulombic unfolding. Cross-sections for low charge states (z = $9^+$ to $12^+$) are smaller in the absence of zinc than compared with those obtained with zinc present (native p53) by up to 4.8% ($12^+$ smaller conformer). This trend reverses at higher charge state, with larger cross-sections in the absence of zinc above z = $12^+$ compared with the equivalent native p53 charge state due to loss of non-covalent structure.

*Table 5.2. Collision cross-sections for p53 with zinc removed obtained on the Synapt HDMS & MoQTOF instruments. Values are an average of three replicates and stated with twice the standard deviation of the mean (2σ). Species not observed indicated by n.o.*

| Charge state | Synapt p53 with Zn removed collision cross-section (Å$^2$) | MoQTOF p53 with Zn removed collision cross-section (Å$^2$) |
|---|---|---|
| 8 | n.o. | 1450 ± 16<br>1566 ± 19 |
| 9 | 2464 ± 25 | 1963 ± 31 |
| 10 | 2558 ± 27 | 1763 ± 15<br>1820 ± 12 |
| 11 | 2760 ± 24<br>2931 ± 44 | 1953 ± 25<br>2067 ± 37 |
| 12 | 3124 ± 29<br>3712 ± 37 | n.o. |
| 13 | 3344 ± 32<br>3856 ± 30 | 3309 ± 31<br>3614 ± 27 |
| 14 | 3655 ± 27 | n.o. |
| 15 | 3542 ± 41<br>4115 ± 43 | n.o. |

### 5.3.11 Ion Mobility-Mass Spectrometry of p53 with Zinc Removed on the MoQTOF Instrument

MoQTOF cross-sections for five charge states ($z = 8^+$ to $11^+$ and $13^+$; $z = 12^+$ was data was of a poor quality and has been omitted) of p53 with zinc removed are presented in Table 5.2 alongside Synapt values. A cross-section plot is shown in Figure 5.11 and shows that multiple gas-phase conformers were observed for a number of charge states. Cross-sections are all smaller than Synapt values, however at charge state $z = 13^+$, the smaller conformer is only 1.0% smaller, possibly presenting the same gas-phase conformer in both instruments. At low charge states ($z = 8^+$ to $11^+$), cross-sections for zinc removed p53 are smaller than those for the equivalent charge for zinc present native p53 by up to 12.3% for the larger $z = 10^+$ conformer.

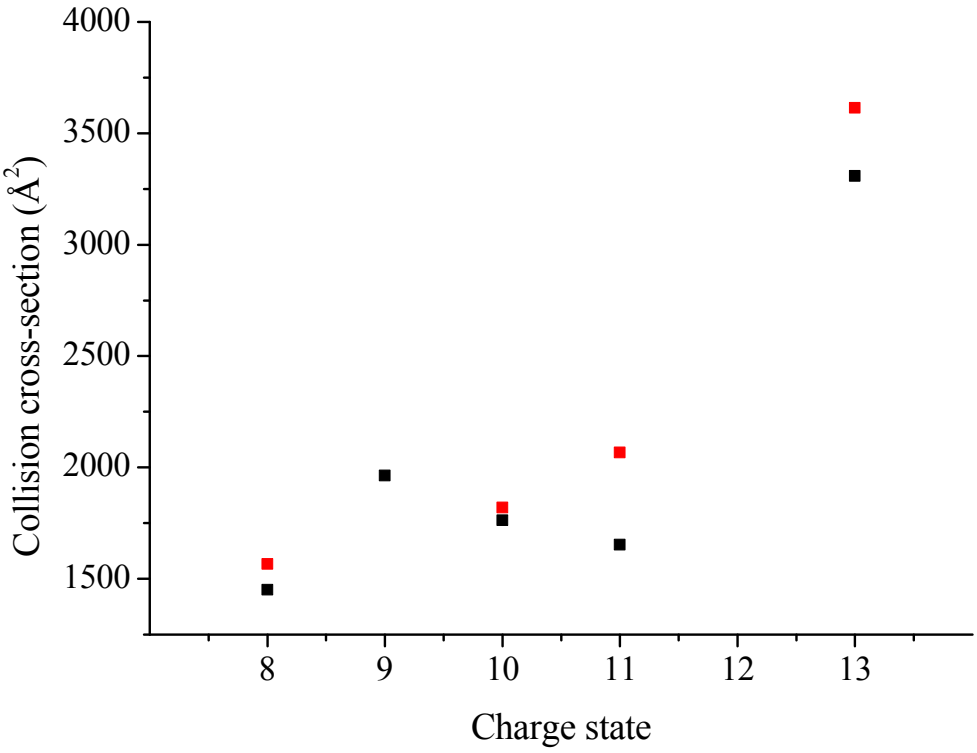



*Figure 5.11. Cross-section plot for p53 with zinc removed data obtained on the MoQTOF instrument. Four charge states (z = 8$^+$, 10$^+$, 11$^+$ and 13$^+$) revealed two gas-phase conformations.*

### 5.3.12 Comparison of Synapt & MoQTOF Ion Mobility-Mass Spectrometry of p53 with Zinc Removed

Synapt data displays a near-linear increase in cross-section with charge to complement unfolding due to Coulombic repulsion. MoQTOF data is again smaller than Synapt, but shows a more sustained cross-section at low charge, with a sudden large unfolding transition from z = 11$^+$ to 13$^+$. This can be interpreted with little insight as z = 12$^+$ data was unavailable. With respect to the larger Synapt data, considerations of the RF field in the T-wave device may also provide larger cross-sections for higher charge species *i.e.* protein molecules with a higher charge will unfold more in the RF potential well. Figure 5.12 combines Synapt and MoQTOF cross-section data for p53 without zinc for ease of comparison.

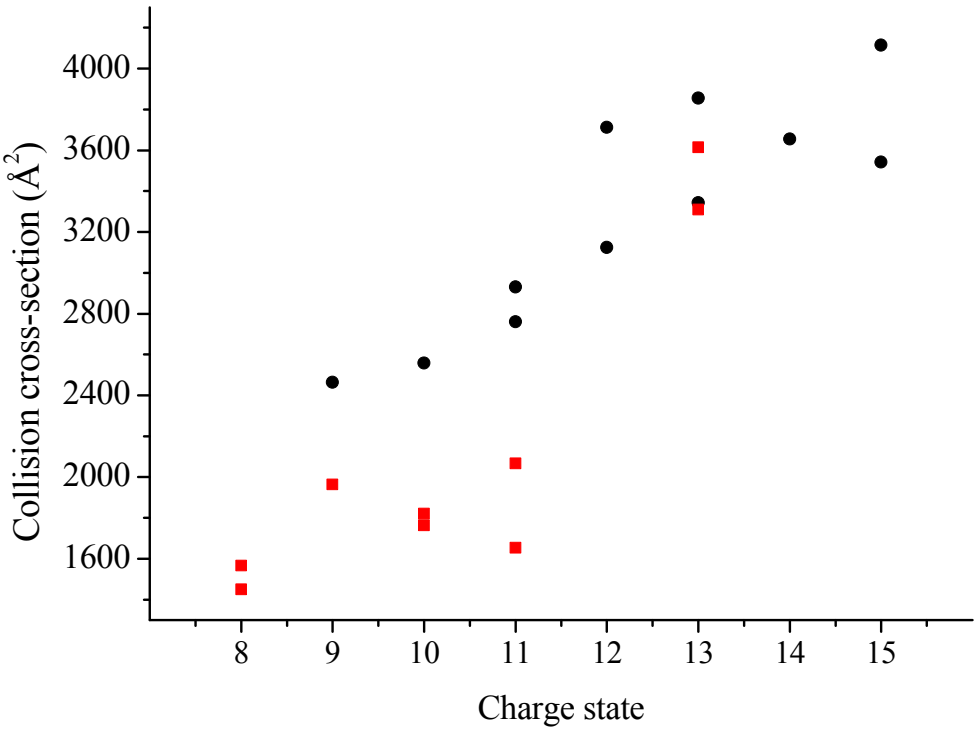



*Figure 5.12. Combined cross-section data for Synapt HDMS (black circles) and MoQTOF (red squares) of p53 with zinc removed. All cross-sections calculated from MoQTOF data are more compact than Synapt values.*

## 5.4 p53 Protein Conclusions

Interrogating protein-metal gas-phase conformations by mass spectrometry and ion mobility-mass spectrometry allows structural rearrangements to be probed. The DNA-binding core domain of p53 coordinates a zinc ion which stabilises the domain. Removal of this zinc ion causes a conformational change that provides an overall more compact structure at low charge states, possibly due to gas-phase collapse now that the cavity filled by the zinc ion is empty. Upon protonation, p53 with zinc present and removed adopt similar gas-phase conformations due to unfolding of the protein indicating that regions removed from the zinc binding cavity are unfolding.

The MoQTOF instrument provided consistently smaller cross-sections for all species over the entire charge state range when compared with cross-sections obtained using the Synapt HDMS instrument. This has been attributed to greater ion desolvation in the source region, and the possibility of additional protein unfolding in the trap region of the Synapt instrument.

## 5.5 Anterior Gradient-2 Protein Introduction

Anterior gradient-2 (AGR2) protein has been found upregulated in a range of cancers including prostate cancer;[20] oestrogen receptor-positive breast cancer;[21] and fibrolammelar carcinoma.[22] The exact mechanisms by which it is involved in cancer cell metastasis are a topic of intense research, but it is involved with the downregulation of the p53 DNA-damage response. Other functions of AGR2 (and homologs) involve intestinal develeopment[23] and limb regeneration in salamanders.[24] The protein was originally identified in the African clawed frog, *Xenopus laevis*, expressed during cement gland development,[25] and was identified a few years later in a proteomics screen for overexpressed proteins in metaplastic tissue of a condition called Barrett's oesophagus.[26, 27] Oesophageal epithelium is subject to damage from bile acid reflux, promoting normal tissue injury, resulting in the development of Barrett's epithelium.[26] Of note, it is these selection pressures (acid and bile stresses) that are being placed upon p53 gene mutation[28] which suggests a unique microenvironment to study and potentially identify novel p53 modifiers.[29]

AGR2 has two theoretical isoforms: a full-length isoform with an N-terminal hydrophobic leader sequence; and a mature form with the N-terminal removed by proteolytic processing. Removal of the N-terminal leader sequence of AGR2 allows the protein to exit the endoplasmic-reticulum, permitting the protein to exist in both the cytoplasm and nucleus, where correlation with AGR2 inhibition of p53 is thought to be initiated.[26] Studies by Fourtouna *et al.*[30] have shown that for a number of tumour cell lines, p53 accumulates in the nucleus upon UV irradiation. Transfection with AGR2 and further irradiation showed redistribution of p53 to cytosolic and membrane regions, providing potential evidence for AGR2 to mediate the nuclear exclusion of p53. If this relocation pathway can be disrupted, p53 may be able to accumulate in the nucleus and subsequently destroy the cancerous cell. AGR2 therefore represents a potentially novel drug target class, with its inhibition potentially leading to p53 activation.

A 12mer peptide aptamer that can interact with AGR2 *via* a peptide-binding groove was identified, containing a bioactive hexapeptide PTTIYY core.[30] Studies revealed that binding of this aptamer perturbed subcellular localisation of p53 by stimulating relocalisation of p53 to nuclear fractions. These results showed that AGR2 can mediate a change in nuclear-cytoplasmic localisation of p53 protein consistent with the link between p53 nuclear localisation and its function as a transcription factor. Control peptide aptamers were made containing mutations in the hexapeptide core to: PTTIYA, PTTIAY, PATIYY, PATIAA and the truncated PTTI.

FT-ICR-MS analysis of AGR2 under oxidative conditions revealed a covalently-bound homodimer formed by a disulfide bridge between residues Cys81.[31] Mutation of this cysteine residue to alanine prevented dimer formation, and reduced the activity of peptide aptamer binding, indicating that the residue is required to maintain a highly active peptide-binding core. The research also suggested that

the dimer interface may protect AGR2 from detection by protein degradation machinery and that AGR2 dimer to monomer conversion may form a physiological switch to either stabilise or degrade the protein. Dimerisation requires oxidation and it has been proposed that in a pro-oxidising environment like that commonly observed in cancer cells, AGR2 may be stabilised in a dimeric state.

## 5.6 AGR2 Experimental

AGR2 has a calculated mass of 21,041 Da and the peptide aptamer sequence masses are shown in Table 5.3 as calculated using MassLynx peptide editor. IM-MS data was obtained for AGR2 under native-like conditions with source temperature 80$^{o}$C; sample cone 135V; extractor cone 110V; 1.5 kV spray voltage; ~3.1 Torr cell pressure; 124 µs pusher period; and 40 µs pulse width.

*Table 5.3. Peptide aptamer masses used in complex with AGR2.*

| Aptamer peptide sequence (after *RRMKWKKSGSG*) | Average mass (Da) |
|---|---|
| PTTI | 1733.07 |
| PATIAA | 1845.20 |
| PTTIYA | 1967.33 |
| PTTIAY | 1967.33 |
| PATIYY | 2029.40 |
| PTTIYY | 2059.42 |

## 5.7 AGR2 results

### 5.7.1 Mass spectrometry of AGR2 Under Native-Like Conditions

50 µM AGR2 in 20 mM ammonium acetate (pH 6.85) was sprayed using nano-ESI under positive mode ionisation conditions and the mass spectrum is shown in Figure 5.13. A large number of gas-phase species were observed with monomer, dimer and trimer species present. Monomer species dominated the mass spectrum with a charge state distribution ranging from z = $6^+$ to $13^+$, with the z = $7^+$ most abundant. The spectrum indicates a relatively low charge density as would be expected from a protein sprayed under native-like conditions. Dimer species are also prevalent throughout the spectrum with peaks assigned to z = $13^+$ to $23^+$. Evenly-charged dimers have concomitant *m/z* values with singly-charged monomers with half the charge. Peaks have therefore been assigned with care. A small population of trimer peaks have also been assigned indicating successful transfer of another solution-phase multimer to the gas phase. A narrow charge state distribution of z = $15^+$ to $18^+$ has been assigned due to successful matching of z = $16^+$ and $17^+$ *m/z* values (3947.1 and 3715.0 *m/z* respectively).

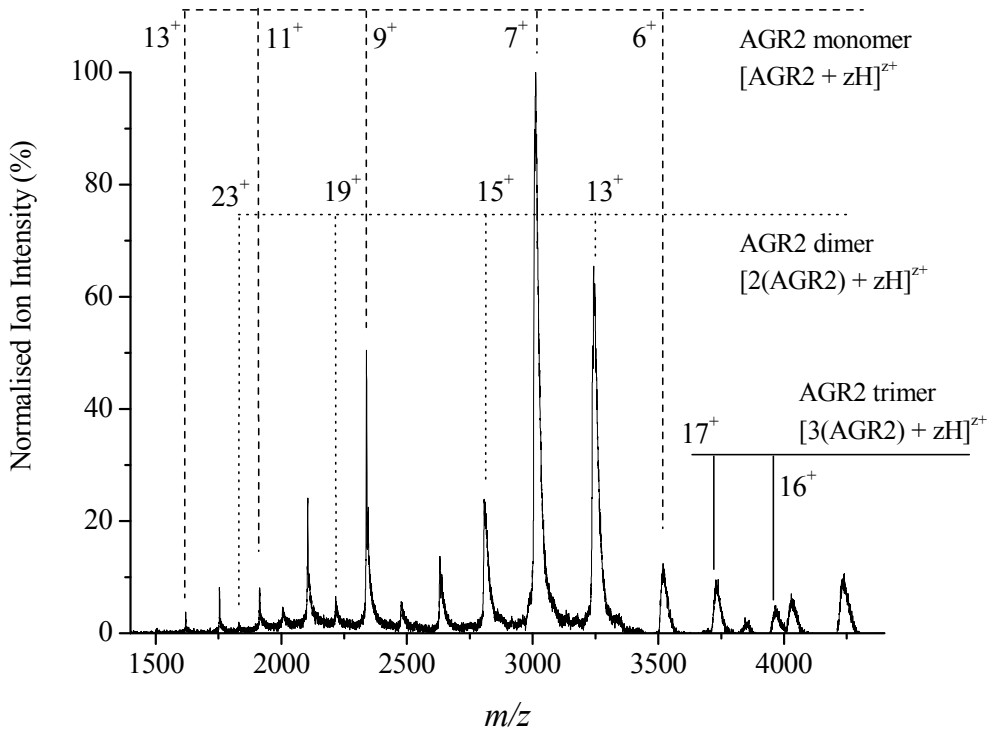



*Figure 5.13. AGR2 mass spectrum sprayed under native-like conditions (pH 6.85) showing monomer, dimer and trimer gas-phase species over the range 1500 to 4300 m/z.*

### 5.7.2 Ion Mobility-Mass Spectrometry of AGR2 Under Native-Like Conditions

Fifteen gas-phase AGR2 species were successfully characterised by IM-MS under native-like conditions. Eight monomer species ($z = 6^+$ to $13^+$) with two conformers for each charge state; five dimer species ($z = 13^+$, $15^+$, $19^+$, $21^+$ and $23^+$) with single gas-phase conformations; and two trimer species ($z = 17^+$ and $19^+$) with single gas-phase conformations. A number of ionic species observed by traditional mass spectrometry could not be characterised by IM-MS due to poor signal or poor data upon analysis (*e.g.* $z = 17^+$ dimer). Collision cross-sections for monomer species are presented in Table 5.4; dimer and trimer species in Table 5.5.

*Table 5.4. Collision cross-sections for AGR2 monomers sprayed under native-like conditions. Values are an average of two replicates & stated with twice the standard deviation of the mean (2σ).*

| Monomer charge state | Collision cross-section ($Å^2$) conformer 1 | Collision cross-section ($Å^2$) conformer 2 |
|---|---|---|
| 6 | 1190.4 ± 18.2 | 1354.7 ± 24.5 |
| 7 | 1308.2 ± 14.6 | 1408.3 ± 20.1 |
| 8 | 1419.4 ± 20.4 | 1627.1 ± 25.3 |
| 9 | 1687.0 ± 12.4 | 1812.7 ± 30.0 |
| 10 | 1737.5 ± 18.9 | 2046.3 ± 13.4 |
| 11 | 1911.3 ± 16.3 | 2216.1 ± 18.9 |
| 12 | 2159.3 ± 17.2 | 2394.2 ± 32.4 |
| 13 | 2292.4 ± 24.6 | 2450.2 ± 14.6 |

*Table 5.5. Collision cross-sections for AGR2 dimers and trimers sprayed under native-like conditions. Values are an average of two replicates & stated with twice the standard deviation of the mean (2σ). Species not observed indicated by n.o.*

| Charge state | Dimer collision cross-section ($Å^2$) | Trimer collision cross-section ($Å^2$) |
|---|---|---|
| 13 | 2703.9 ± 42.4 | n.o. |
| 15 | 3045.3 ± 53.6 | n.o. |
| 17 | n.o. | 3698.7 ± 56.2 |
| 19 | 3504.2 ± 54.9 | 3934.8 ± 71.4 |
| 21 | 4081.6 ± 46.3 | n.o. |
| 23 | 4374.0 ± 62.8 | n.o. |

Figure 5.14 presents raw experimental data for the two $z = 6^+$ monomer conformations with cross-sections 1190.4$Å^2$ and 1354.7$Å^2$ respectively.

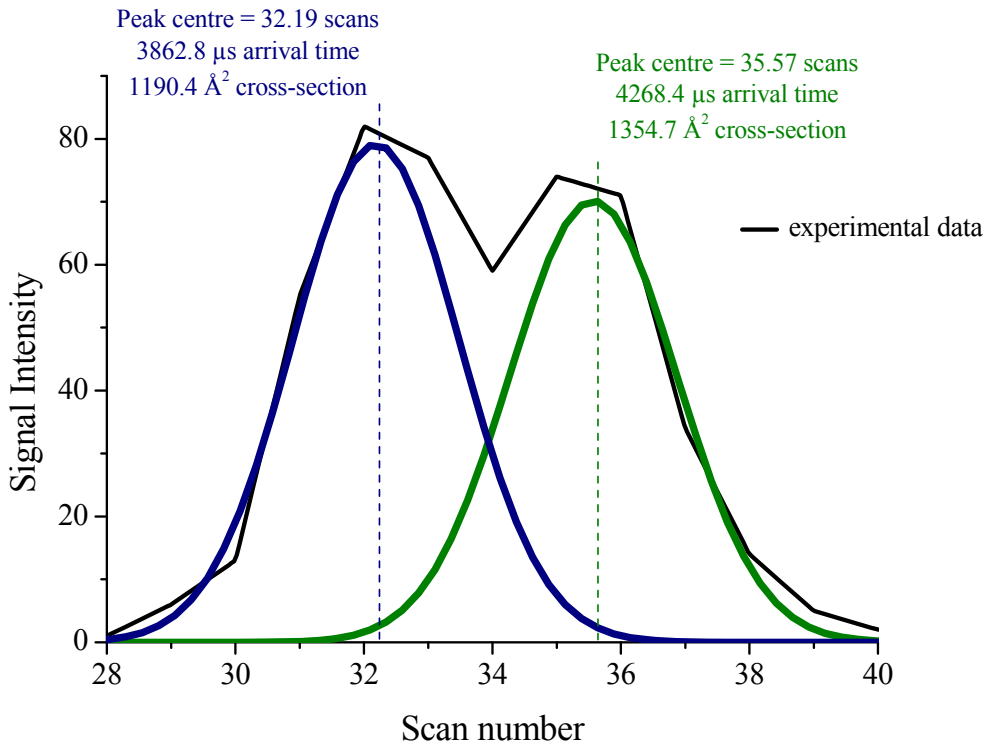



*Figure 5.14. Raw experimental data for both conformers of AGR2 monomer z = 6$^+$ showing arrival time and cross-sections.*

Figure 5.15 presents cross-sections as a function of charge state for all multimeric AGR2 species derived under these conditions.

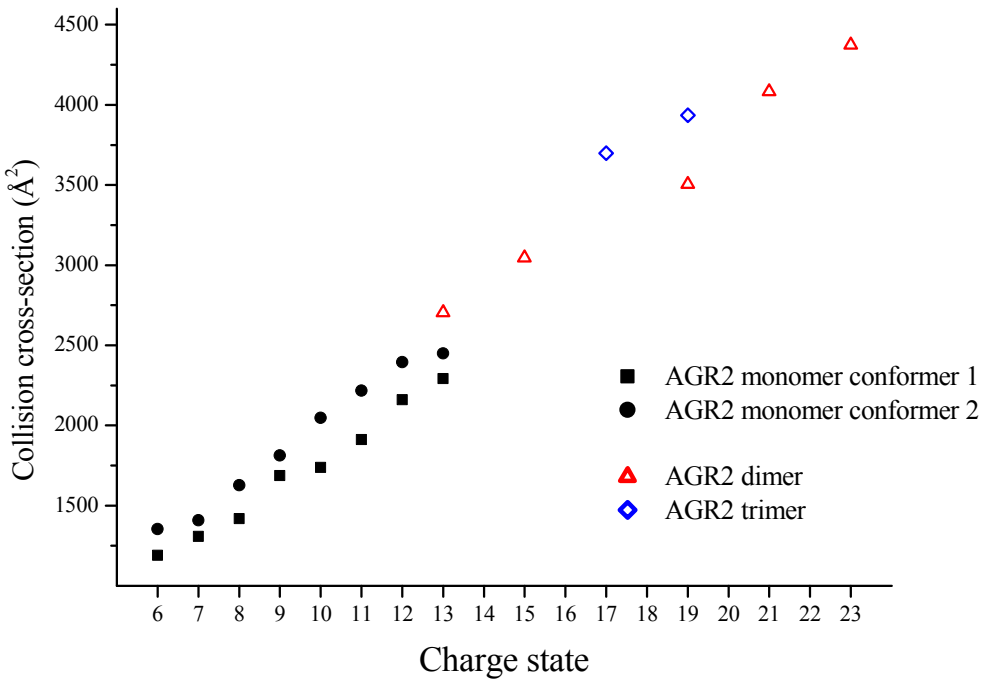



*Figure 5.15. Plot of collision cross-section against charge state for all multimeric species of AGR2 obtained under native-like conditions (pH 6.85).*

From Figure 5.15 it can be observed that increasing charge increases collision cross-section approximately linearly due to Coulomb unfolding of the protein. Protonation of solvent accessible basic residues promotes unfolding over the entire charge state region, with no evidence for regions of stability *i.e.* no obvious plateaus in the plot. Dimer cross-sections are smaller than double the possible constituent monomer cross-sections, indicating some molecular arrangement upon association. Trimer cross-sections may be regarded as relatively compact for such a large (~63.1 kDa) multimeric protein. Protonation of the trimer species from $z = 17^+$ to $19^+$ results in a 6.0% cross-section increase. This increase is small due to the large number of basic residues available on all three subunits exposed to solvent.

Raw experimental data for AGR2 monomer charge state $z = 11^+$ is shown in Figure 5.16. This clearly shows the two conformers (cross-sections 1911.3Å$^2$ and 2216.1Å$^2$) that have been obtained by fitting two separate curves to the experimental data ($R^2 = 0.9903$).

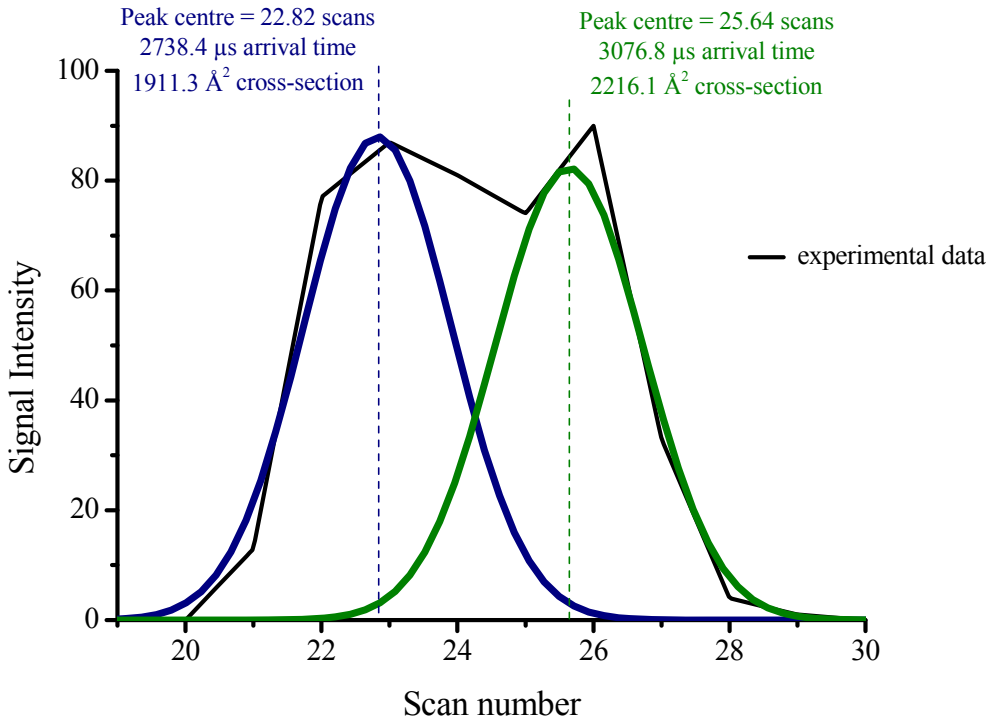



*Figure 5.16. Raw experimental data for the 11$^+$ charge state of AGR2. Two conformers were elucidated.*

### 5.7.3 Mass Spectrometry of AGR2 in Ammonium Acetate Containing 10% Isopropanol

50 µM AGR2 in 20 mM ammonium acetate containing 10% isopropanol by volume (pH 5.5) was sprayed using nano-ESI under positive mode ionisation conditions. The mass spectrum is shown in Figure 5.16. Monomer species with a charge state distribution similar to the native-like conditions is observed, with z = 6$^+$ to 13$^+$, however the most intense peak is the z = 9$^+$ (*c.f.* z = 7$^+$ under native-like conditions) *i.e.* a shift to higher charge state for the most abundant peak. A distribution of dimer species with z = 13$^+$ to 15$^+$ is observed. No trimer species are observed.

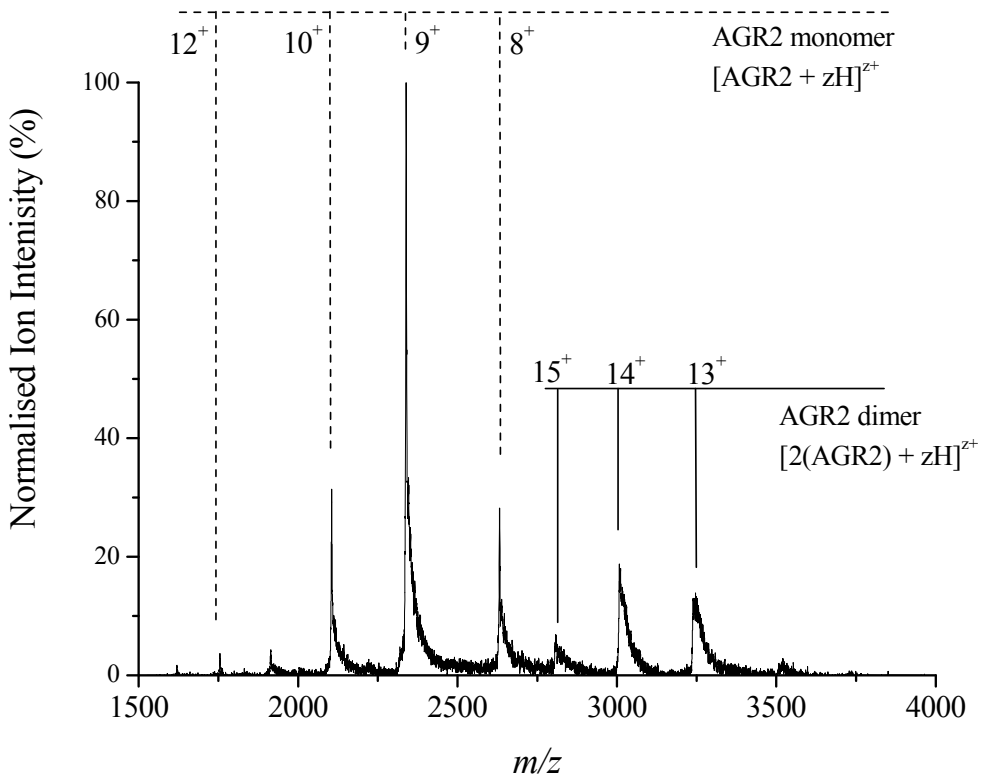



*Figure 5.17. AGR2 mass spectrum sprayed from 10% isopropanol and 20 mM ammonium acetate solution (pH 5.5) showing monomer and dimer gas-phase species (range 1500 to 4000 m/z).*

Upon addition of isopropanol, fewer multimeric gas-phase species are observed in the mass spectra and there is a shift in the most abundant monomer species to higher charge. This would imply a slight denaturing of the protein and may be reflected in the collision cross-sections calculated by IM-MS. Isopropanol may reduce the propensity of forming higher order multimers by causing structural rearrangement of exposed residues or regions that could interact with another AGR2 molecule. Gas-phase interactions are dominated by electrostatic interactions. If these are reduced by the presence of isopropanol, fewer higher order species will be formed.

### 5.7.4 Ion Mobility-Mass Spectrometry of AGR2 in Ammonium Acetate Containing 10% Isopropanol

Ten gas-phase AGR2 species were successfully characterised by IM-MS from a solution containing 10% by volume isopropanol 20 mM ammonium acetate (pH 5.5): seven monomer species ($z = 7^+$ to $13^+$) with two conformers for each charge state and three dimer species ($z = 13^+$ to $15^+$) with single gas-phase conformations. Collision cross-sections for all species are presented in Table 5.6.

*Table 5.6. Collision cross-sections for AGR2 monomers and dimers sprayed from 10% isopropanol in buffer. Values are an average of two replicates & stated with twice the standard deviation of the mean (2σ). Species not observed indicated by n.o.*

| Charge state | Monomer collisional cross-section ($Å^2$) conformer 1 | Monomer collisional cross-section ($Å^2$) conformer 2 | Dimer collisional cross-section ($Å^2$) |
|---|---|---|---|
| 7 | 1292.0 ± 28.4 | 1455.0 ± 39.1 | n.o. |
| 8 | 1636.2 ± 22.4 | 1694.8 ± 28.3 | n.o. |
| 9 | 1700.8 ± 34.6 | 1829.2 ± 22.1 | n.o. |
| 10 | 1795.6 ± 32.1 | 1905.4 ± 41.6 | n.o. |
| 11 | 1825.2 ± 36.7 | 2033.2 ± 24.9 | n.o. |
| 12 | 2024.3 ± 30.5 | 2212.4 ± 34.7 | n.o. |
| 13 | 2401.2 ± 32.6 | 2643.2 ± 35.7 | 2666.2 ± 48.7 |
| 14 | n.o. | n.o. | 2714.1 ± 49.2 |
| 15 | n.o. | n.o. | 2787.4 ± 42.9 |

Figure 5.18 presents cross-sections as a function of charge state for all multimeric AGR2 species derived under these conditions.

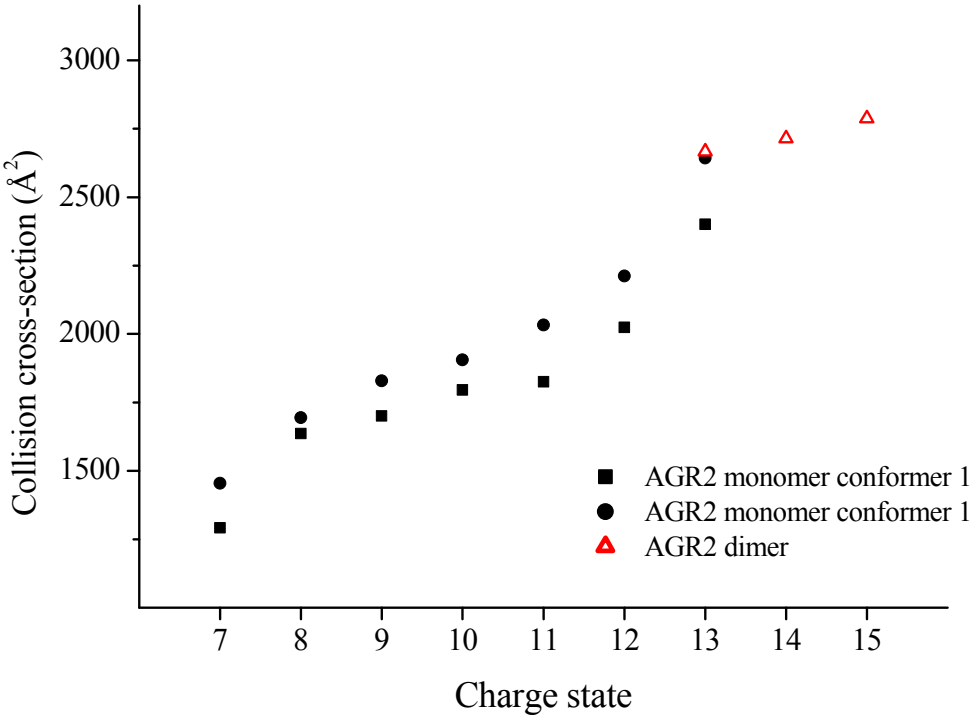



*Figure 5.18. Plot of collision cross-section against charge state for all multimeric species of AGR2 obtained from an ammonium acetate solution containing 10% isopropanol (pH 5.5).*

Collision cross-sections calculated for AGR2 containing isopropanol show a similar trend as those produced under native-like conditions, with increasing charge increasing cross-section. Comparison with cross-sections produced under native-like conditions show that addition of 10% isopropanol causes a slight contraction in cross-section for the majority of charge states. This contraction (more compact cross-section) is most pronounced at intermediate charge states where protonation is most likely causing local unfolding within domains. Isopropanol is perhaps stabilising helical structure in small regions, rather than causing a global stabilisation of the molecule. At the highest charge states ($z = 13^+$), the cross-section of the second isopropanol solution conformer is larger than that of the native-like solution and possibly attributed to structural movements within molecule due to isopropanol having exposed a region that, upon protonation, has caused a larger unfolding event.

### 5.7.5 Mass Spectrometry of AGR2 at pH 4.0

50 μM AGR2 in 50/50 by volume methanol/ammonium acetate buffer containing 2% formic acid (pH 4.0) was sprayed using nano-ESI under positive mode ionisation conditions. The mass spectrum is shown in Figure 5.19. As would be expected under decreased pH conditions, a wide charge state distribution, with $z = 8^+$ to $20^+$, is observed indicating an unfolded protein. However, the protein is not significantly denatured as it still retains compact unfolded structures as indicated by the dominance of the low charge states. A bimodal distribution, with centres at $z = 9^+$ and $12^+$ is observed implying two solution-phase conformers (most likely a near-native structured state, and an unfolded structure) have been transferred to the gas phase. Associated with the lower charge states ($z \leq 11^+$) adducts are observed that are assigned to non-specific binding of sodium and potassium cations. Tailing on the $z = 9^+$ and $8^+$ peaks is due to incomplete desolvation of more folded protein molecules.

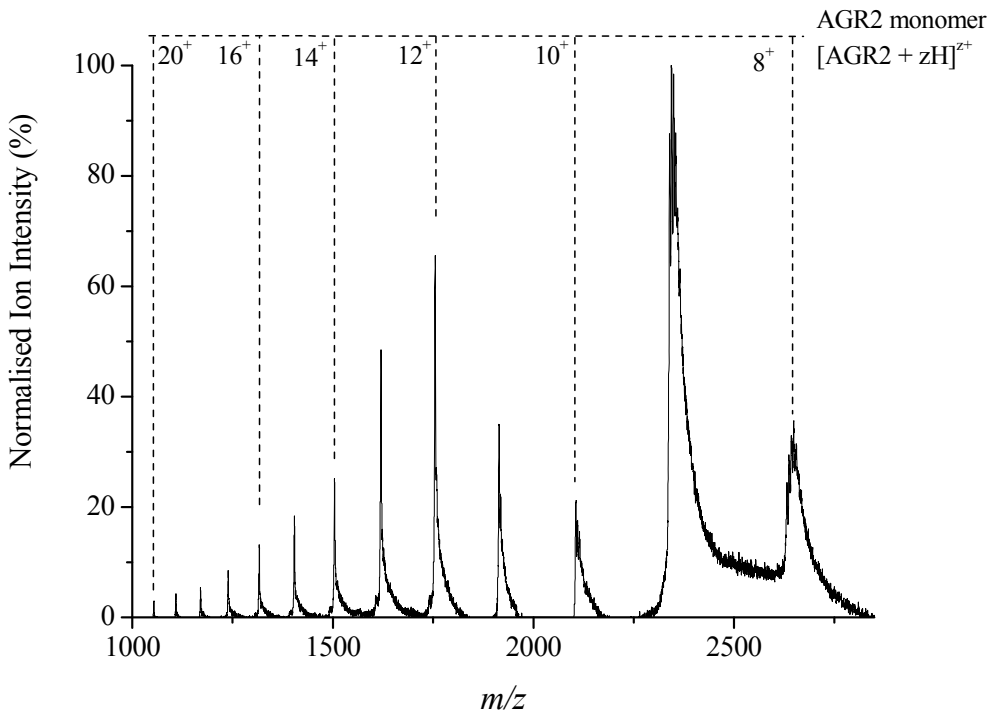



*Figure 5.19. AGR2 mass spectrum sprayed from 50/50 by volume methanol/ammonium acetate solution with 2% formic acid added (pH 4.0) showing only monomer gas-phase species over the range 1000 to 2800 m/z.*

### 5.7.6 Mass Spectrometry Studies of Peptide Aptamers RRMKWKKSGSGPxxxxx

The potential AGR2-inhibitor aptamer, RRMKWKKSGSGPTTIYY, was analysed by mass spectrometry and will be referred to using the six most C-terminal residues, PTTIYY. Subsequent aptamers will be referred to as modifications of this hexapeptide region, but all contain the RRMKWKKSGSG eleven amino acid sequence.

#### 5.7.6.1 Mass Spectrometry Investigation of Peptide Aptamer RRMKWKKSGSGPTTIYY

The mass spectrum for PTTIYY is presented in Figure 5.20. Peaks can be assigned to three intact peptide molecules with z = $2^+$ to $4^+$, with the z = $3^+$ charge state most abundant. A b ion series indicating successive loss of both C-terminal tyrosine residues, $b_{15}$ and $b_{16}$ with z = $3^+$, has also been assigned.

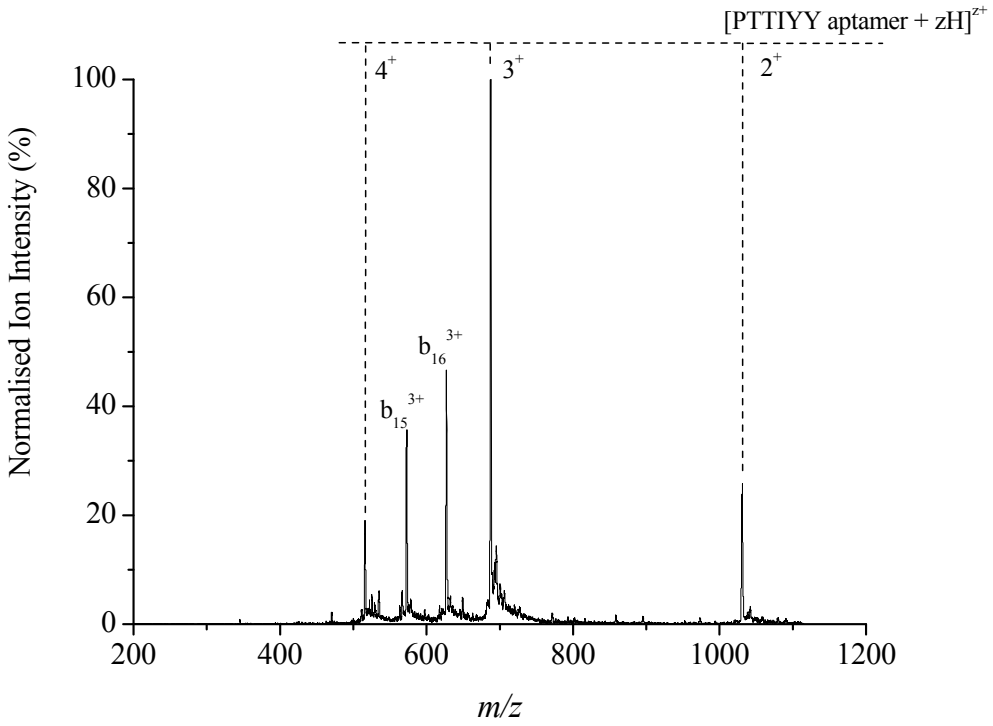



*Figure 5.20. Mass spectrum of aptamer PTTIYY over the range 200 to 1200 m/z.*

### 5.7.6.2 Mass Spectrometry Investigation of Peptide Aptamers RRMKWKKSGSGPTTIYA and RRMKWKKSGSGPTTIAY

The bioactive hexapeptide region was modified with one tyrosine mutated for an alanine residue at both positions. The mass spectra for these isomeric aptamers are presented in Figures 5.21 and 5.22 respectively.

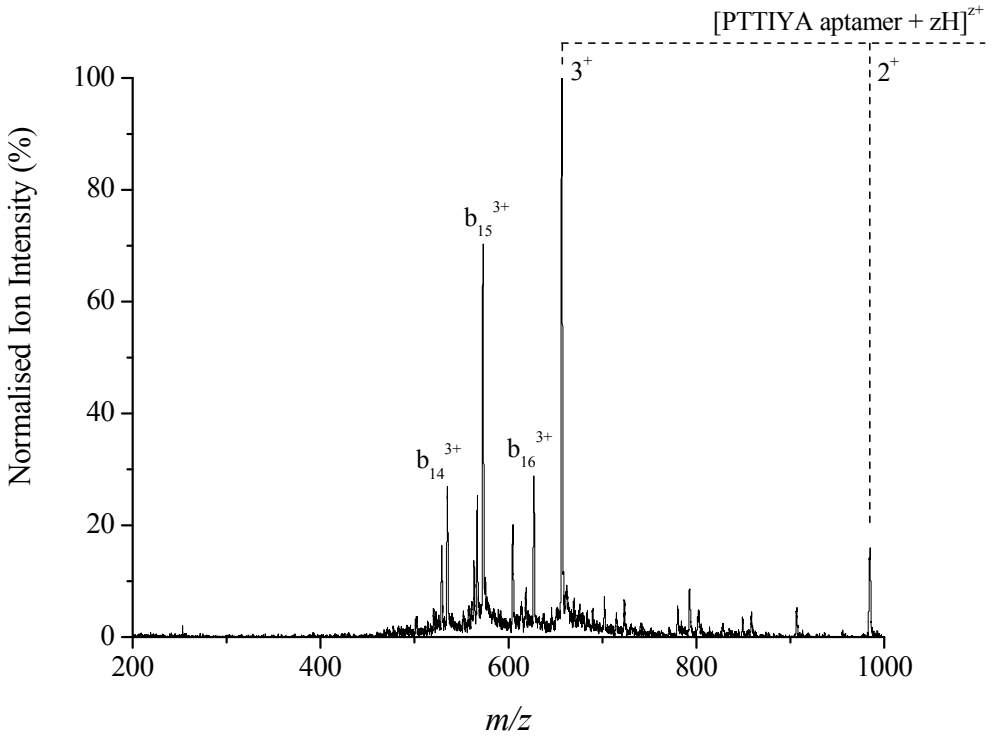



*Figure 5.21. Mass spectrum of aptamer PTTIYA over the range 200 to 1000 m/z.*

The mass spectrum of aptamer PTTIYA (Figure 5.21) has two charge states of the full length peptide with $z = 2^+$ and $3^+$. Three further peaks can unambiguously be identified as a b ion series with $z = 3^+$, indicating loss of the C-terminal alanine ($b_{16}$); alanine and tyrosine ($b_{15}$); alanine, tyrosine and isoleucine ($b_{14}$) residues respectively. No further peaks could be unambiguously assigned.

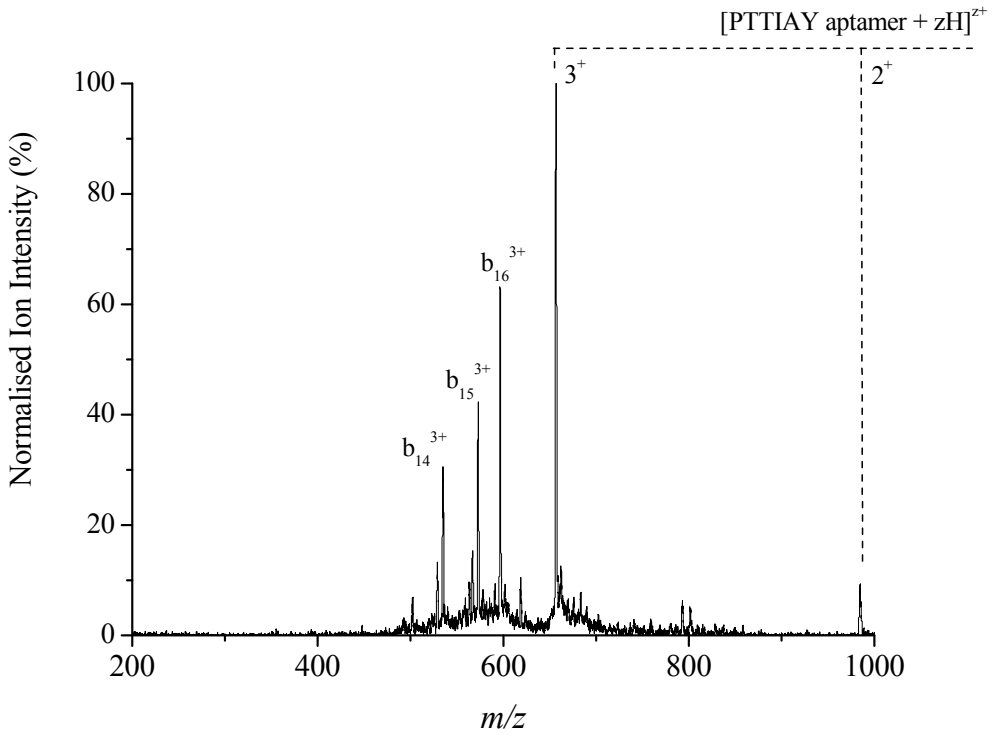



*Figure 5.22. Mass spectrum of aptamer PTTIAY over the range 200 to 1000 m/z.*

The mass spectrum of aptamer PTTIAY (Figure 5.22) has two charge states of the full length peptide with $z = 2^+$ and $3^+$. Three further peaks can unambiguously be identified as a b ion series with $z = 3^+$, indicating loss of the C-terminal tyrosine ($b_{16}$); tyrosine and alanine ($b_{15}$); and tyrosine, alanine and isoleucine ($b_{14}$) residues respectively as was observed for the PTTIYA aptamer in Figure 5.20. No further peaks could be unambiguously assigned.

### 5.7.6.3 Mass Spectrometry Investigation of Peptide Aptamer RRMKWKKSGSGPATIYY

The next modification to the hexapeptide bioactive core involved mutation of the first threonine residue to alanine to produce the PATIYY variant. The mass spectrum for this aptamer is presented in Figure 5.23.

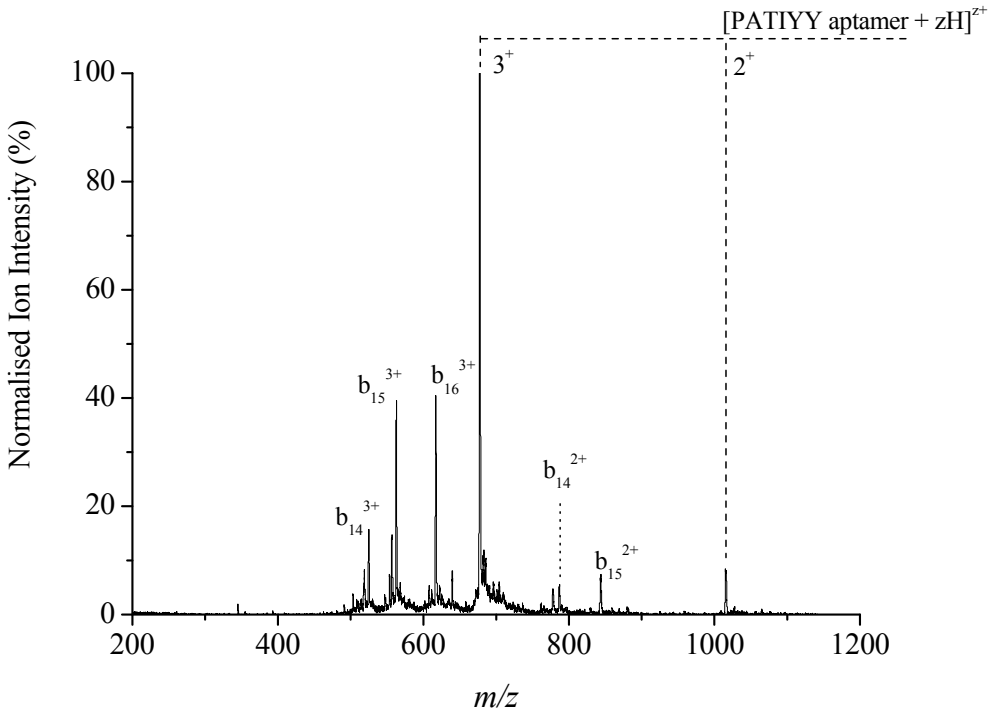



*Figure 5.23. Mass spectrum of aptamer PATIYY over the range 200 to 1200 m/z.*

A greater number of peaks can be assigned in the spectrum for PATIYY than previous aptamers, with two charge states ($z = 2^+$ and $3^+$) again present. Two sets of b ions are observed. $b_{14\text{-}16}$ with $z = 3^+$ indicates loss of the C-terminal tyrosine ($b_{16}$); two tyrosines ($b_{15}$); and two tyrosines and isoleucine ($b_{14}$) residues respectively. $b_{14\text{-}15}$ with $z = 2^+$ indicates loss of two tyrosines ($b_{15}$); and two tyrosines and isoleucine ($b_{14}$) residues respectively. No further peaks could be unambiguously assigned.

### 5.7.6.4 Mass Spectrometry Investigation of Peptide Aptamer RRMKWKKSGSGPATIAA

Mutation of the first threonine to alanine, and both C-terminal tyrosine residues to alanine produced the PATIAA aptamer. The mass spectrum for this aptamer is presented in Figure 5.24.

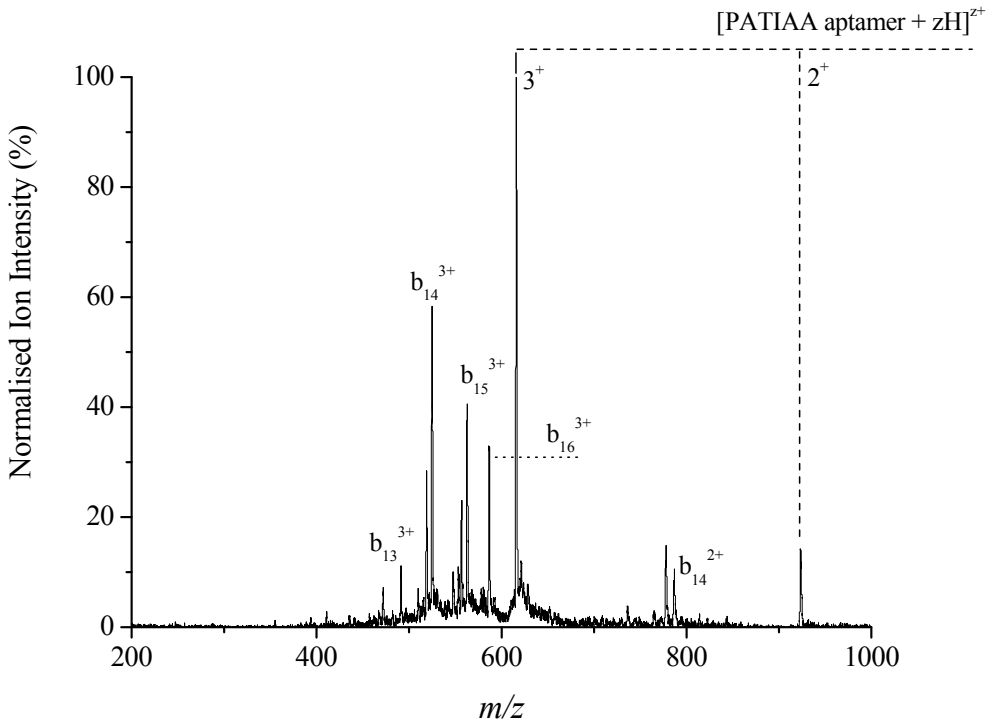



*Figure 5.24. Mass spectrum of aptamer PATIAA over the range 200 to 1000 m/z.*

In the PATIAA mass spectrum, two charge states ($z = 2^+$ and $3^+$) are present. Two sets of b ions are observed. $b_{13\text{-}16}$ with $z = 3^+$ indicates loss of the C-terminal tyrosine ($b_{16}$); two tyrosines ($b_{15}$); two tyrosines and isoleucine ($b_{14}$); and two tyrosines, isoleucine and threonine ($b_{13}$) residues respectively. A single peak assigned to $b_{14}$ with $z = 2^+$ indicates loss of and two tyrosines and isoleucine ($b_{14}$) residues. No further peaks could be unambiguously assigned.

### 5.7.6.5 Mass Spectrometry Investigation of Peptide Aptamer RRMKWKKSGSGPTTI

A truncated form of the PTTIYY hexapeptide was produced and characterised by mass spectrometry (Figure 5.25).

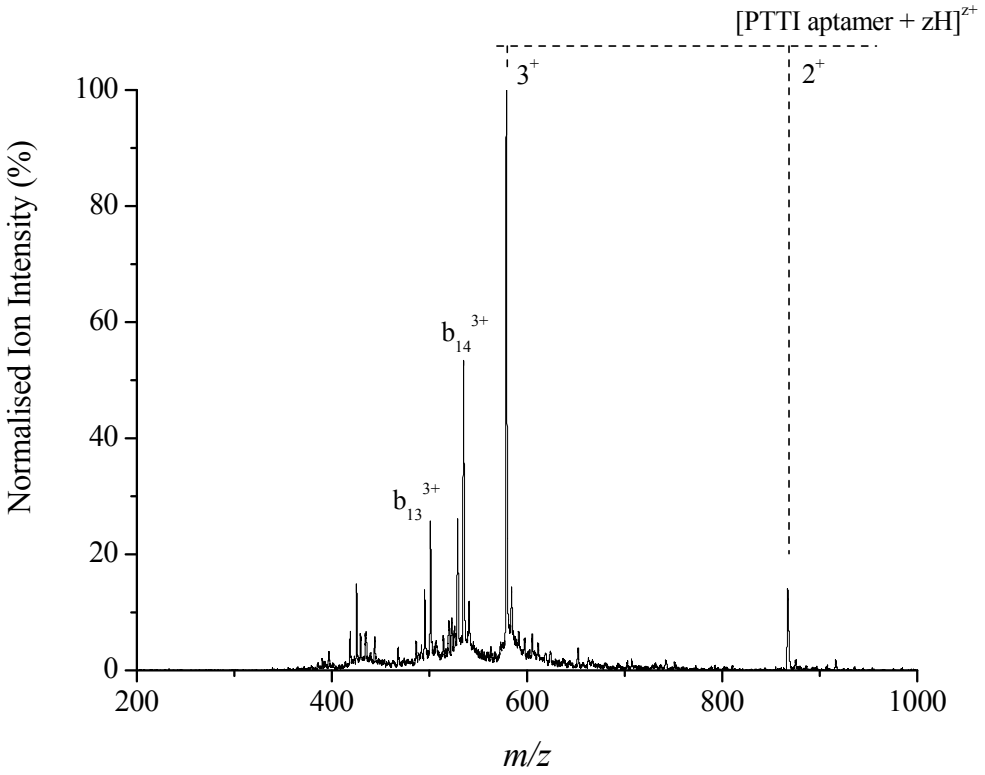



*Figure 5.25. Mass spectrum of aptamer PTTI over the range 200 to 1000 m/z.*

The PTTI mass spectrum displays two charge states ($z = 2^+$ and $3^+$). Two b ions have been assigned, $b_{13}$ and $b_{14}$ with $z = 3^+$, indicating loss of the C-terminal isoleucine ($b_{14}$); and isoleucine and tyrosine ($b_{13}$) residues. No further peaks could be unambiguously assigned.

### 5.7.7 Ion Mobility-Mass Spectrometry Studies of Peptide Aptamers

Collision cross-sections for intact charge states (*i.e.* not b ions) of all six aptamers were calculated and are presented in Table 5.7. Two gas-phase conformers were detected for some aptamer charge states.

*Table 5.7. Collision cross-sections for all six aptamer intact molecules. Values are an average of two replicates & stated with twice the standard deviation of the mean (2σ). Species not observed indicated by n.o.*

| Species & charge state | Collision cross-section conformer 1 (Å$^2$) | Collision cross-section conformer 2 (Å$^2$) |
|---|---|---|
| PATIYY, $2^+$ | 363.4 ± 4.1 | n.o. |
| PATIYY, $3^+$ | 380.6 ± 3.2 | n.o. |
| PTTIYA, $2^+$ | 247.6 ± 1.3 | n.o. |
| PTTIYA, $3^+$ | 377.8 ± 4.3 | n.o. |
| PATIAA, $2^+$ | 233.1 ± 5.1 | 319.2 ± 2.1 |
| PATIAA, $3^+$ | 366.9 ± 2.1 | n.o. |
| PTTI, $2^+$ | 260.1 ± 4.8 | 335.8 ± 5.1 |
| PTTI, $3^+$ | 360.0 ± 6.1 | n.o. |
| PTTIAY, $2^+$ | 238.4 ± 3.7 | 309.1 ± 6.4 |
| PTTIAY, $3^+$ | 379.1 ± 4.2 | n.o. |
| PTTIYY, 2+ | 348.9 ± 8.1 | n.o. |
| PTTIYY, 3+ | 403.8 ± 6.3 | 504.0 ± 7.2 |
| PTTIYY, 4+ | 462.2 ± 9.1 | n.o. |

Cross-sections for the z = $2^+$ aptamers are plotted in Figure 5.26 for comparison between the mutated constructs.

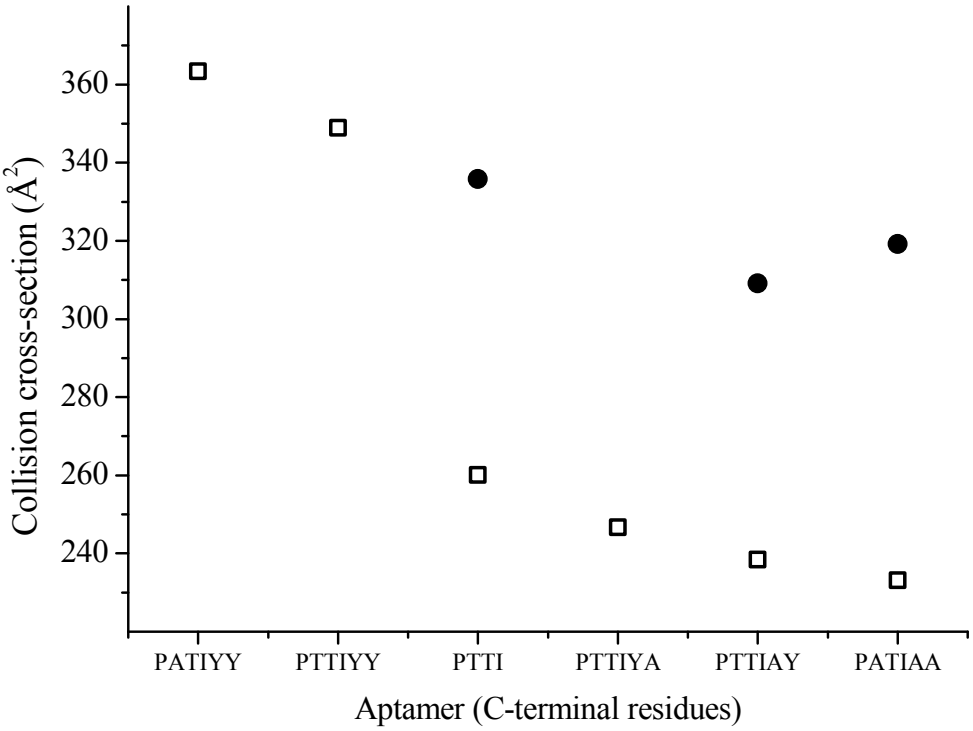



*Figure 5.26. Collision cross-sections for all six aptamers (z = 2$^+$). Two conformers (unfilled squares and filled circles) were elucidated for three of aptamers.*

From Figure 5.26 it can be observed that small variations in the C-terminal residues can change the cross-section by a significant amount. Alanine residues are known to promote helix formation, perhaps explaining the smallest cross-section for one conformation of aptamer PATIAA. The related aptamer PATIYY has the largest cross-section observed possibly due to unfolding of the helix and the C-terminal tyrosines projecting away from the backbone. This effect is also observed with the related aptamers PTTIYY and PTTI. Again, the presence of the C-terminal tyrosines promotes a greater cross-section by 88.8 Å$^2$. The remaining two aptamers, PTTIYA and PTTIAY differ only in the position of the last two residues, with a cross-section difference of 9.2 Å$^2$, indicating adoption of similar gas-phase structures.

## 5.8 Mass Spectrometry & Ion Mobility-Mass Spectrometry Studies of the AGR2 Complex with PTTIYY

There is promising evidence that the PTTIYY aptamer, when complexed with AGR2, disrupts the pathway that leads to downregulation of the tumour suppressor protein p53. This mechanism is still not fully understood, but using mass spectrometry to characterise the initial binding stoichiometry may lead to a further understanding.

### 5.8.1 Mass Spectrometry Results for the AGR2 Complex with PTTIYY

Five different gas-phase species were observed upon analysing the AGR2:PTTIYY solution:

i) unbound PTTIYY peptide ($z = 2^+$);
ii) AGR2 monomer;
iii) AGR2 dimer;
iv) AGR2:PTTIYY complex (1:1);
v) $(AGR2)_2$:PTTIYY complex (2:1).

The mass spectrum presented in Figure 5.27 shows the $z = 9^+$ monomer as the most abundant peak. The relatively low charge density of all the species present indicates that the protein is in a native-like state and not unfolded. This is supported by the presence of the complex, not only in monomeric form, but also as a dimeric. Only one aptamer is observed to bind, to both monomeric and dimeric forms, which may indicate the binding site is blocked by the second monomer after association to the complexed monomer. The intensity of the unbound peptide aptamer may be indicative of peptide that has dissociated from the AGR2 (as there is sufficient unbound monomer of greater intensity to bind to), peptide that was not able to bind to AGR2 due to conformational restrictions, or finally that spectrometer settings were tuned to be more selective towards higher *m/z* species.

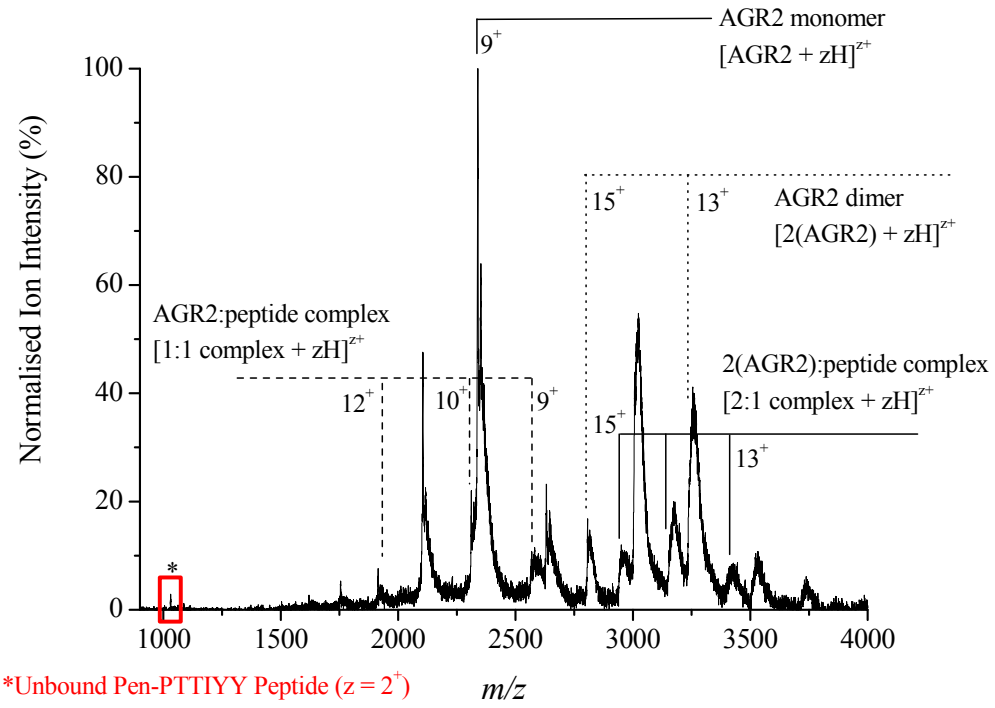



*Figure 5.27. Mass spectrum of AGR2:PTTIYY complex over the range 900 to 4000 m/z showing a large number of gas-phase species.*

Collision cross-sections for eight monomer and three dimer AGR2 species are presented in Table 5.8. The $z = 7^+$ monomer and $z = 14^+$ dimer have concomitant *m/z* values but two species could be distinguished using IM-MS, allowing determination of both values.

*Table 5.8. Collision cross-sections for AGR2 monomer and dimer species. Values are an average of two replicates & stated with twice the standard deviation of the mean (2σ). Species not observed indicated by n.o.*

| Charge state | Monomer collision cross-section ($Å^2$) | Dimer collision cross-section ($Å^2$) |
|---|---|---|
| 6 | 1314.7 ± 18.2 | n.o. |
| 7 | 1400.6 ± 24.3 | n.o. |
| 8 | 1691.6 ± 24.2 | n.o. |
| 9 | 1728.1 ± 19.2 | n.o. |
| 10 | 1806.5 ± 14.6 | n.o. |
| 11 | 2086.6 ± 26.4 | n.o. |
| 12 | 2195.7 ± 28.8 | n.o. |
| 13 | 2369.7 ± 29.6 | 2764.2 ± 43.5 |
| 14 | n.o. | 2797.1 ± 36.8 |
| 15 | n.o. | 3072.7 ± 45.5 |

To assist in characterising the AGR2 monomer, a plot of cross-section is presented in Figure 5.28 showing: AGR2 monomer from native-like conditions, AGR2 monomer from 10% isopropanol solution, and AGR2 monomer not in complex with PTTIYY aptamer.

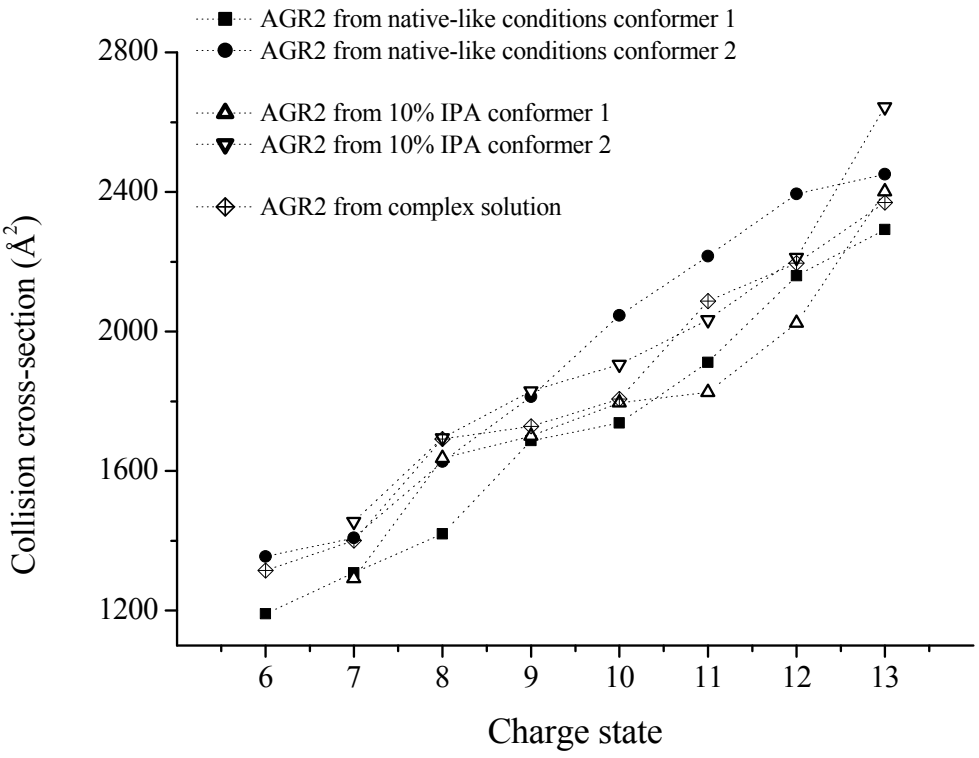



*Figure 5.28. Cross-section plot over a range of charge states for all monomeric AGR2 species obtained. Dotted lines are for guidance only and do not imply a linear trend.*

AGR2 monomer (conformer 1) from native-like conditions retains the most compact structure over the vast majority of charge states, with only $z = 11^+$ and $12^+$ being slightly less compact compared to one conformer from the solution sprayed from isopropanol. At $z = 10^+$ to $12^+$, conformer 2 adopts the largest (least compact) structure of all monomers observed and is possibly attributed to monomer that has dissociated from dimer species. The second conformer elucidated from the isopropanol solution retains the second largest cross-section over the charge state range and indicates protein unfolding perhaps due to mild protein denaturation and small local unfolding. The AGR2 monomer cross-sections obtained from the sample containing the aptamer PTTIYY appear to be accommodated between both the native-like and isopropanol samples suggesting that protein is no longer in its native-like geometry, nor has it unfolded to a large extent. This could be due to a short-lived binding event with the PTTIYY aptamer, inducing a conformational change, and then dissociation of the aptamer in the gas phase producing an aptamer-free monomer with a conformation different from that of the native-like monomer that has had no contact with the aptamer. Cross-sections for all three species do appear to follow a similar unfolding trend at low, intermediate and high charge states.

Collision cross-sections for four AGR2:PTTIYY (1:1) and three $(AGR2)_2$:PTTIYY (2:1) complex species are presented in Table 5.9.

*Table 5.9. Collision cross-sections for AGR2:PTTIYY monomeric and dimeric complex species. Values are an average of two replicates & stated with twice the standard deviation of the mean (2$\sigma$). Species not observed indicated by n.o.*

| Charge state | AGR2:PTTIYY (1:1) complex collision cross-section ($Å^2$) | $(AGR2)_2$:PTTIYY (2:1) complex collision cross-section ($Å^2$) |
|---|---|---|
| 9 | 1835.8 ± 40.2 | n.o. |
| 10 | 1875.7 ± 38.9 | n.o. |
| 11 | 1982.2 ± 48.7 | n.o. |
| 12 | 2336.8 ± 47.6 | n.o. |
| 13 | n.o. | 2780.6 ± 32.4 |
| 14 | n.o. | 3361.8 ± 45.1 |
| 15 | n.o. | 3463.9 ± 47.3 |

Cross-sections for AGR:PTTIYY species are compared with non-complexed species under native-like conditions in Figure 5.29 to show any conformational change undergone upon binding aptamer binding.

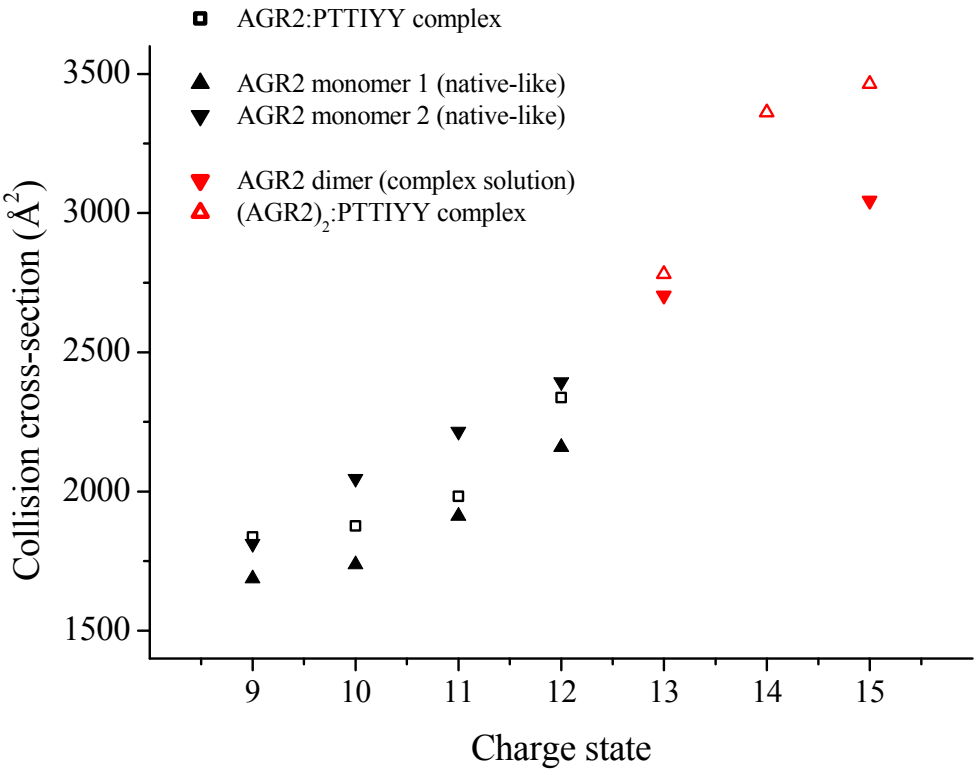



*Figure 5.29. Comparison of cross-sections between PTTIYY-bound (open shapes) and unbound (filled shapes) species of monomer (black shapes) and dimer (red shapes) against native-like monomer and dimer.*

### 5.8.2 Ion Mobility-Mass Spectrometry Results for the AGR2 Complex with PTTIYY (1:1)

Cross-sections for the 1:1 complex over four charge states were elucidated and compared with unbound AGR2 monomer under native-like conditions. At the lowest charge state obtained, $z = 9^+$, the cross-section for the complex is larger than both native-like conformers (by 23.1 and 148.8 $Å^2$) indicating a conformational change has occurred within the protein structure upon binding the aptamer. This conformational change is difficult to interpret at the remaining three charge states as the complex cross-section resides between the two gas-phase conformers of the native-like monomer. If the larger monomer conformer is assumed to come from dissociation of the dimer complex and comparisons are made with the smaller conformer, the complex assumes a larger cross-section across the entire charge state range, indicating a binding conformation transition upon aptamer association. The largest difference (177.5 $Å^2$) is observed at $z = 12^+$.

### 5.8.3 Ion Mobility-Mass Spectrometry Results for the AGR2 Complex with PTTIYY (2:1)

Cross-sections for the 2:1 complex over three charge states were elucidated and compared with dimeric AGR2 in the absence of PTTIYY under native-like conditions. The dimer:PTTIYY complex has larger cross-sections over all three charge states, indicating that upon aptamer binding the protein structure increases in size to accommodate the peptide at its binding site. A large conformational transition occurs between charge states $z = 13^+$ and $14^+$ with a change of 581.2 $Å^2$ in cross-section. Protonation of the $z = 14^+$ species to the $z = 15^+$ induced a shift in cross-section of 102.1 $Å^2$, approximately 17.6% of the previous transition ($z = 13^+ \rightarrow 14^+$). Unfortunately, a cross-section for the $z = 14^+$ dimer under native-like conditions could not be extracted. It can be observed that the $z = 13^+$ charge state for both complexed and uncomplexed dimer species differ by only 76.7 $Å^2$, whereas the $z = 15^+$ species differ by 418.6 $Å^2$. Protonation of the complex causes a greater increase in cross-section and is probably due to protonation of the monomer subunit *not* involved in aptamer binding, thereby causing unfolding. This is supported by the presence of the larger monomer conformer under native-like conditions.

Unbound aptamer with $z = 2^+$ was also observed in the mass spectrum and a cross-section of 342.2 $Å^2$ was elucidated. This compares well with the cross-section obtained from spraying the aptamer alone (348.9 $Å^2$) indicating gas-phase structure adoption is similar.

## 5.9 AGR2 Protein Conclusions

AGR2 has been successfully transferred to the gas phase under native-like conditions. Monomer, non-covalently bound dimer and trimer species have all been observed, as well as complexes with the aptamer peptide PTTIYY. Cross-sections for all species, including free aptamers, have been elucidated and show that any conformational change in the complex is small when compared with the unbound monomer. This small change could mean a restructuring of the protein to expose favourable interactions regions or residues, producing a molecular structure with similar cross-section as the unbound monomer. Aptamer peptides all show a degree of residue loss from the C-terminal with mutation indicating loss of structural stability in the gas-phase.

Further in-depth mass spectrometry studies, such as protein-ligand binding to produce dissociation constants, and the use of higher aptamer concentrations may provide further structural and biological information on this potential anticancer agent.

## 5.10 Chapter 5. References

1. Vousden, K. H.; Lane, D. P., p53 in health and disease. *Nat. Rev. Mol. Cell Biol.* **2007,** 8, (4), 275-83.

2. Lane, D. P., p53, guardian of the genome. *Nature* **1992,** 358, (6381), 15-16.

3. Showalter, S. A.; Bruschweiler-Li, L.; Johnson, E.; Zhang, F.; Bruschweiler, R., Quantitative Lid Dynamics of MDM2 Reveals Differential Ligand Binding Modes of the p53-Binding Cleft. *J. Am. Chem. Soc.* **2008,** 130, (20), 6472-6478.

4. Dey, A.; Verma, C. S.; Lane, D. P., Updates on p53: modulation of p53 degradation as a therapeutic approach. *Br. J. Cancer* 98, (1), 4-8.

5. Dey, A.; Tergaonkar, V.; Lane, D. P., Double-edged swords as cancer therapeutics: simultaneously targeting p53 and NF-[kappa]B pathways. *Nat. Rev. Drug Discov.* **2008,** 7, (12), 1031-1040.

6. Alarcon-Vargas, D.; Ronai, Z., p53-Mdm2--the affair that never ends. *Carcinogenesis* **2002,** 23, (4), 541-7.

7. Harris, S. L.; Levine, A. J., The p53 pathway: positive and negative feedback loops. *Oncogene* **2005,** 24, (17), 2899-908.

8. Levine, A. J.; Hu, W.; Feng, Z., The p53 pathway: what questions remain to be explored? *Cell Death Differ.* **2006,** 13, (6), 1027-1036.

9. Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P., Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **1996,** 274, (5289), 948-53.

10. Hainaut, P.; Mann, K., Zinc Binding and Redox Control of p53 Structure and Function. *Antioxidants & Redox Signaling* **2001,** 3, (4), 611-623.

11. Meplan, C.; Richard, M. J.; Hainaut, P., Metalloregulation of the tumor suppressor protein p53: zinc mediates the renaturation of p53 after exposure to metal chelators in vitro and in intact cells. *Oncogene* **2000,** 19, (46), 5227-36.

12. Reed, M.; Wang, Y.; Mayr, G.; Anderson, M. E.; Schwedes, J. F.; Tegtmeyer, P., p53 domains: suppression, transformation, and transactivation. *Gene Expr.* **1993,** 3, (1), 95-107.

13. Qilu Ye, H. W. J. Z. Q. W. Z. J., The complex structure of calmodulin bound to a calcineurin peptide. *Proteins: Structure, Function, and Bioinformatics* **2008,** 73, (1), 19-27.

14. Canadillas, J. M. P.; Tidow, H.; Freund, S. M. V.; Rutherford, T. J.; Ang, H. C.; Fersht, A. R., Solution structure of p53 core domain: Structural basis for its instability. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** 103, (7), 2109-2114.

15. Olivier, M.; Eeles, R.; Hollstein, M.; Khan, M. A.; Harris, C. C.; Hainaut, P., The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum. Mutat.* **2002,** 19, (6), 607-14.

16. Versluis, C.; Heck, A. J. R., Gas-phase dissociation of hemoglobin. *Int. J. Mass Spectrom.* **2001,** 210-211, 637-649.

17. Griffith, W. P.; Kaltashov, I. A., Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry. *Biochem.* **2003,** 42, (33), 10024-33.

18. Jurchen, J. C.; Williams, E. R., Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. *J. Am. Chem. Soc.* **2003,** 125, (9), 2817-2826.

19. Jurchen, J. C.; Garcia, D. E.; Williams, E. R., Further studies on the origins of asymmetric charge partitioning in protein homodimers. *J. Am. Soc. Mass Spectrom.* **2004,** 15, (10), 1408-1415.

20. Zhang, J.-S.; Gong, A.; Cheville, J. C.; Smith, D. I.; Young, C. Y. F., AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. *Genes, Chromosomes and Cancer* **2005,** 43, (3), 249-259.

21. Innes, H. E.; Liu, D.; Barraclough, R.; Davies, M. P. A.; O'Neill, P. A.; Platt-Higgins, A.; de Silva Rudland, S.; Sibson, D. R.; Rudland, P. S., Significance of the metastasis-inducing protein AGR2 for outcome in hormonally treated breast cancer patients. *Br. J. Cancer* **2006,** 94, (7), 1057-1065.

22. Vivekanandan, P.; Micchelli, S. T. L.; Torbenson, M., Anterior gradient-2 is overexpressed by fibrolamellar carcinomas. *Human Path.* **2009,** 40, (3), 293-299.

23. Chang, J.; Chance, M. R.; Nicholas, C.; Ahmed, N.; Guilmeau, S.; Flandez, M.; Wang, D.; Byun, D.-S.; Nasser, S.; Albanese, J. M.; Corner, G. A.; Heerdt, B. G.; Wilson, A. J.; Augenlicht, L. H.; Mariadason, J. M., Proteomic changes during intestinal cell maturation in vivo. *J. Proteomics* **2008,** 71, (5), 530-546.

24. Kumar, A.; Godwin, J. W.; Gates, P. B.; Garza-Garcia, A. A.; Brockes, J. P., Molecular Basis for the Nerve Dependence of Limb Regeneration in an Adult Vertebrate. *Science* **2007,** 318, (5851), 772-777.

25. Aberger, F.; Weidinger, G.; Grunz, H.; Richter, K., Anterior specification of embryonic ectoderm: the role of the Xenopus cement gland-specific gene XAG-2. *Mech. Dev.* **1998,** 72, (1-2), 115-130.

26. Pohler, E.; Craig, A. L.; Cotton, J.; Lawrie, L.; Dillon, J. F.; Ross, P.; Kernohan, N.; Hupp, T. R., The Barrett's Antigen Anterior Gradient-2 Silences the p53 Transcriptional Response to DNA Damage. *Mol. Cell. Proteomics* **2004,** 3, (6), 534-547.

27. Yagui-Beltran, A.; Craig, A. L.; Lawrie, L.; Thompson, D.; Pospisilova, S.; Johnston, D.; Kernohan, N.; Hopwood, D.; Dillon, J. F.; Hupp, T. R., The human oesophageal squamous epithelium exhibits a novel type of heat shock protein response. *Eur. J. Biochem.* **2001,** 268, (20), 5343-5355.

28. Leedham, S. J.; Preston, S. L.; McDonald, S. A. C.; Elia, G.; Bhandari, P.; Poller, D.; Harrison, R.; Novelli, M. R.; Jankowski, J. A.; Wright, N. A., Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. *Gut* **2008,** 57, (8), 1041-1048.

29. Little, T. J.; Nelson, L.; Hupp, T. R., Adaptive Evolution of a Stress Response Protein. *PLoS ONE* **2007,** 2, (10), e1003.

30. Fourtouna, A.; Murray, E.; Nicholson, J.; Maslon, M. M.; Pang, L.; Dryden, D. T. F.; Hupp, T. R., The Anterior Gradient-2 Pathway as a Model for Developing Peptide-Aptamer Anti-Cancer Drug Leads that Stimulate p53 Function. *Curr. Chem. Bio.* **2009,** 3, (2), 124-137.

31. Murray, E.; Clarke, D. J.; Florance, H. V.; MacKay, C. L.; Langridge-Smith, P.; Hupp, T. R., Proteasomal-dependent Degradation of Anterior Gradient-2 is Regulated by an Oxidant-Responsive Dimerisation Motif. *Mol. Cell Proteomics* **2009,** Submitted.

# Chapter 6. Mass Spectrometry & Ion Mobility-Mass Spectrometry Studies of Factor H

*This Chapter investigates two distinct modules of the complement control protein, Factor H. Collision cross-sections for the modules have been obtained using two ion mobility instruments – the MoQTOF and a Synapt HDMS mass spectrometer. The first Factor H module, region 10-15, has been interrogated and cross-sections elucidated for the protein containing up to six covalently-bound N-acetylglucosamine sugar molecules. Cross-sections for the second module, the C-terminal 19-20 region, have also been obtained. This module may be an important polyanion binding site for the entire factor H molecule. Characterising the structure of these two regions by IM-MS may further information to the overall structure of the factor H protein.*

## 6.1 Factor H Introduction

Factor H (fH) is a large complement control protein which circulates in plasma and directs the immune system to act on pathogens rather than damaging host tissue. The main role of this soluble glycoprotein is to regulate the Alternative Pathway of the complement system, a biochemical cascade that removes pathogens from an organism, and can operate without antibody participation.[1, 2] fH (155 kDa, 1213 amino acid residues) is composed of twenty complement control protein modules (1 to 20) joined in a "string-of-beads" arrangement by nineteen short, potentially flexible, linking sequences of three to eight residues.[3] Variations in polyanion binding sites within the protein are connected to a number of complement-mediated diseases including atypical haemolytic uremic syndrome[4] and age-related macular degeneration.[5] Mapping of these polyanion sites have been rigorously investigated by Schmidt *et al.*[3] using ($^{1}$H, $^{15}$N) heteronuclear single quantum coherence nuclear magnetic resonance imaging of $^{15}$N-labelled proteins. In this technique, a spectral cross-peak is expected for each backbone amide group; for small folded proteins, discrete and well dispersed cross-peaks are expected. Due to the size of the entire protein, discrete modules are being produced by recombinant technology and analysed separately to build up a picture of the entire molecule. The results presented in this Chapter are the first mass spectrometry and ion mobility-mass spectrometry investigations of any region of the protein.

This Chapter presents results for two regions of complement control protein modules: region 10-15 and region 19-20 (the C-terminal domain) denoted fH 10-15 and fH 19-20 respectively. The solution structure for fH 19-20 has been determined by NMR experiments.[6] Binding site location and affinity of glycosaminoglycans (GAGs), sulfated heparin-derived tetrasaccharides, have also been investigated for many of the complement control protein regions, including fH 10-15 and fH 19-20.[3] Understanding the structure of the protein and the conformation(s) it adopts to bind GAGs and C3b (a cleavage product of complement component 3) are fundamental to learning more of the mechanisms

Factor H employs in its regulatory role. Ion mobility combined with mass spectrometry lends itself well to such needs.

## 6.2 Factor H 10-15 and 19-20 Experimental

Factor H regions 10-15 and 19-20 were sprayed under native-like conditions from 20 mM ammonium acetate containing 10% by volume isopropanol (pH 6.0) at a concentration of ~25 μM. Two ion mobility-mass spectrometers were used to characterise the proteins by mass spectrometry and ion mobility-mass spectrometry: the MoQTOF and a Synapt HDMS instrument.

### 6.2.1 MoQTOF Instrument Settings

Factor H 10-15 solution was sprayed using nano-ESI with a capillary voltage of 1.85 kV; source temperature 80$^{o}$C; sample cone 140V; extractor cone 115V; source hexapole pressure 1.94 x $10^{-1}$ mbar; pusher period 108 μs; and pulse width 40 μs. Factor H 19-20 solution was sprayed using nano-ESI with the same settings as fH 10-15 other than: capillary voltage of 1.94 kV; and extractor cone 122V. The cell pressure for both experiments was ~3.2 Torr; cell temperature was ~302 K. Nine drift voltages (60V, 50V, 40V, 35V, 30V, 25V, 20V, 15V and 10V) were measured for both proteins.

### 6.2.2 Synapt HDMS Instrument Settings

Synapt HDMS instrument settings are presented in Chapter 2. A Triversa Nanomate (Advion, Ithaca, NY, U.S.A.) was used to produce electrosprayed ions from 20 mM ammonium acetate containing 10% by volume isopropanol (pH 6.0). Protein concentrations were ~25 μM.

## 6.3 Factor H 10-15 Results

Results have been set out in two sections for each protein: the first compares the mass spectrometry data obtained on the two difference instruments; the second compares the ion mobility-mass spectrometry data. Comparisons are made from preliminary data between the two instruments used with respect to variations in collision cross-sections.

### 6.3.1 Factor H 10-15 – Synapt HDMS Mass Spectrometry

Under the solvent conditions, a wide charge state range was observed in the mass spectrum (Figure 6.1) indicative of a more unfolded structure than would be expected spraying from buffer.

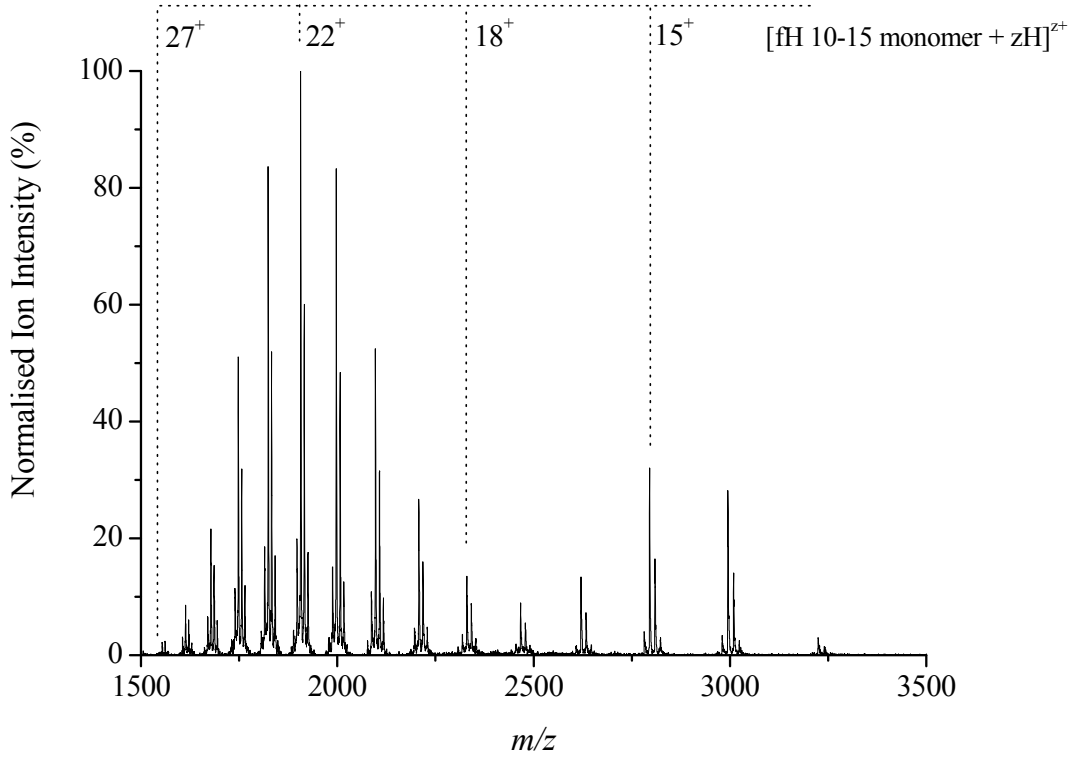



*Figure 6.1. Mass spectrum of native-like fH 10-15 sprayed from 20 mM ammonium acetate and 10% by volume isopropanol (pH 6.0) on a Synapt HDMS instrument.*

The wide charge state distribution ranges from z = $13^+$ to $27^+$ with a bimodal distribution centred on z = $15^+$ and $22^+$. For each charge state, a series of peaks indicated two to six covalently-bound N-acetylglucosamine (GlcNAc) sugar molecules ($M_R$ = 220 Da) with peaks separated by 203 *m/z* (Figure 6.2). The discrepancy between the mass of the sugar molecule and the peak separation (17 Da) is due to water lost in the condensation reaction required to covalently attach the sugar to asparagine residues. These are artefacts from the expression method. Within the mass spectrum, the most abundant species is attributed to four sugar molecules bound.

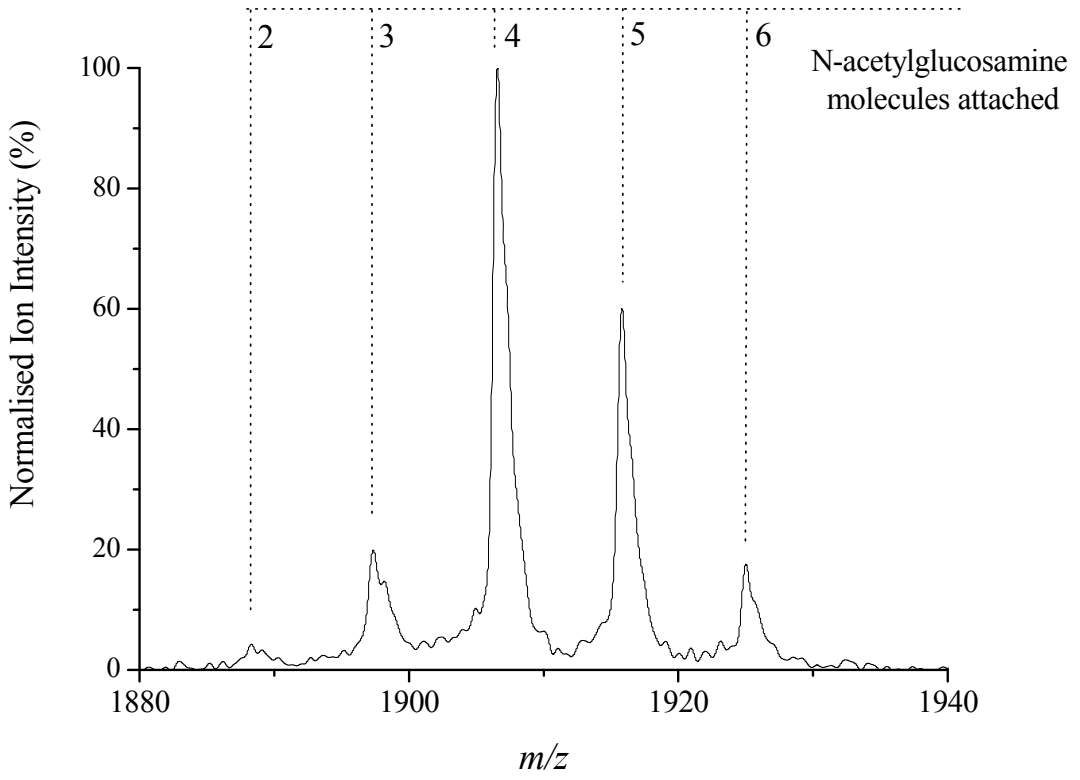



*Figure 6.2. fH 10-15 $22^+$ charge state showing two to six covalently-bound N-acetylglucosamine sugar molecules.*

### 6.3.2 Factor H 10-15 – MoQTOF Mass Spectrometry

fH 10-15 was sprayed under the same solvent conditions (pH 6.0) as those on the Synapt and is shown in Figure 6.3. A similarly wide charge state distribution was observed but shifted to z = $12^+$ to $26^+$. A trimodal distribution with peaks with centres at z = $14^+$, $17^+$ and $21^+$ was also observed. The relative abundance of the lower charge states indicates a more native-like protein providing further experimental evidence that the MoQTOF is able to successfully transfer relatively large (~ 41 kDa) molecular complexes into the gas phase. Two to five bound GlcNAc sugar molecules were observed with four sugar bound molecules still the most abundant. A peak for six sugar bound molecules was poorly resolved, but present.

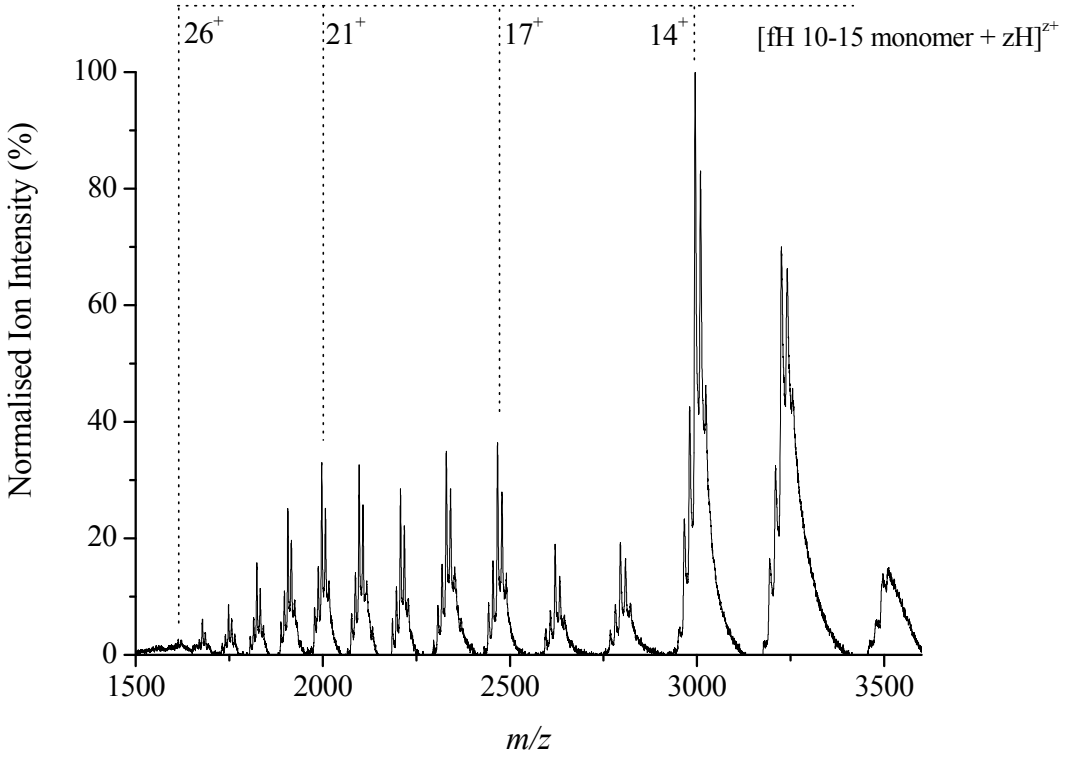



*Figure 6.3. Mass spectrum of native-like fH 10-15 sprayed from 20 mM ammonium acetate and 10% by volume isopropanol (pH 6.0) on the MoQTOF instrument.*

### 6.3.3 Factor H 10-15 – Synapt HDMS Ion Mobility-Mass Spectrometry

Collision cross-sections for four fH 10-15 species were elucidated and are presented in Table 6.1. Three to six GlcNAc bound sugar molecules are observed on each charge state with four bound sugars dominating. At low charge states, cross-sections for three, four and five bound sugars to the protein are either the same of within 5% of each other indicating structural similarity. As charge increases, collision cross-section increases and a large rate of cross-section change is observed between $z = 18^+$ to $22^+$ for all species indicating domain unfolding. This is observed most with the four sugar bound species with an increase of 1876Å$^2$ from $z = 18^+$ (2615Å$^2$) to $z = 22^+$ (4491Å$^2$), and further charge sequestering at $z = 22^+$ to $26^+$ shows a small increase in cross-section.

*Table 6.1. Synapt HDMS collision cross-sections for three to six N-acetylsglucosamine bound to fH 10-15 over the charge state range z = 13$^+$ to 26$^+$. Values are an average of three replicates and stated with twice the standard deviation of the mean (2$\sigma$). Species not observed indicated by n.o.*

| Charge state | fH 10-15 +3 GlcNAc (Å$^2$) | fH 10-15 +4 GlcNAc (Å$^2$) | fH 10-15 +5 GlcNAc (Å$^2$) | fH 10-15 +6 GlcNAc (Å$^2$) |
|---|---|---|---|---|
| 13 | n.o. | 2183 ± 24 | 2183 ± 15 | n.o. |
| 14 | 2196 ± 10 | 2196 ± 22 | 2196 ± 19 | 2225 ± 24 |
| 15 | 2301 ± 18 | 2202 ± 19 | 2202 ± 36 | 2236 ± 41 |
| 16 | 2397 ± 11 | 2323 ± 21 | 2342 ± 12 | 2396 ± 17 |
| 17 | 2481 ± 14 | 2460 ± 34 | 2460 ± 11 | 2519 ± 19 |
| | 3032 ± 11 | 2889 ± 37 | 2922 ± 19 | 2938 ± 26 |
| | 3713 ± 17 | 3713 ± 12 | 3639 ± 24 | 3664 ± 31 |
| 18 | 3155 ± 19 | 2615 ± 16 | 2636 ± 28 | 2718 ± 36 |
| | 3941 ± 13 | 3169 ± 13 | 3200 ± 27 | 3348 ± 37 |
| 19 | 3544 ± 19 | 3443 ± 17 | 3494 ± 28 | 3510 ± 38 |
| | 3974 ± 11 | 3988 ± 19 | 4072 ± 23 | 4030 ± 41 |
| 20 | 3669 ± 24 | 3841 ± 39 | 3857 ± 32 | 3858 ± 42 |
| | 4130 ± 31 | 4340 ± 47 | 4310 ± 34 | 4339 ± 19 |
| 21 | 3861 ± 25 | 3879 ± 31 | 3916 ± 32 | 4162 ± 20 |
| | 4377 ± 17 | 4425 ± 30 | 4439 ± 18 | n.o. |
| 22 | 4474 ± 18 | 4491 ± 41 | 4457 ± 31 | 4508 ± 45 |
| 23 | 4519 ± 41 | 4556 ± 45 | 4556 ± 14 | 4575 ± 48 |
| 24 | 4643 ± 58 | 4623 ± 38 | 4623 ± 18 | 4541 ± 55 |
| 25 | 4629 ± 28 | 4651 ± 36 | 4628 ± 48 | 4759 ± 53 |
| 26 | 4775 ± 20 | 4775 ± 31 | 4751 ± 45 | 4774 ± 56 |

### 6.3.4 Factor H 10-15 – MoQTOF Ion Mobility-Mass Spectrometry

MoQTOF data for fH 10-15 provided a large number of cross sections due to the presence of covalently-bound GlcNAc sugar molecules. Cross-sections for two, three, four and five bound sugar molecules were acquired over a charge state range of z = 13$^+$ to 25$^+$ (ion mobility signal for z = 12$^+$ and 26$^+$ was too weak to obtain cross-sectional data) with the trend in increasing cross-section increasing with charge (Table 6.2).

*Table 6.2. MoQTOF collision cross-sections (Å$^2$) for two to five N-acetylsglucosamine bound to fH 10-15 over the charge state range z = 13$^+$ to 26$^+$. Values are an average of three replicates and stated with twice the standard deviation of the mean (2$\sigma$). Species not observed indicated by n.o.*

| Charge state | fH 10-15 +2 GlcNAc (Å$^2$) | fH 10-15 +3 GlcNAc (Å$^2$) | fH 10-15 +4 GlcNAc (Å$^2$) | fH 10-15 +5 GlcNAc (Å$^2$) |
|---|---|---|---|---|
| 13 | 2570 ± 28 | 2597 ± 21 | 2600 ± 17 | 2608 ± 10 |
| 14 | 2546 ± 22 | 2559 ± 24 | 2538 ± 18 | 2577 ± 14 |
| 15 | 2820 ± 28 | 2761 ± 35 | 2722 ± 13 | 2738 ± 17 |
| 16 | 2970 ± 21 | 2821 ± 36 | 2778 ± 24 | 2793 ± 17 |
| 17 | 3005 ± 29 | 2972 ± 24 | 2853 ± 25 | 2741 ± 18 |
| 18 | 3105 ± 22 | 3028 ± 29 | 2992 ± 21 | 3132 ± 25 |
| 19 | 3194 ± 30 | 3387 ± 22 | 3319 ± 20 | 3308 ± 29 |
| 20 | 3430 ± 34 | 3573 ± 21 | 3547 ± 34 | 3827 ± 31 |
| 21 | 3577 ± 44 | 3660 ± 38 | 3919 ± 42 | 3550 ± 38 |
| 22 | 3564 ± 46 | 3750 ± 42 | 4157 ± 48 | 3945 ± 39 |
| 23 | 3675 ± 42 | 3895 ± 48 | 4035 ± 45 | 3881 ± 42 |
| 24 | 3973 ± 53 | 4154 ± 53 | 4065 ± 49 | 4122 ± 53 |
| 25 | n.o. | 4616 ± 54 | 4102 ± 51 | 4331 ± 54 |

At low charge states, z = 13$^+$ and 14$^+$, cross-sections for all four species are within two standard deviations of the mean cross-section at that charge state (z = 13$^+$ average cross-section = 2594.1 Å$^2$ ± 33.3Å$^2$; z = 14$^+$ average cross-section = 2555.4Å$^2$ ± 33.8Å$^2$). This infers a similar structure for all species and suggests that the covalently-bound sugar molecules are possibly folded back against the protein surface (as opposed to projecting away from the protein surface – it is assumed this would increase cross-section to a greater extent) and interacting with available hydrogen bonds and electrostatic regions to reduce their exposed surface area. Similarities in cross-section cease as charge increases above z = 15$^+$. fH 10-15 cross-sections with two bound GlcNAc molecules had the smallest cross-sections overall, whereas four and five bound sugar molecules to fH 10-15 showed greater increase in cross-section at higher charge. Over the entire charge state range, three bound GlcNAc showed the greatest increase in cross-section from 2597.6Å$^2$ at z = 13$^+$, to 4616.2Å$^2$ at z = 25$^+$ ($\Delta_{\text{cross-section}}$ = 2018.6Å$^2$). These observations may be explained by the negative sugar moiety having sufficient regions to interact with the protein at low charge states when well folded and compact, but as unfolding is induced by charge addition, the sugar molecules have fewer favourable interactions with the protein and add to the cross-section as they project away from the backbone into space. Between z = 17$^+$ and 20$^+$, all four species exhibit a possible unfolding event as cross-sections increase at a greater rate than elsewhere (425.0, 600.8, 694.5 and 1086.7Å$^2$ for two, three, four and five bound GlcNAc fH 10-15 respectively). The largest of these increases, for the five bound sugar fH 10-15

molecule, may indicate that a number of the covalently-bound sugars are no longer interacting closely with the protein as these non-specific binding sites have been lost through protein unfolding, with bond angles and distances for favourable interactions now out of range.

### 6.3.5 Comparison of Synapt HDMS and MoQTOF Ion Mobiltiy Data for Factor H 10-15

Preliminary Synapt fH 10-15 collision cross-section values are more compact at low charge states and more extended at high charge states than MoQTOF data. Experimental results to date have provided more compact cross-sections for all species with the MoQTOF instrument (Figure 6.4). Comparing Synapt and MoQTOF mass spectra of fH 10-15, Synapt spectra indicate a more denatured protein (dominant charge state at $z = 22^{+}$).

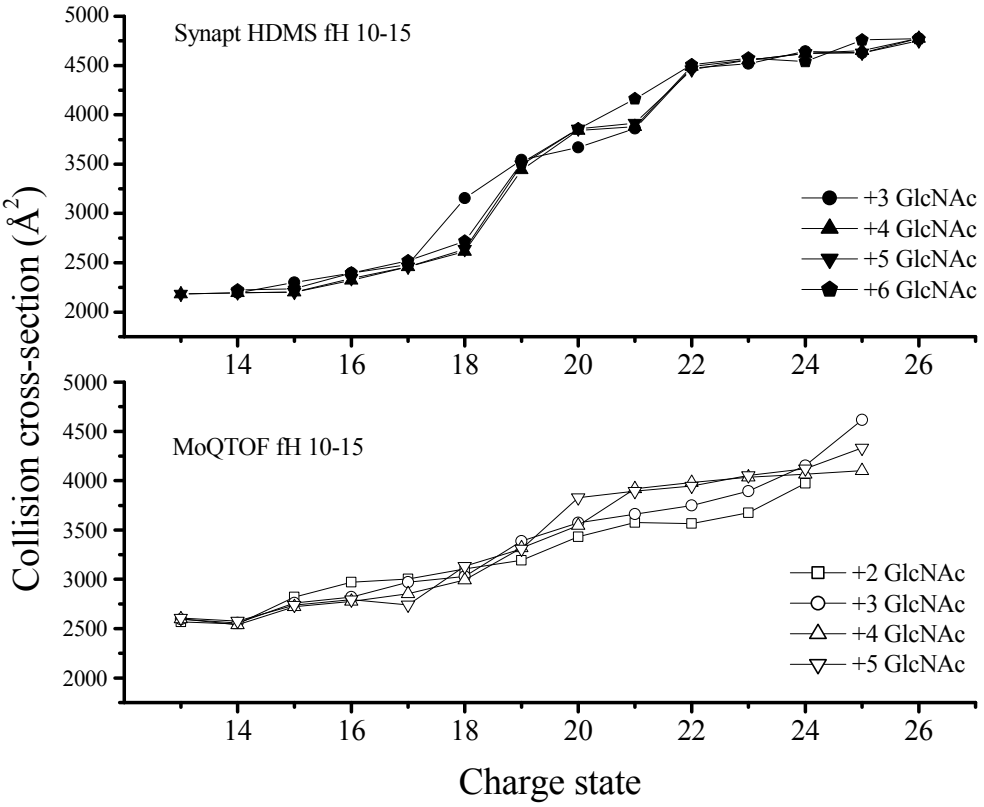



*Figure 6.4. Factor H 10-15 cross-section plots for data acquired on the Synapt HDMS and MoQTOF instruments showing the change in cross-section with charge state and with the number of sugar bound molecules*

Denatured proteins in solution interact with the solvent to a greater extent than native proteins, with residues projecting into the solvent to maximise available interactions. Upon transfer to the gas phase, desolvation occurs and the denatured protein collapses and electrostatic interactions, rather than hydrophilic or hydrophobic interactions, dominate the final geometry adopted. This may be the case

for the formation of more compact geometries at the low charge states of fH 10-15 (Synapt data) as the less constrained denatured protein geometries in solution collapse in the gas phase.

## 6.4 Factor H module 19-20 Results

### 6.4.1 Factor H module 19-20 – Synapt HDMS Mass Spectrometry

Four charge states were observed from z = $7^+$ to $10^+$ with the $8^+$ charge state dominant as shown in Figure 6.5. This narrow charge state distribution and relatively low charge density indicates a native-like solution phase structure transferred to the gas phase.

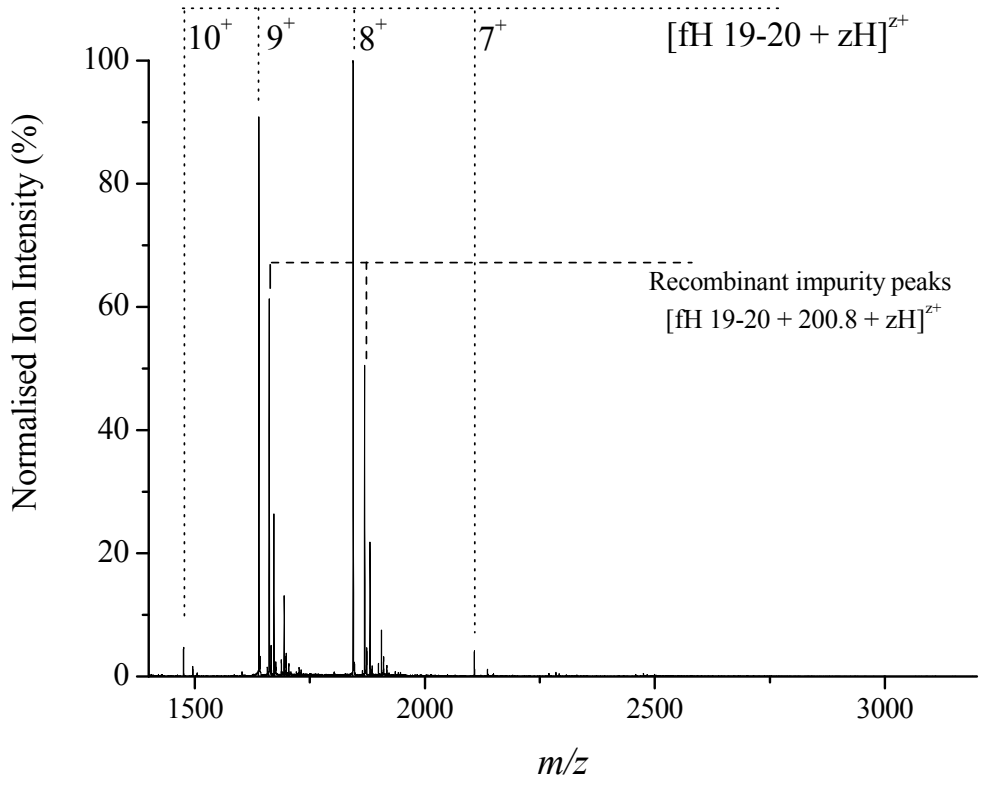



*Figure 6.5. Mass spectrum of native-like fH 19-20 sprayed from 20 mM ammonium acetate and 10% by volume isopropanol (pH 6.0) on the Synapt HDMS instrument.*

During the protein expression method, two additional residues (alanine; $M_R$ 71.08 Da; and glutamic acid; $M_R$ 129.10) have been covalently incorporated into the protein population as a recombinant impurity. This is observed as a peak at 200.8 *m/z* higher than the native protein peak for all four charge states.

### 6.4.2 Factor H 19-20 – MoQTOF Mass Spectrometry

fH 19-20 was sprayed on the MoQTOF from the same solvent conditions as the Synapt and produced five monomer charge states ($z = 6^+$ to $10^+$) as shown in Figure 6.6. The relatively narrow charge state distribution centred on the most abundant peak, $z = 9^+$, again indicates an overall folded protein structure. Upon interrogating the mass spectrum above 2000 *m/z*, a plethora of dimeric species can be elucidated. Due to the recombinant impurity of two additional amino acid residues (E and A), three possible dimer species may be produced:

i) Dimer with both monomers not containing the recombinant impurity;
ii) Dimer with one monomer subunit having no impurity and one monomer containing the impurity;
iii) Dimer with both monomers containing the impurity.

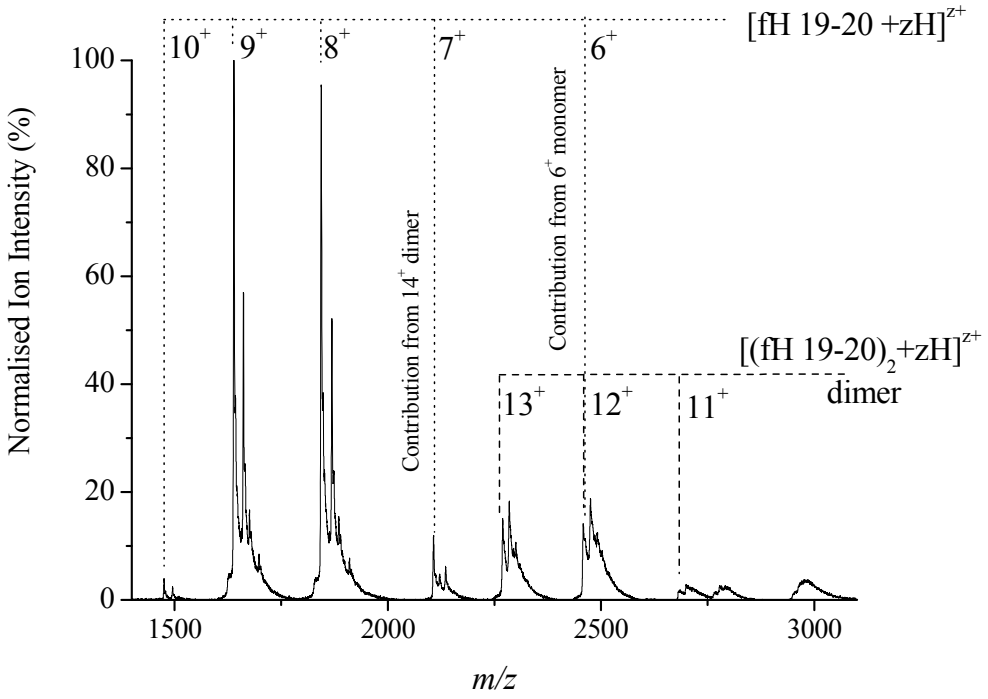



*Figure 6.6. Mass spectrum of native-like fH 19-20 sprayed from 20 mM ammonium acetate and 10% by volume isopropanol (pH 6.0) on the MoQTOF instrument.*

All three dimer species were observed for $z = 13^+$ (2269.3, 2284.7 and 2300.0 *m/z* respectively). From the intensity of the surrounding species, dimer peaks have been assigned with confidence with $z = 11^+$ to $14^+$. Evenly charged dimer *m/z* values are concomitant with those of a monomer with half the charge, so care has been taken when assigning monomers with $z = 6^+$ and $7^+$ and their respective dimers with $z = 12^+$ and $14^+$.

### 6.4.3 Comparison of Synapt HDMS & MoQTOF Mass Spectrometry Data for Factor H 19-20

Both instruments successfully transferred native-like fH 19-20 into the gas phase as shown by the narrow charge state distributions. Protein with and without the recombinant error were clearly observed on all peaks. MoQTOF data showed a considerably greater number of gas-phase species including all three dimer forms as well as poorly resolved higher *m/z* species such as the $5^+$ monomer.

### 6.4.4 Factor H 19-20 – Synapt HDMS Ion Mobility-Mass Spectrometry

Collision cross-sections for four charge states, z = $7^+$ to $10^+$, of fH 19-20 were calculated with a single conformer characterised for z = $7^+$ and $10^+$ and two conformers characterised for z = $8^+$ and $9^+$ as shown in Table 6.3. Cross-section increases near-linearly with charge with an overall increase from $7^+$ to $10^+$ of 22%, indicative of an initially well-structured protein undergoing minor structural rearrangement. The two conformers of charge states $8^+$ and $9^+$ are of similar collision cross-section (1729.6Å$^2$ and 1743.6Å$^2$ respectively) and may have adopted similar gas-phase structures. Using a wave gradient of 8-10V, the z = $13^+$ dimer containing no recombinant impurity was elucidated with a cross-section of 2322.1Å$^2$.

*Table 6.3. Factor H 19-20 collision cross-sections obtained using the Synapt HDMS instrument. Values are an average of three replicates and stated with twice the standard deviation of the mean (2$\sigma$). Species not observed indicated by n.o.*

| Charge state | Synapt collision cross-section (Å$^2$) conformer 1 | Synapt collision cross-section (Å$^2$) conformer 2 |
|---|---|---|
| 7 | 1328.1 ± 28.5 | n.o. |
| 8 | 1443.4 ± 31.2 | 1729.6 ± 49.3 |
| 9 | 1556.2 ± 29.5 | 1743.6 ± 42.1 |
| 10 | 1699.3 ± 42.1 | n.o. |

### 6.4.5 Factor H 19-20 – MoQTOF Ion Mobility-Mass Spectrometry

Cross-sections for five monomer charge states, z = $6^+$ to $10^+$, and four dimer charge states, z = $11^+$ to $14^+$, could be elucidated using the MoQTOF instrument (Table 6.4). Two conformers for most monomer and dimer charge states were elucidated. Monomer cross-sections ranged from 1227.2Å$^2$ for the smallest conformer of z = $6^+$, to 1904.4Å$^2$ for the largest conformer with z = $10^+$.

*Table 6.4. Collision cross-sections for monomeric and dimer species of factor H 19-20 obtained used the MoQTOF instrument. Multiple values at the same charge state indicate another conformer was detected. Values are an average of three replicates and stated with twice the standard deviation of the mean (2σ). Species not observed indicated by n.o.*

| Charge state | Monomer cross-section ($Å^2$) | Dimer no impurity cross-section ($Å^2$) | Dimer 1 EA impurity cross-section ($Å^2$) | Dimer 2 EA impurity cross-section ($Å^2$) |
|---|---|---|---|---|
| 6 | 1227.2 ± 29.2<br>1343.4 ± 21.2 | n.o. | n.o. | n.o. |
| 7 | 1214.8 ± 19.4<br>1295.2 ± 18.6 | n.o. | n.o. | n.o. |
| 8 | 1328.5 ± 23.3<br>1479.4 ± 29.3 | n.o. | n.o. | n.o. |
| 9 | 1364.7 ± 21.5<br>1545.5 ± 13.2 | n.o. | n.o. | n.o. |
| 10 | 1667.2 ± 31.4<br>1904.4 ± 19.3 | n.o. | n.o. | n.o. |
| 11 | n.o. | 1933.2 ± 24.3<br>2240.8 ± 34.2 | n.o. | n.o. |
| 12 | n.o. | n.o. | 2198.2 ± 38.1<br>2442.3 ± 25.4 | n.o. |
| 13 | n.o. | 2279.4 ± 39.2<br>2449.6 ± 41.3 | 2314.6 ± 36.1<br>2562.8 ± 39.2 | 2304.5 ± 34.1<br>2457.2 ± 33.1 |
| 14 | n.o. | n.o. | 2414.1 ± 31.2<br>2606.0 ± 42.3 | n.o. |

Cross-section values obtained from the MoQTOF are smaller than obtained from the Synapt but there is much greater agreement in values. Cross-sections for $z = 7^+$ to $10^+$ for Synapt and MoQTOF are between 0.7 and 2.5% for one conformer at each of these charge states. As discussed in the fH 19-20 mass spectrometry section, three dimeric species were observed. For $z = 11^+$, only the impurity-free dimer (*i.e.* both monomer subunits do not have additional glutamic acid or alanine residues present) was interrogated producing cross-sections of 1933.2$Å^2$ and 2240.8$Å^2$ as further dimer species could not be resolved.

Figure 6.7 presents raw experimental data for the $z = 11^+$ dimer (no impurity). Arrival times of 3048.0 µs and 3397.2 µs for the two conformers, with cross-sections 1933.2$Å^2$ and 2240.8$Å^2$ respectively, have been determined. Raw data presented was collected at a drift voltage of 20V.

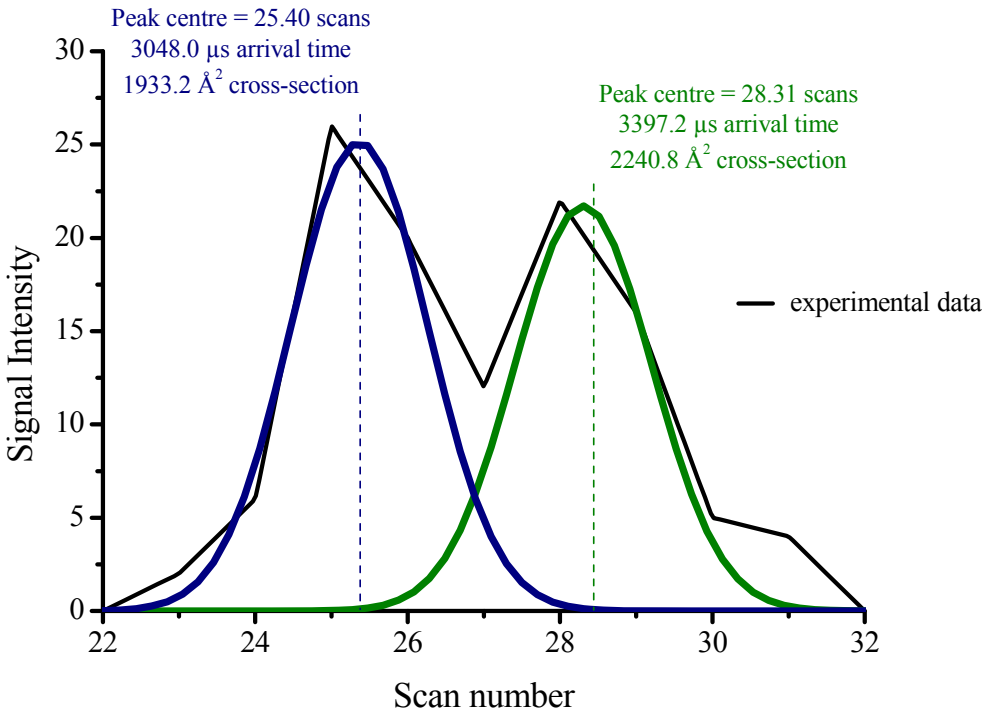



*Figure 6.7. Raw experimental data for the z = 11$^+$ dimer (no impurity) showing two conformers with cross-sections 1933.2Å$^2$ and 2240.8Å$^2$ respectively.*

For z = 12$^+$ and 14$^+$, the dimer containing one impurity monomer could be unambiguously identified (2475.0 and 2121.5 *m/z* for z = 12$^+$ and 14$^+$ respectively) in the ion mobility-mass spectrum and two conformers for each charge state were elucidated. Using Origin 7.5 curve fitting software, Gaussian peaks were fitted to the arrival time distributions of each drift voltage (within a charge state) to extract overlapping dimer and monomer information. This method, along with observing the peak intensity in the mass spectrum, allowed cross-sections to be calculated for evenly-charged monomer and dimer species. Finally, the z = 13$^+$ charge state provided full elucidation of all three dimeric species as the *m/z* values did not interfere with any other species. Figure 6.8 presents the cross-sections for all dimer species as plot of charge state.

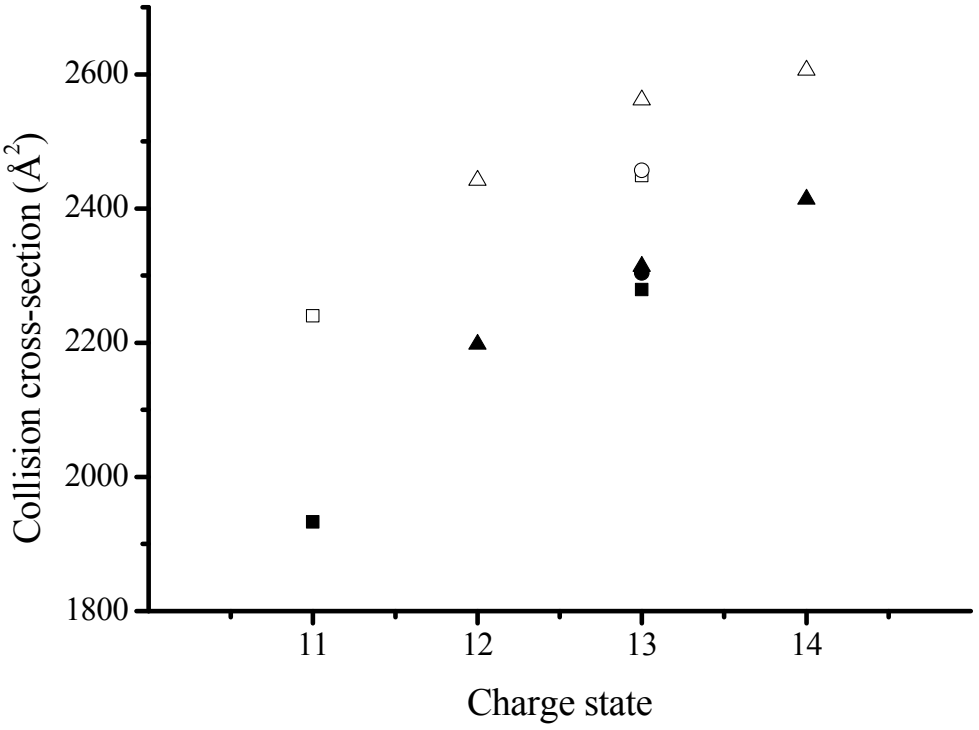



*Figure 6.8. Cross-section plot of factor H 19-20 dimeric species. Square shapes show the dimer species containing no recombinant impurity (z = 11$^+$ and 13$^+$). Triangular shapes show the dimer species that have one monomer containing the recombinant impurity (z = 12$^+$, 13$^+$ and 14$^+$). Circular shapes show the dimer species that have both monomers containing the recombinant impurity (z = 13$^+$). All species show two gas-phase conformations for each charge state.*

Protein dimer species free from recombinant impurities have the most compact geometry of the three dimeric species, with one conformer having a cross-section (2449.6Å$^2$) similar to the addition of the two smaller monomer conformers of 6$^+$ and 7$^+$ (2442.0Å$^2$). The other dimer conformer (2279.4Å$^2$) cross-section is smaller than mere addition of two monomer subunits and may provide evidence for folding and structural arrangement. Overall change in cross-section upon incorporating one or two monomer subunits with the recombinant impurity is not more than 4.4% - a small contribution to cross-section as would be expected for addition of only two residues to a protein chain.

### 6.4.6 Comparison of Synapt HDMS & MoQTOF Ion Mobility Data for Factor H 19-20

Cross-sections were calculated for four monomer charge states using Synapt data; two having a single conformer and two having two conformers. No dimer cross-sections could be produced at the 8V wave height used, however using a wave height gradient of 8V to 10V produced a single conformer for dimer with z = 13$^+$. MoQTOF data allowed a larger number of gas-phase species to be analysed and cross-sections elucidated for five monomer charge states with two conformations for each. Figure

6.9 is a comparison plot for monomer fH 19-20 species cross-sections obtained on the Synapt and MoQTOF instruments.

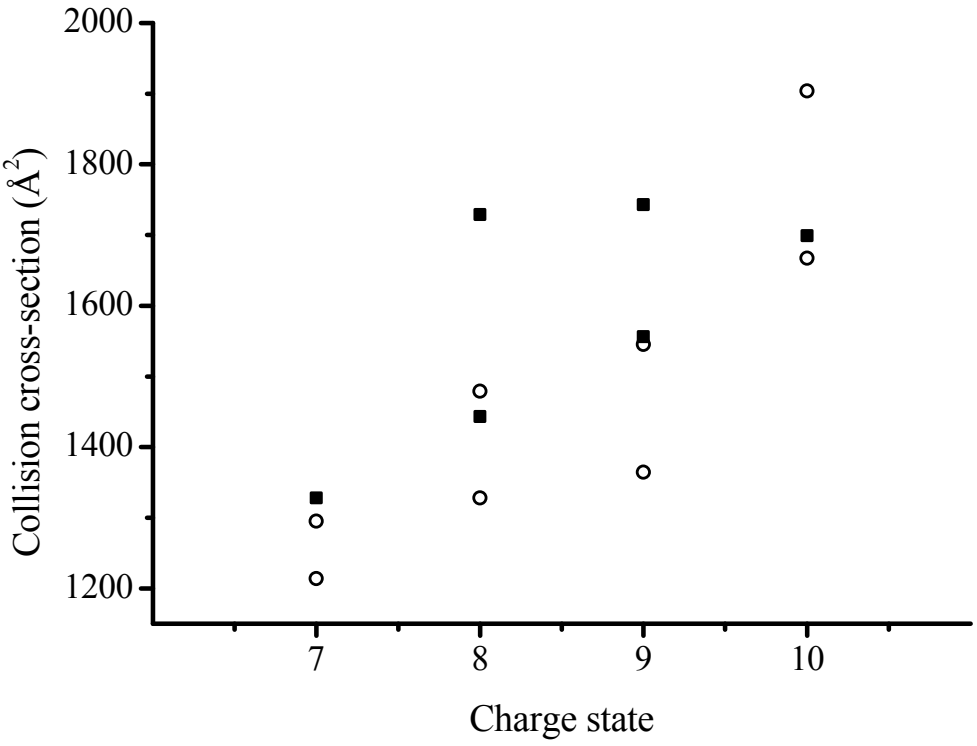



*Figure 6.9. Comparison cross-section plot for factor H 19-20 monomeric species between Synapt HDMS (closed squares) and MoQTOF (open circles) instruments over the charge state range z = $7^+$ to $10^+$.*

From the plot, it is clear that Synapt cross-section values are overall larger for most charge states with a number of close agreements (z = $8^+$, $9^+$ and $10^+$). Smaller cross-sections from MoQTOF data are attributed to better desolvation within the source region, but may also come from gas-phase dissociation of the dimer species, however this would be to a less extent as a greater range of monomer charge states would be expected.

## 6.5 Computational Comparisons for Factor H 19-20

Published structures are available from the Research Collaboratory for Structural Bioinformatics (RCSB)[7] for module fH 19-20: NMR structure (PDB ID: 2BZM[6]) and XRC structure (PDB ID: 2G7I[8]). No structures to date exist for module fH 10-15. Using MOBCAL,[9, 10] computational methods can provide theoretical collision cross-sections using NMR structure or XRC atomic coordinates which can then be compared to experimental data. For the NMR structure of fH 19-20, a cross-section of 1924Å$^2$ was calculated using the exact hard-sphere scattering method and a value of 1907Å$^2$ was

calculated using the trajectory method. These values are in good agreement with one of the $z = 10^+$ monomer conformers (1904 $Å^2$). Native protein transfer from solution to gas phase is likely to result in gas-phase collapse due to desolvation, causing residue side chains that were solvated to collapse back against the protein, thereby reducing the collision cross-section. Cross-sections calculated using the Trajectory methods are frequently slightly smaller than those calculated using the EHSS method, and has been reported by others.[11, 12]

For the XRC structure of fH 19-20, a cross-section of 1823$Å^2$ was calculated using the exact hard-sphere scattering method and a value of 1801$Å^2$ was calculated using the trajectory method. These values lie between the $z = 9^+$ and $10^+$ collision cross-sections showing that condensed-phase data can compare well with gas-phase compact sturtcures.

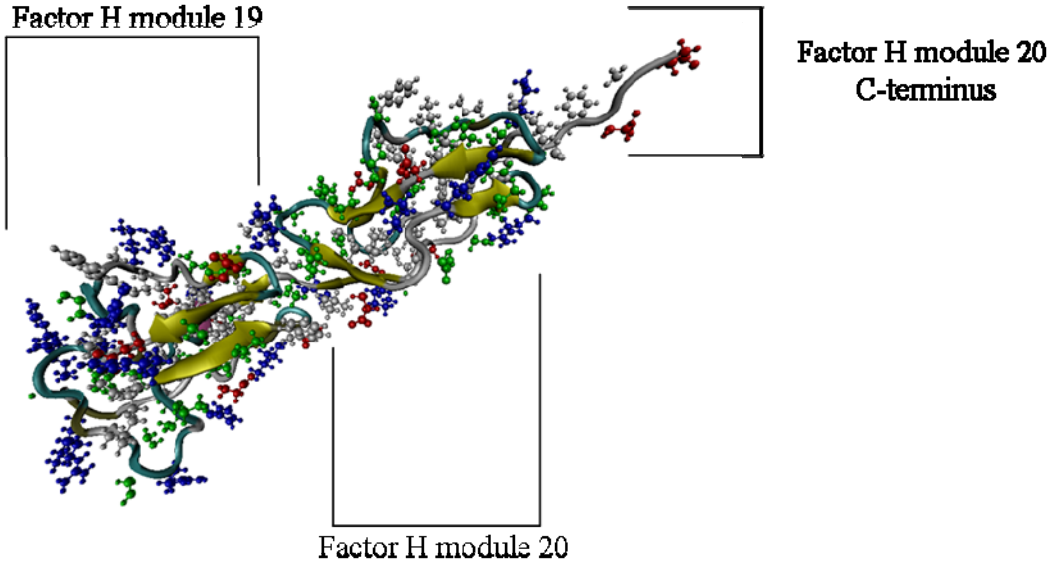



*Figure 6.10. X-ray crystal structure 2G7I of factor H 19-20 modified and to show the folding of the two consecutive modules. The sheet-like structure of the folded backbone is shown by the yellow arrows; amino acid residues are shown (aromatic – white; aliphatic – green; acidic – red; basic – blue) using VMD software.*[13]

## 6.6 Factor H Conclusions

fH 10-15 mass spectra differed with the Synapt base peak at $z = 22^+$ compared with $z = 14^+$ on the MoQTOF. Current evidence suggests native-like protein structure displays a narrow charge state distribution with a base peak of low charge state, indicating a more denatured protein obtained on the Synapt instrument. A covalent artefact of the expression system provided interesting multiple peaks at each charge state with three to six molecules of N-acetylglucosamine sugar bound. Subsequent addition of sugar molecules did not significantly affect the protein cross-section due to their small size compared to the large protein size. MoQTOF mass spectra displayed the same covalent artefacts however two to five molecules of the sugar were resolved. Cross-sections for $z \geq 20^+$ were more compact from MoQTOF data, however low charge states ($z = 12^+$ to $19^+$) from Synapt data were more compact. This may be explained by considering the denatured protein in solution (as shown by the Synapt mass spectrometry data) being less conformationally constrained and, when transferred to the gas phase, undergoes greater hydrophobic collapse to produce more compact low charge state geometries.

Finally, the C-terminal region of factor H (19-20) was investigated. Mass spectrometry data from the Synapt HDMS instrument provided evidence for a native-like protein with a narrow charge state of low charge (base peak $z = 8^+$). MoQTOF data provided the same charge state distribution, but with contributions from gas-phase dimer species which were resolved using enhanced Gaussian fittings to arrival time distributions. A recombinant error in the protein expression method produced a population of protein molecules with an additional two amino acid residues (alanine and glutamic acid) which could be clearly observed in the mass spectrum. These additional peaks proved useful in characterising monomer and dimer contributions at concomitant *m/z* values. Cross-sections for monomer subunits were in good agreement (less than 2.5%) between Synapt and MoQTOF data, and all conformers (bar the larger $z = 10^+$ conformer) were more compact than the computationally-derived cross-section from the NMR structure. Dimer cross-sections could not be elucidated from Synapt data without the use of a gradient ramp, however up to three dimer species cross-sections could be extracted at charge states $z = 11^+$ to $14^+$ from MoQTOF data.

The work presented in this Chapter shows the potential for mass spectrometry in studying regions of much larger molecules. Further complementary NMR and MS work may provide information on the entire Factor H molecule (modules 1 to 20) by "stitching" together results from the studies of constitutive sections, until such time as technology allows interrogation of the entire molecule.

## 6.6 Chapter 6. References


1. Walport, M. J., Complement. First of two parts. *N. Engl. J. Med.* **2001,** 344, (14), 1058-66.

2. Walport, M. J., Complement. Second of two parts. *N. Engl. J. Med.* **2001,** 344, (15), 1140-4.

3. Schmidt, C. Q.; Herbert, A. P.; Kavanagh, D.; Gandy, C.; Fenton, C. J.; Blaum, B. S.; Lyon, M.; Uhrin, D.; Barlow, P. N., A New Map of Glycosaminoglycan and C3b Binding Sites on Factor H. *J. Immunol.* **2008,** 181, (4), 2610-2619.

4. Buddles, M. R.; Donne, R. L.; Richards, A.; Goodship, J.; Goodship, T. H., Complement factor H gene mutation associated with autosomal recessive atypical hemolytic uremic syndrome. *Am. J. Hum. Genet.* **2000,** 66, (5), 1721-2.

5. Hageman, G. S.; Anderson, D. H.; Johnson, L. V.; Hancox, L. S.; Taiber, A. J.; Hardisty, L. I.; Hageman, J. L.; Stockman, H. A.; Borchardt, J. D.; Gehrs, K. M.; Smith, R. J.; Silvestri, G.; Russell, S. R.; Klaver, C. C.; Barbazetto, I.; Chang, S.; Yannuzzi, L. A.; Barile, G. R.; Merriam, J. C.; Smith, R. T.; Olsh, A. K.; Bergeron, J.; Zernant, J.; Merriam, J. E.; Gold, B.; Dean, M.; Allikmets, R., A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** 102, (20), 7227-32.

6. Herbert, A. P.; Uhrin, D.; Lyon, M.; Pangburn, M. K.; Barlow, P. N., Disease-associated Sequence Variations Congregate in a Polyanion Recognition Patch on Human Factor H Revealed in Three-dimensional Structure. *J. Biol. Chem.* **2006,** 281, (24), 16512-16520.

7. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucl. Acids Res.* **2000,** 28, (1), 235-242.

8. Jokiranta, T. S.; Jaakola, V. P.; Lehtinen, M. J.; Parepalo, M.; Meri, S.; Goldman, A., Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *Embo. J.* **2006,** 25, (8), 1784-94.

9. Shvartsburg, A. A.; Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996,** 261, (1-2), 86-91.

10. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996,** 100, (40), 16082-16086.

11. Jarrold, M. F., Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Acc. Chem. Res.* **1999,** 32, (4), 360-367.

12. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H., Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **2008,** 22, (20), 3297-304.

13. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Molec. Graphics* **1996,** 14, (1), 33-38.

# Chapter 7. Conclusions to Exploring Gas-Phase Protein Conformations by Ion Mobility-Mass Spectrometry

Mass spectrometry and ion mobility-mass spectrometry studies of a range of peptides and proteins have been characterised, predominantly using an in-house modified instrument, the MoQTOF, but also with the only currently commercially available ion mobility-mass spectrometer, the Synapt HDMS. Changes in solvent composition and pH; source pressure; ligand binding and gas-phase dissociation mechanisms have been investigated for molecules that have molecular masses in the range ~2 kDa to ~64 kDa, and have shown that the MoQTOF instrument is capable of producing cross-sections that compare well with the Synapt HDMS instrument. Collision cross-sections for a number of ionic species over a wide charge state range provide evidence for protein unfolding in the gas-phase to adopt more extended conformations. For a number of species of the same charge state, more than one gas-phase conformer was elucidated.

In Chapter 3, three different species of cytochrome *c* were characterised: monomers (~12.4 kDa), dimers (~24.8 kDa) and trimers (~37.2 kDa). Compact monomer species were produced from native-like and aqueous solvent conditions at low concentration. Extended monomer cross-sections were produced as a result of protein unfolding upon addition of organic solvent and acid. Increasing protein concentration under aqueous conditions produced higher order multimeric species of dimers and trimers. Cross-sections for these species were obtained and reported for the first time. These species were stable in the gas-phase, with relatively compact structures showing structural arrangement of the monomer subunits had occurred. Cross-section trends of the dimer species were compared with known gas-phase dimer dissociation behaviour where asymmetric and symmetric dissociation occurs depending upon the charge of the dimer species. Comparisons with computationally-derived cross-sections indicated low charge state conformations in the gas-phase compare well with condensed-phase calculated cross-sections. A number of fragments produced by native electron capture dissociation, a novel gas-phase dimer dissociation technique, were characterised and their cross-sections elucidated, revealing that fragments formed are from cleavages adjacent to the haem ring. Unfolding of the monomer subunits and subsequent fragment is still not well understood, but the cross-sections obtained complement the proposed unfolding and fragmentation mechanism.

In Chapter 4, data presenting the successful transfer and characterisation of the ~64 kDa tetrameric protein haemoglobin was shown. At elevated pH levels, the tetramer complex dominated the mass spectra, showing little dissociation and compact gas-phase structures that were smaller than calculated cross-sections from condensed-phase structures. This may be attributed to gas-phase collapse of the protein in the absence of bulk solvent. Two dimer species were characterised and collision cross-sections obtained. The dimer species are heterodimers and differ by the presence of either two or one haem ligand. Cross-sections for both species are similar and show low charge density, indicating an

ordered structure. The presence of one of the dimers, the so-called pseudo-holo heterodimer missing one haem ligand, has been proposed as a crucial intermediate in the assembly process of the tetramer *in vivo*. This work shows that this intermediate is in a compact structured form capable of binding the final haem ligand, and then would be able to associate to the intact tetramer. Upon decreasing the solvent composition to denaturing conditions, the constituent monomer subunits could be studied and compared to show that in the absence of the haem ligand, both adopt extended conformations at high charge state. The α monomer in the presence of the haem ligand displayed a low charge density, indicating a structured conformation that may enable the more unstructured $\beta_{apo}$ monomer subunit to use it as a template so it can adopt a structure capable of haem ligand binding. Comparisons with calculated cross-sections show that low charge state gas-phase species have cross-sections comparable with condensed-phase structures.

In Chapter 5, two proteins that are linked *in vivo* have been characterised in the presence and absence of small binding molecules. The first protein, the DNA-binding core domain of p53 (~24.6 kDa), contains a zinc ion that is thought to stabilise the DNA-binding region. Cross-sections were obtained on both the MoQTOF and Synapt HDMS instrument in the presence and absence of zinc (removed by the chelating agent phenanthroline) and showed gas-phase collapse of the structure in the absence of zinc. Cross-sections compared between instruments showed the MoQTOF values were consistently more compact. This is potentially attributed to a combination of increased desolvation in the source region of the MoQTOF, and to minor protein unfolding in the trap region of the Synapt HDMS instrument. Chapter 5 also presented work on anterior gradient-2 protein (~21 kDa) which downregulates p53 through potential cellular relocalisation. If AGR2 was inhibited from doing so, thereby allowing upregulation of p53 and possible tumour suppression, then this pathway could provide novel anticancer properties. The protein formed higher order multimers under native-like solvent conditions of monomer, dimer and trimer species. Cross-sections for all species were obtained and again showed relatively compact structural arrangement. Peptide aptamers intended to target AGR2 to inhibit its downregulation of p53 were characterised by MS and IM-MS for the first time, and cross-sections were obtained. Upon addition of one aptamer having the bioactive hexapeptide core PTTIYY to the AGR2 protein, two complexes were observed of the 1:1 (protein to peptide) and 2:1 stoichiometries. Cross-sections were obtained for these species.

In Chapter 6, two regions of the large complement control protein Factor H were interrogated by MS and IM-MS using the MoQTOF and Synapt HDMS instrument. Factor H region 10-15 (~40.9 kDa) was analysed and shown to contain a number of covalently-bound sugar molecules. Cross-sections were obtained for three to six bound sugar molecules on the Synapt, and two to five bound on the MoQTOF. Factor H region 19-20 (14.7 kDa) was also analysed and shown to contain a recombinant impurity of two additional amino acids. This proved to be of use as dimeric species were successfully transferred to the gas phase for analysis using the MoQTOF and the impurities allowed three distinct dimer species, of similar structure, to be characterised. Comparison with calculated cross-sections

from NMR and XRC data for Factor H 19-20 indicates that the gas-phase structure has a cross-section closer to that of the solution-phase structure.

## Future Work

The MoQTOF instrument can be considered an organic instrument, not in terms of bioanalytical chemistry performed on it, but in that it is evolving constantly due to the work being achieved using it now and the future direction the research group and the biological macromolecule community steer it. In less than five years, the instrument has changed from a traditional QTOF to an ion mobility instrument capable of interrogating molecules in excess of 60 kDa. This figure will climb ever higher as improvements are made to regions of the instrument.

Probing the gas-phase stability of the cytochrome *c* multimers using CID would reveal more about the structure and interactions involved in these higher order assemblies. Increased control of experimental conditions and data analyses may provide complete coverage of all dimer charge states (including even numbered charge states). The MoQTOF drift cell can now be cooled and heated reproducibly, therefore enabling a number of further experiments relating to temperature-dependent protein unfolding to be undertaken.

Gas-phase dissociation studies of haemoglobin could reveal ionic species not obtained in this Thesis, such as trimers or homodimers. Ion mobility separation of tetramer species, and then collision-induced dissociation may assist in revealing species that contribute to the assembly/disassembly mechanism, thereby providing further evidence to this as-yet undefined pathway. Further comparisons between drift cell and travelling wave ion mobilities need to be completed to further understand the differences, and similarities, between these two instruments and their effect on measuring protein structure.

IM-MS and MS studies of the two other p53 domains (the C- and N-terminal domains) could reveal more about the overall structure of the protein. Binding of other metals (such as cobalt) and comparing cross-sections may provide further information on the zinc binding region of the DNA-binding core domain. With respect to AGR2, the use of MS in ligand binding studies could be used to produce dissociation constants ($K_d$ values) to study the relative binding affinities of potential peptide aptamers that may, ultimately, provide a new anticancer therapy.

Finally, there are twelve further regions (regions 1 to 9 and 16 to 18) of the Factor H protein which have not been studied by MS or IM-MS as of yet. If interrogated by NMR, MS, IM-MS and *in silico* techniques, the entire 155 kDa protein could soon be characterised and its roles in self-recognition further understood.

## Outlook – Tomorrow's Protein Studies

It is hopefully clear from this Thesis that the past two decades have provided huge advances in studying biological molecules in the gas phase by mass spectrometry-based techniques. If similarly heroic advances are made in the next two decades, structural mass spectrometry will be an even more exciting and routine bioanalytical tool used in laboratories. The shifting focus to larger biological molecules and complex assemblies is being driven forward by a number of world class laboratories such as Professors Carol Robinson, Albert Heck, Lars Konermann and Igor Kaltashov to name a few field-leaders. The need to better understand and relate solution phase properties to these complexes in the gas phase and ultimately apply them to *in vivo* observations is still, potentially, a number of years away.

Mass spectrometer developments from the research groups of Professors David Clemmer, Michael Bowers and Michael Gross continue to provide better and more interesting techniques to characterise molecules. Relatively recent instrument developments, such as the orbitrap, are now common in many laboratories interesting in proteomic and genomic applications. These instrument development research group provide competition, as well as collaboration, with industrial companies such as Waters Micromass, Bruker Daltonics, Thermo Scientific and Agilent, producing new ideas and new instruments with which to study proteins.

# Appendix A

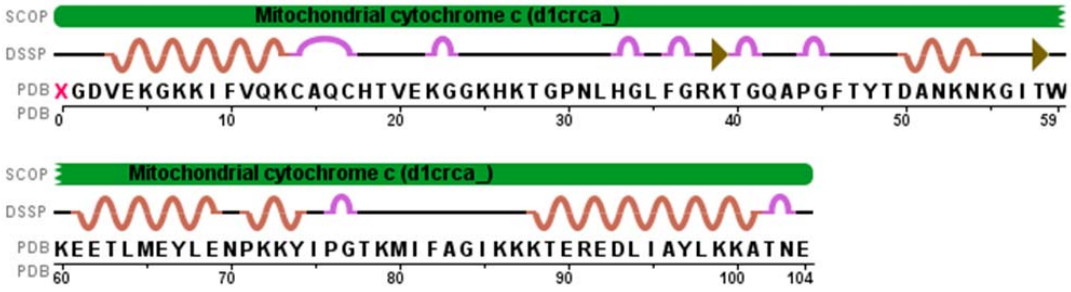
Table of amino acids showing their 3 and 1 letter code, monoisotopic and average masses.

| Amino Acid | 3 Letter code | 1 Letter Code | Monoisotopic Mass | Average Mass |
|---|---|---|---|---|
| Glycine | Gly | G | 54.021464 | 57.05 |
| Alanine | Ala | A | 71.037114 | 71.08 |
| Serine | Ser | S | 87.032029 | 87.08 |
| Proline | Pro | P | 97.052764 | 97.12 |
| Valine | Val | V | 99.068414 | 99.07 |
| Threonine | Thr | T | 101.04768 | 101.1 |
| Cysteine | Cys | C | 103.00919 | 103.1 |
| Leucine | Leu | L | 113.08406 | 113.2 |
| Isoleucine | Ile | I | 113.08406 | 113.2 |
| Asparagine | Asn | N | 114.04293 | 114.1 |
| Aspartic acid | Asp | D | 115.02694 | 115.1 |
| Glutamic acid | Gln | Q | 128.05858 | 128.1 |
| Lysine | Lys | K | 128.09496 | 128.2 |
| Glutamine | Glu | E | 129.0496 | 129.1 |
| Methionine | Met | M | 131.04048 | 131.2 |
| Histidine | His | H | 137.05891 | 137.1 |
| Phenylalanine | Phe | F | 147.06841 | 147.2 |
| Arginine | Arg | R | 156.10111 | 156.2 |
| Tyrosine | Tyr | Y | 163.06333 | 163.2 |
| Tryptophan | Trp | W | 186.07931 | 186.2 |

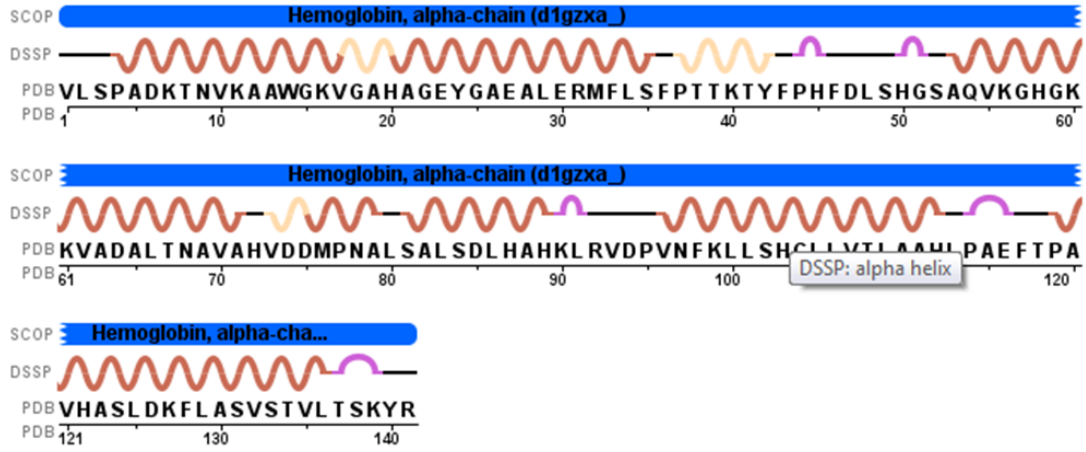
# Appendix B

The following appendix displays the amino acid composition of some proteins discussed in this Thesis. Basic secondary structure is also displayed above the amino acid code. The RCSB PDF file is provided after the protein name.

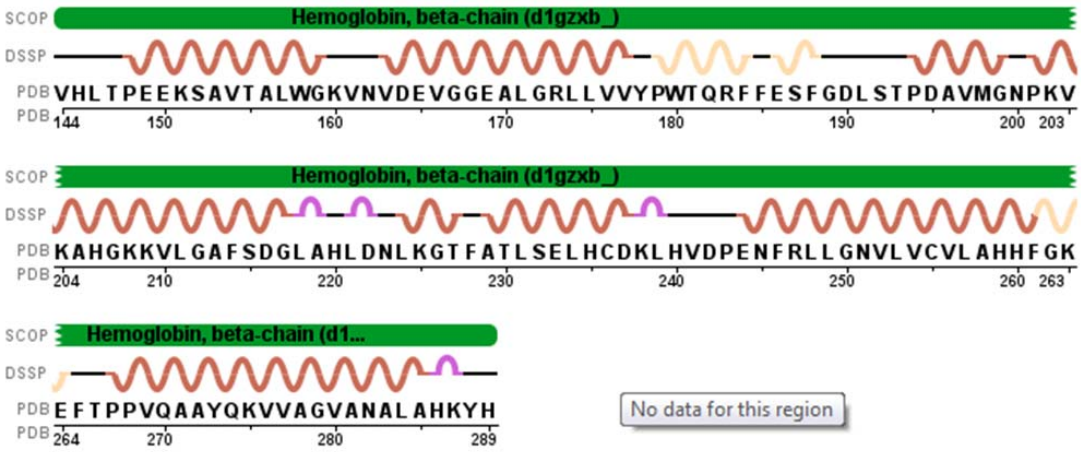
## B.1 Cytochrome *c* – 1crc



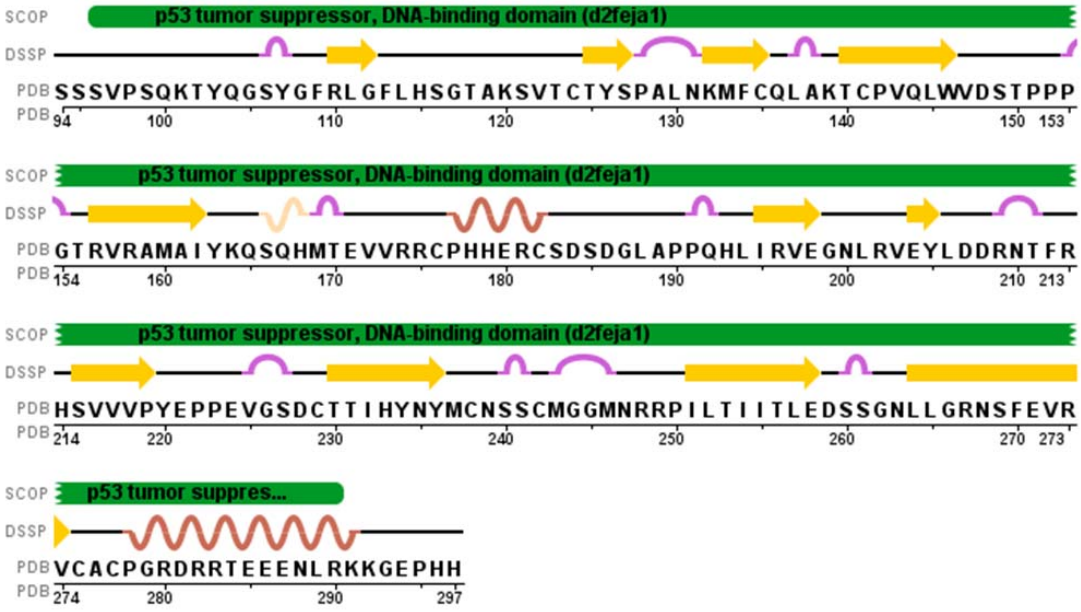

## B. 2 Haemoglobin chain A – 1gzxa



## B.3 Haemoglobin chain B – 1gzxb




## B.4 DNA-binding core domain of p53 – 2fej



## B.5 Factor H module 19-20 – 2g7i