Mass Spectrometry-Based Studies for the Investigation of Protein Structure and Dynamics

Matthew Edgeworth Bsc (Hons)

A thesis submitted for the degree of Doctor of Philosophy

University of Warwick

School of Life Sciences

November 2014

“I always think that the chances of finding out what really is going on are so absurdly remote that the only thing to do is to say hang the sense of it and just keep yourself occupied.” Douglas Adams, *The Hitchhikers Guide to the Galaxy*

Contents

Table of Figures

Acknowledgements

First and foremost I would like to thank my supervisors, professor Jim Scrivens and Dr Daniel Higazi for the opportunity to undertake this research project and for their support and encouragement over the last four years. I would also like to thank the University of Warwick, AstraZeneca and MedImmune for providing the funding for the project.

A page is hardly enough space to thank everyone who has helped me during the course of the last four years, but I would like to draw special attention to those that have had the biggest impact on me.

The Scrivens group has downsized during my time, with each group member leaving for a new and exciting challenge. Whilst they are no longer here with me, I will always be grateful for the contributions of Fran, Elle, Nisha, Krisztina, Andy and Jonathon for helping me in the early days, and making the lab such an enjoyable place to be. I would also like to thank Imtiaaz and Owen for their support during the final stages.

The ABC characterisation group at MedImmune, Cambridge, has been like a second home to me over the last two years, and I would like to thank everyone there for making me feel so welcome and included. I would like to give a special mention to Dr Jonothan Phillips. Without your guidance, and patience, I would not be the scientist I am today, and this thesis would be at least one chapter lighter!

Mum and Dad; you've provided me with more than just financial support, but also the encouragement I've needed to see this through. Without you I would not have made it this far, and I would not be the man I am today.

I have also been extremely fortunate to have had a number of close friends to help me get by, and so to all of you I say thank you, particularly Nicola, Mike, Sophie, Adam and Kevin for making my home life so fun and imaginative.

And finally I would like to thank Laura. You helped me through a particularly dark time of my PhD, and I feel that I am now emerging into a brighter future thanks to you.

Declaration

I hereby declare that this thesis, submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy and entitled "*Mass Spectrometry-Based Studies for the Investigation of Protein Structure and Dynamics*", represents my own work and has not been previously submitted to this or any other institution for any degree, diploma or other qualification. Work undertaken by my colleagues or collaborators is explicitly stated where appropriate.

Matthew Edgeworth

November 2014

Summary

The use of mass spectrometry to study of proteins and their non-covalent complexes has grown during the last few decades thanks to the introduction of soft ionisation techniques capable of preserving the weak molecular bonds present in higher order protein structure. The integration of the shape selective technique ion mobility with mass spectrometry has allowed the study of conformation and dynamics of native proteins. The work presented in this thesis focuses on the use of ion mobility mass spectrometry to investigate both protein conformation, and conformational differences induced by mutation.

Prolyl oligopeptidase family enzymes are proteases characterised by their unique function; these enzymes are only capable of digesting short amino acid sections of no more than thirty amino acids. These proteins have been implicated in a number of neurological disorders and have been targeted as potential drug candidates. PREP is an 80.7 kDa monomeric protein that has been targeted as a potential drug candidate. PREP has been crystallised in two distinct conformations, open and closed, and it has been suggested that the protein exists in equilibrium between the two states in solution, with only the open conformation allowing substrate entry via a domain separation mechanism. Ion mobility mass spectrometry has been used to confirm the presence of both conformations, and to investigate the effect of ligand binding on conformation. It was found that, at lower energy states, PREP was only capable of adopting a single conformation, and that more extended conformations were only present following activation of the protein. Binding of ligand appeared to increase the relative stability of the protein. DPP IV, is a larger, dimeric, protein from the same family. Unlike PREP, DPP IV has only been crystallised in a single conformation and it was proposed that small loop movements, rather than domain separation, allowed substrate entry. Ion mobility measurements show only a single conformation of DPP IV, consistent with no large conformational changes, supporting this hypothesis.

Haemoglobin is the main protein involved in gas transport in mammalian systems, taking oxygen from the lungs to the tissues of the body. Disorders of haemoglobin represent the most common of all inherited disorders, with an

estimated 7% of the global population being carriers for a haemoglobin disorder. A previous study by Scarff *et al* used ion mobility mass spectrometry to identify conformational differences between normal HbA and HbS, the haemoglobin mutant responsible for sickle cell disease. Recent experimental improvements in sample preparation, data collection and data processing have been used in this research to provide improved experimental information. HbA, HbS and HbC, a third haemoglobin variant known to form crystals within erythrocytes, were investigated using ion mobility mass spectrometry. Calibration of ion mobility data using native protein standards indicated that the structural differences between HbA and HbS were smaller than previously reported, and that the CCS measurements of the two proteins were similar for the native charge states. The HbC molecule does however adopt a smaller conformation. All three proteins unfold as a factor of increased protonation, with HbS and HbC showing evidence of adopting a more extended conformation at lower charge states when compared to HbA, suggesting possible differences in protein stability. These stability differences were investigated using collision-induced activation of the protein, the results suggesting that HbA is more resistant to unfolding that either HbS or HbC.

Monoclonal antibodies represent a new generation of biotherapeutic capable of high specificity and selectivity, with diminished risk of inducing a host immune response. Antibodies therapeutics have an added benefit of interacting with host cellular systems, promoting host immune response. Engineering of antibodies has become well established to improve or diminish antibody-receptor binding in order to increase or abolish this interaction. Four antibody Fc variants have been studied using a combination of ion mobility mass spectrometry and hydrogen deuterium exchange mass spectrometry; a wild type, a TM mutant, a YTE mutant and a TMYTE double mutant. Previous studies have shown that the introduction of both TM and YTE mutations leads to a decrease in the thermal stability of the protein, and it was an aim of the study to provide structural information to explain this thermal destabilisation. Ion mobility mass spectrometry measurements suggest that there is little change in global conformation between the four variants. HDX results show that mutation introduces changes in local conformation across the protein, with increases in

deuterium uptake observed at sites distant to the mutation sites. One region, located between the TM and YTE mutation sites, showed a greater than additive increase in deuterium uptake in TMYTE mutant compared with either TM or YTE mutants, indicating a region that hade become destabilised in the double mutant. These changes could be responsible for the observed loss of thermal stability.

Abbreviations

A

Ab	Antibodies
ADA	Anti-drug antibodies
ADC	Analogue-to-digital
ADCC	Antibody dependent cellular cytotoxicity
ADP	Antibody dependent phagocytosis
amu	accurate mass unit
ATD	Arrival time distribution

B

β	Probability factor of exchange

C

c	Enhanced duty cycle coefficient
C1	Complement protein 1
C1q	Complement protein component q
C1r	Complement protein component r
C1s	Complement protein component s
C-domain	Constant domain
C-terminal	Carboxy- terminal amino acid
CCS	Collision cross section
CD	Circular dichroism

CDC	Complement-dependent cytotoxicity
CDEM	Continuous dynode electron multiplier
CDR	Complimentarity determining regions
CEM	Chain ejection method
C_H1	Constant domain, heavy chain 1
C_H2	Constant domain, heavy chain 2
C_H3	Constant domain, heavy chain 3
CI	Chemical Ionisation
CID	Collision induced dissociation
C_L	Constant domain, light chain
CRM	Charge residue model

D

DC	Direct current
DCIMS	Drift cell ion mobility spectrometry
DE	Delayed Extraction
DPP IV	Dipeptidyl Peptidase IV
DSC	Density scanning caliometry

E

EDC	Enhanced duty cycle
EHSS	Exact hard sphere scattering
EI	Electron impact

EM	Electron multiplier
ESI	Electrospray isonisation
ETD	Electron transfer dissociation

F

Fab	Fragment antigen binding
FAIMS	High-field asymmetric waveform ion mobility spectrometry
Fc	Fragment crystallisable
FcγR I, II, III	Fc gamma receptor I, II and III
FcRn	Fc neonatal receptor
F_D	Folded protein, deuterated
F_H	Folded protein, protonated
FI	Field Ionisation
FR	Flanking regions
FTICR	Fourier transform ion cyclotron resonance

G

Gal	Galactose
GlcNAc	N-acetylglucosamine

H

Hb	Haemoglobin
HbA	Normal haemoglobin, $\alpha_2\beta_2$

HbA2	Variant Haemoglobin, $\alpha_2\delta_2$
HbC	Variant haemoglobin, $\alpha_2\beta^c{}_2$
HbS	Sickle cell haemoglobin, $\alpha_2\beta^s{}_2$
HDX	Hydrogen/ deuterium exchange
HAS	Human serum albumin

I

IEM	Ion evaporation mechanism
Ig	Immunoglobulin
IgA	Immunoglobulin α
IgD	Immunoglobulin δ
IgE	Immunoglobulin ε
IgG	Immunoglobulin γ
IgM	Immunoglobulin μ
IMS	Ion mobility spectrometry
IM-MS	Ion mobility mass spectrometry
IT	Ion trap

K

K	Kelvin
K_0	Reduced mobility
k_1	Kinetics of unfolding
k_{-1}	Kinetics of folding

k_2	Intrinsic chemical exchange for an unstructured peptide
k_b	Boltzmann's Constant
k_{ex}	Kinetics of exchange
k_f	Kinetics of exchange of the folded form
k_u	Kinetics of exchange of the unfolded form
k_{unf}	Constant used to describe the kinetics of unfolding

M

μ	Reduced mass of the ion-neutral pair
m	Mass
mAb	Monoclonal antibody
Man	Mannose
MALDI	Matrix assisted laser desorption ionisation
mBar	Millibar
MCP	Multichannel plate
MD	Molecular dynamics
m_g	Mass of buffer gas
μL	Microlitre
mL	Millilitre
mM	Millimolar
μM	Micromolar
MS	Mass spectrometry

ms	milliseconds
MS/MS	Tandem MS
m/z	mass-to-charge ratio

N

N	Background gas density
n	Scan number
NMR	Nuclear magnetic resonance
N-terminal	Amino- terminal amino acis

O

Ω	Rotationally averaged cross section

P

PA	Projection approximation
PDB	Protein data bank
PDMS	Plasma desorption mass spectrometry
POP	Prolyl Oligopeptidase
PREP	Prolyl Endopeptidase
PSA	Projected superposition approximation
PWHM	Peak width half maximum

Q

Q	Quadrupole

Qtof	Quadrupole - time of flight
QqQ	Triple quadrupole

R

RF	Radio Frequence

S

SAXS	Small angle scattering
SRIG	Stacked ring ion guide

T

T	Buffer gas temperature
t_d	Drift time
t'_d	Corrected drift time
TDC	Time-to-digital
TM	Trajectory method
TOF	Time of Flight
TOT-TOF	Tandem time of flight time of flight
TWIMS	Travelling wave ion mobility spectrometry

U

U_D	Unfolded protein, deuterated
U_H	Unfolded protein, protonated

V

V	Volt
V-domain	Variable domain
V_H	Variable domain, heavy chain
V_L	Variable domain, light tchain

Z

z	charge
ze	Ionic charge
ZPP	Z-pro-prolinal

Chapter 1 : Introduction

1.1 Mass spectrometry

1.1.1 Mass spectrometry

Mass spectrometry (MS) is an analytical technique that measures the mass-to-charge ratio (*m/z*) of gas phase ions. A typical mass spectrometry system shown in Figure 1-1 consists of an ionisation source for generating ions, a mass analyser to measure the *m/z* of these ions and a detector to register the number of ions at each *m/z* value, with a post-detection computer system which may be used to process and/or interpret the data (Cristoni and Bernardi 2003, Heck and van den Heuvel 2004). A number of inlet methods may be employed to introduce the sample into the system.

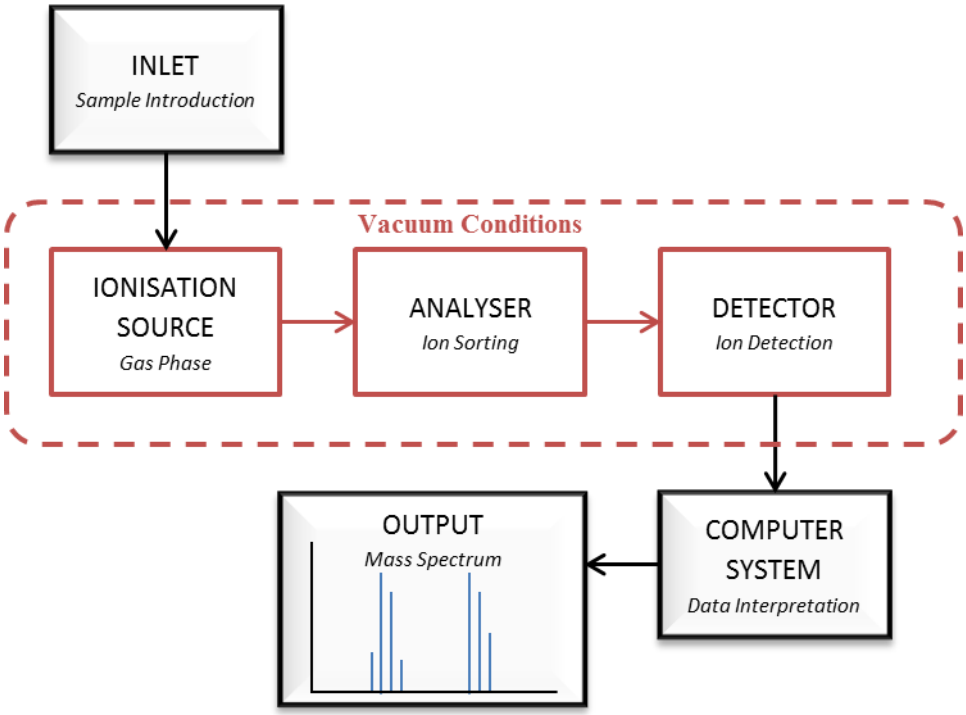


Figure 1-1 – Schematic of a mass spectrometer.

Ions produced in the source are analysed prior to MS detection. Information collected by the detector is processed by a computer system to generate a spectrum.

A number of possible components may be employed for each element, and by varying these components it is possible to construct a mass spectrometry system suitable for a wide range of analytical applications.

1.1.2 Ionisation Sources

In the ionisation source the sample is ionised prior to introduction into the mass spectrometer. Many different ionisation methods are suitable for use with mass spectrometry. Some ionisation techniques are highly energetic and transfer a lot of

energy to the analyte, leading to fragmentation of the ion, whilst others are softer and leave the ion intact (de Hoffman and Stroobant 2007).

Initial MS instruments made use of ionisation sources employing electron ionisation (EI), chemical ionisation (CI) or field ionisation (FI). These methods required the analyte to be in the gas phase prior to analysis and thus were only suitable for volatile, thermally stable compounds. Due to these limitations early MS analysis was rarely used for biological applications (Feng, Liu *et al*. 2008)

Modern ionisation sources are capable of transferring an analyte into to the gas phase from either the liquid or solid phase as part of the ionisation process. The development of electrospray ionisation (ESI) (Wong, Meng *et al*. 1988, Fenn, Mann *et al*. 1989), a liquid-phase ionisation method, and matrix-assisted laser desorption ionisation (MALDI) (Karas, Bachmann *et al*. 1987, Karas and Hillenkamp 1988, Tanaka, Waki *et al*. 1988), a solid-phase method, revolutionised the MS field since these techniques were soft enough to allow the study of biomolecules by MS. Such has been the impact of these techniques that the Nobel Prize in Chemistry in 2002 was awarded to John Fenn and Koichi Tanaka for their roles in their development.

Electrospray Ionisation

Electrospray ionisation (ESI) was developed by John Fenn (Fenn, Mann *et al*. 1989). It has since become the principal technique for the study of biomolecules using MS since it is considered to be the softest of the ionisation techniques (leading to the least fragmentation of the analyte) and because, as a liquid phase technique, it is compatible with inlet systems incorporating liquid chromatography (Feng, Liu *et al*. 2008), enabling high throughput analysis of complex biological samples. In an ESI source, shown in Figure 1-2, an analyte in solution is ejected from a capillary upon the application of a strong electric field. A solution containing analyte molecules dissolved in a volatile buffer is driven into the capillary at a flow rate of 1-20 µL/min, depending on the inlet method (Wilm and Mann 1994, Banerjee and Mazumdar 2012). Application of a high voltage, typically 2-6 kV (Banerjee and Mazumdar 2012), applied to the capillary leads to the partial separation of positive and negative electrolyte ions in the solution. In positive ion mode, positive ions migrate to the surface of the liquid, whilst negative ions migrate to the centre of the capillary. Electrostatic repulsion of positive electrolytes, combined with the effect of

the electric field leads to an elongation of the solvent at the tip, generating a characteristic cone (Taylor cone). The tip of this cone, the least stable point, elongates further to form a liquid filament which breaks up to form the individual charged droplets containing the analyte (Kebarle 2000).

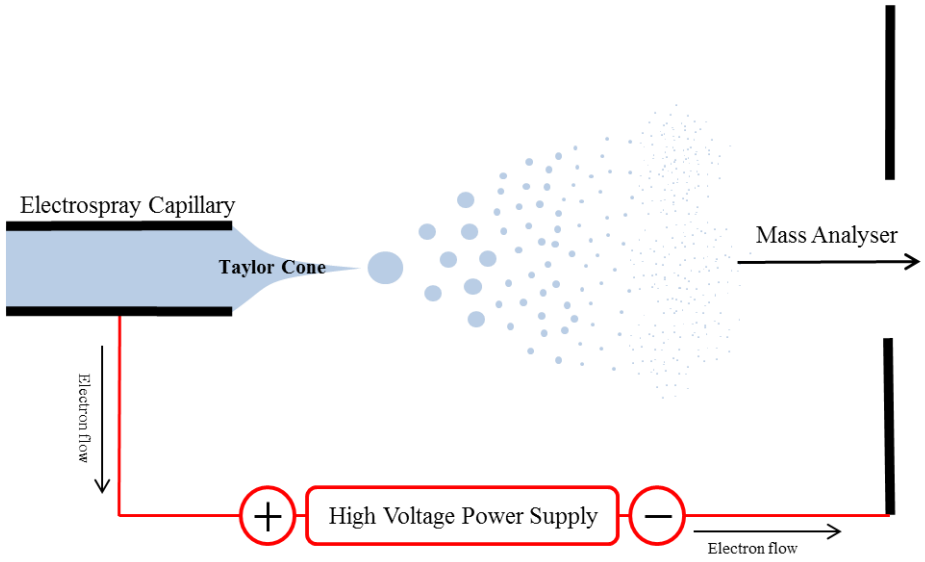


Figure 1-2 – Schematic of an ESI source.
The electric potential applied to the capillary causes the separation of the electrolytes in solution. Electrostatic repulsion causes the elongation of the liquid filament and the formation of liquid droplets.

ESI capillaries generate first generation droplets approximately 1-2 µm (Wilm and Mann 1996) in diameter, each of which may contain many thousands of analyte molecules (Wilm and Mann 1994). An exact mechanism describing how charged ions are produced from these droplets is still under debate, with three main mechanisms having been proposed. The ion evaporation method (IEM) (Iribarne and Thomson 1976) and the charge residue model (CRM) are the two most accepted mechanisms of ion formation., whilst the chain ejection method (CEM) is a more recently proposed mechanism (Konermann, Ahadi *et al*. 2012). These are summarised in Figure 1-3.

Within the source the charged droplets undergo solvent evaporation (Fenn, Mann *et al*. 1989). In order to assist with the evaporation of solvent ESI capillaries are heated to 100-300^{o}C (Banerjee and Mazumdar 2012).

The droplet diameter decreases with evaporation until the Rayleigh limit is reached. This is the point at which the coulombic repulsion of surface charges and the solvent

surface tension are equal. At this point the droplets undergo coulumbic fission into daughter droplets of smaller diameter and lower charge.

The IEM mechanism states that there is a certain threshold at which the charge required for coulombic fission of the droplet is lower than that required for direct evaporation of the charged ions from the droplet. At this point the droplets themselves remain whilst the charged ions evaporate from the droplet and become fully desolvated (Iribarne and Thomson 1976).

The CRM mechanism, in contrast, suggests that this limit does not exist, and indicates that successive coulomb fission and evaporation eventually leads to desolvated charged ions (Dole, Mack *et al*. 1968).

The CEM mechanism is a more recent proposal based on extensive computational molecular dynamics (MD) simulations. In this mechanism it is suggested as a droplet approaches the Rayleigh limit the ions within migrates to the surface of the droplet. One end of the chain is ejected from the droplet, followed by the sequential ejection of rest of the ion by electrostatic repulsion. Unlike the CRM and IEM, the CEM is not dependant on droplet size and therefore may occur in any droplet approaching the Rayleigh limit (Konermann, Ahadi *et al*. 2012).

It is generally accepted that no single mechanism is completely correct; rather that each mechanism is applicable to a different subset of molecules. The CRM approach best explains the ionisation of larger, globular molecules such as intact proteins (Heck and van den Heuvel 2004). MD simulations indicate that, within the charged droplet, excess charge migrates to the surface, whilst large, hydrophobic molecules cluster towards the centre, supporting the CRM mechanism for proteins. The IEM approach best explains the ionisation of smaller, hydrophilic ions, whilst the CEM approach best describes the ionisation of longer, disordered molecules such as polymer chains or unfolded peptides (Konermann, Ahadi *et al*. 2012).

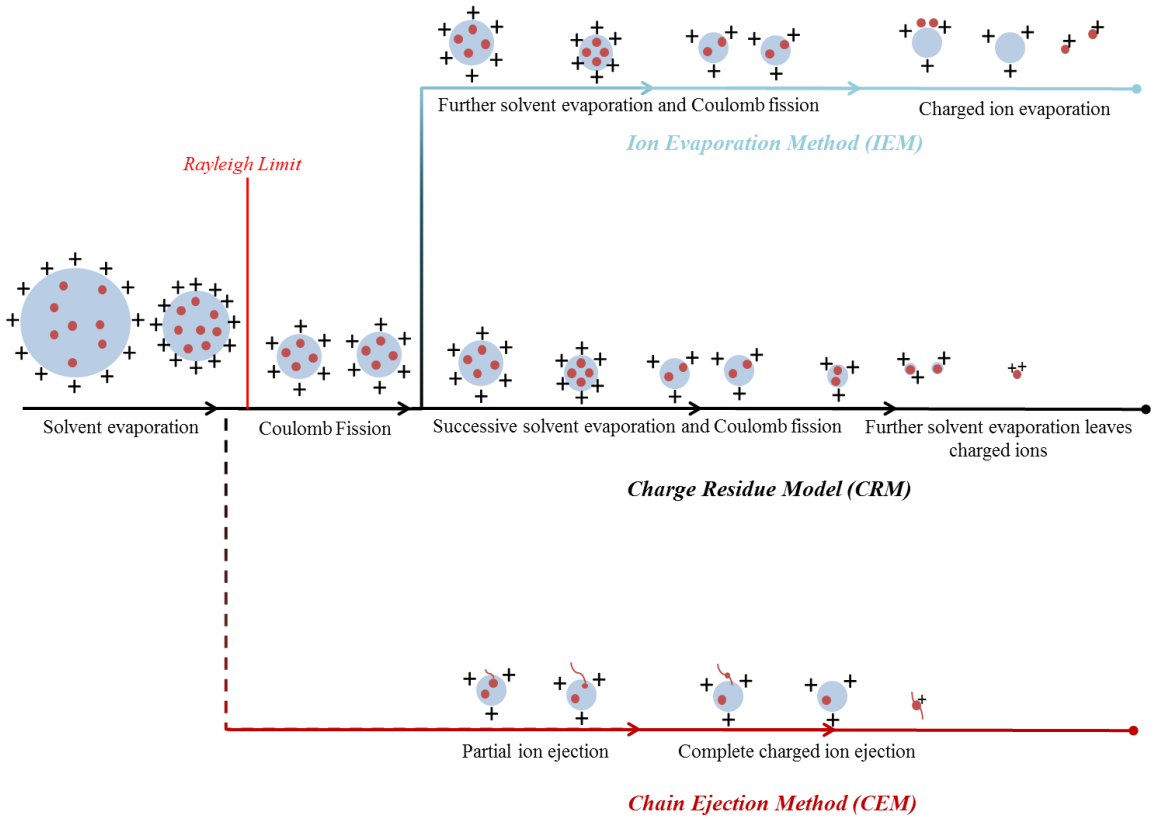


Figure 1-3 – Overview of the three proposed mechanisms of generating charged ions in the electrospray method.
In the IEM, the energy threshold for solvent evaporation eventually becomes higher than the ion evaporation energy leading to ion evaporation from the droplet. In the CRM, all the solvent evaporates from a droplet, leaving only a single ion. In the CEM the ion is ejected from the droplet in sequential steps along the length of the ion. The CEM method occurs as the droplet approaches the Rayleigh limit, and does not always require successive coulombic fission to occur.

Nanoelectrospray Ionisation

In 1996 Wilm and Mann first published data on nanoelectrospray ionisation (nESI) (Wilm and Mann 1996); a version of ESI scaled down to produce smaller droplets from the capillary using lower flow rates. nESI capillaries are far smaller than ESI capillaries, with an internal diameter of 1-4 µm and a tip of ~1 µM (Wilm and Mann 1996) . Droplets are produced from the capillary tip by the same mechanisms as for ESI, with the notable exception that the flow rates for a nESI source are far lower than that of ESI (1.2-30 µL.hr for nESI compared to up to 30 mL/ hour for ESI (Konijnenberg, Butterer *et al*. 2013). As a result of the lower flow rates and smaller capillary spray orifice, nESI droplet size is far smaller than those produced by ESI; nESI droplets are approximately 200 nm in diameter (based on theoretical calculations),100-1000 times smaller than the ESI droplets (Wilm and Mann 1996).

nESI has a number of benefits over conventional ESI. The smaller droplet size of nESI makes solvent evaporation far more favourable, reducing the need for high source temperature which can lead to the thermal decomposition of sample (Konijnenberg, Butterer *et al*. 2013). nESI is also more tolerant to non-volatile salts and buffers typically present in biological samples (Lorenzen and Duijn 2001). Importanly, nESI also requires much lower samples concentrations, and samples volumes (as little as 1 µL of sample at a typical concentration 5-10 µM), and is more amenable to less volatile solvents, such as water, ammonium acetate or ammonium bicarbonate, which makes it ideally suited for the study of biomolecules (Konijnenberg, Butterer *et al*. 2013).

1.1.3 Mass Analysers

Once ions have been produced they must be separated, based on their *m/z,* by a mass analyser. There are a number of potential mass analysers available for use in a mass spectrometer, each of which uses different methods to measure the *m/z* of an ion. All mass analysers use either static or dynamic electric and magnetic fields, alone or in combination, to achieve ion separation and can be divided into two classes: scanning mass analysers, which transmit ions of different masses successively over an extended timescale or simultaneous transmission analysers, which transmit all ions simultaneously whilst achieving separation.

Modern mass analysers include quadrupole (Q), ion trap (IT), time of flight (TOF), Orbitrap and Fourier transform ion cyclotron resonance (FTICR) based approaches. Although they are all capable of measuring *m/z*, different mass analysers have different performance characteristics with some analysers better suited to certain fields of research. Many modern instruments contain more than one mass analyser (hybrid instruments) allowing them to be used in conjunction to increase instrument versatility and provide an option for multiple experimental approaches.

Quadrupoles

The development of the quadrupole mass analyser dates back to the 1950's with initial work carried out by Paul and Steinwegen in 1953. Quadrupoles are in the class of scanning mass analysers. A quadrupole consists of four parallel circular rods. Opposed rods are paired electrically and a radio frequency (RF) voltage is applied between the two pairs. Superimposed over this RF field is a direct current (DC) voltage.

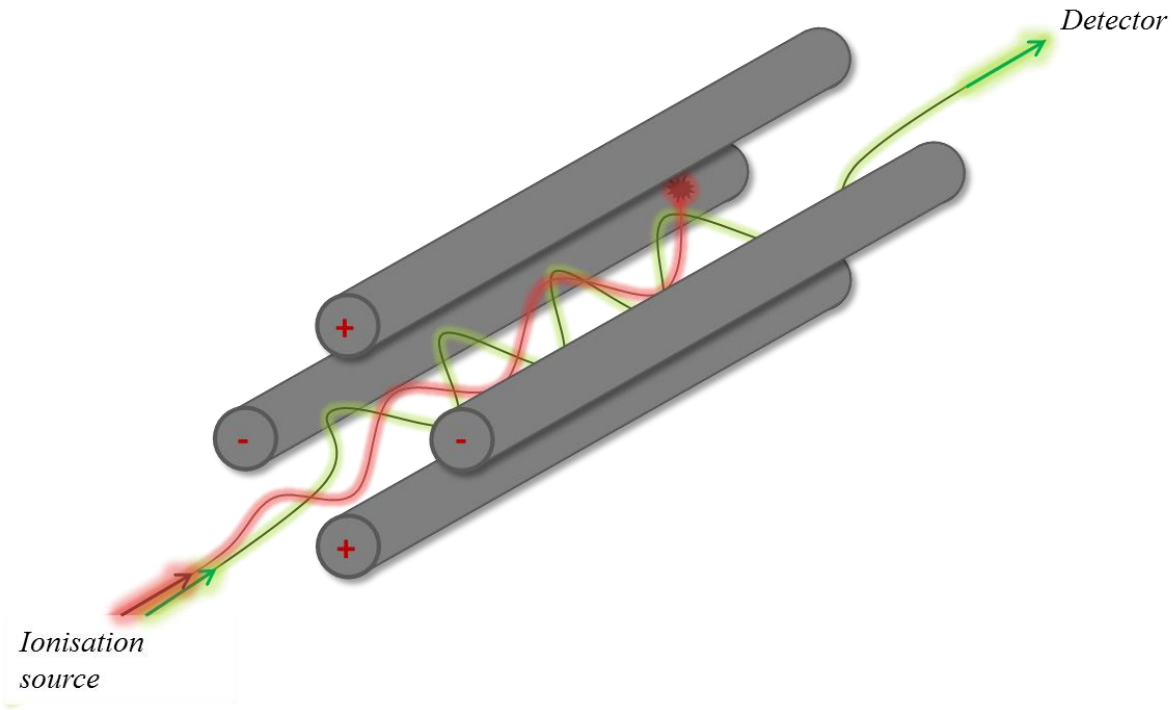


Figure 1-4 - Representation of a quadrupole mass analyser.
Opposed parallel rods are paired with a RF field of the same polarity, with a superimposed DC field. The relationship between teh RF and DC fileds acts as a *m/z* filter, with only one type of ion having a stable trajectory through the rods, shown by the green line. All other ions have an unstable trajectory through the poles and are lost, shown by the red line.

Based on the values on the RF and DC voltages applied, only certain *m/z* ions have a stable trajectory through the analyser, based on their oscillation between the poles. Ions with little oscillation pass directly through the analysers, whilst those with a stronger oscillation may collide with one of the poles and can be lost. This is shown schematically in Figure 1-4. Varying the DC and RF voltages changes which m/z ions have a stable trajectory allowing a quadruple to be utilised as a mass filter. Quadrupoles are typically operated in scanning mode, where DC and RF voltages are varied to allow all ions to pass sequentially through to the detector, but may also be used in a fixed mode, where only a single *m/z* species is allowed through to the detector (de Hoffman and Stroobant 2007)

Time of Flight

TOF mass analysers date back to the late 1940's, however the first publication describing a linear TOF-MS instrument with a resolution high enough to be practical was in 1953. Despite their high sensitivity and potentially unlimited *m/z* range all early TOF analysers suffered with problems with resolution, with a few hundred being the best resolution observable on these early instruments (Mamyrin 2001).

TOF analysers separate ions based on their velocities through a free-field region called a flight tube. As ions enter the analyser they are accelerated by a potential difference towards the flight tube. All ions receive the same kinetic energy, and therefore separation occurs based on Equation 1-1, where t is time to traverse the flight tube, L is the length of the flight tube, m is the mass, z is the number of charges, e is the charge and V_s is the initial acceleration potential.

$$t^2 = \frac{m}{z}\left(\frac{L^2}{2eV_s}\right)$$

Equation 1-1

Given that the terms within the parentheses are constant, the time taken for an ion to traverse the flight tube is proportional to the square root of the *m/z*. This indicates that molecules of low mass, or high charge, will traverse the flight tube faster than large molecules with fewer charges (de Hoffman and Stroobant 2007).

For many years MALDI was the preferred ionisation method for use with TOF analysers, principally due to the significant loss in sensitivity associated with the use

of continuous ionisation methods, such as ESI, compared to the MALDI based pulsed ionisation method (Mirgorodskaya, Shevchenko *et al*. 1994). Throughout the 1980s a number of research groups dedicated time to developing TOF analysers capable of analysing a continuous ion beam without losing sensitivity. These new systems integrated orthogonal acceleration regions prior to the TOF analyser which collected ions prior to their acceleration into the TOF analyser, facilitating their use with a number of continuous ionisation techniques (Mirgorodskaya, Shevchenko *et al*. 1994, Guilhaus, Selby *et al*. 2000).

A number of factors in early instrument design led to a broad ion packet arriving at the detector directly resulting in low resolution. The introduction of delayed extraction (DE) and reflectron technology eliminated many of these problems and hugely improved the resolution of TOF analysers.

DE, first explored by Wiley and McLaren (Wiley and McLaren 1955), was designed to overcome the effect of kinetic energy differences of ions on resolution. Ions are given a short time (hundreds of nanoseconds to a few miliseconds) to expand into the field free environment of the source prior to being subject to a voltage pulse. By introducing a delayed pulse in this manner, ions of the same *m/z* are first allowed to separate based on their internal kinetic energy before the pulse is applied. Ions with higher energy drift further from the source of the pulse and thus receive less of a kinetic boost than those closer to the pulse, equalising the kinetic energy differential between the ions (Wiley and McLaren 1955).

The reflectron, developed by Mamyrin, uses a decelerating field that acts as an ion mirror slowing ions of the same *m/z* value and deflecting them back towards the detector, as illustrated in Figure 1-5. Ions of higher kinetic energy penetrate deeper into the decelerating that ions of lower kinetic energy but same *m/z*. The additional time spent within the field corrects the kinetic differential between the m/z ions, much in the same way as a DE, ensuring a narrower ion packet arriving arrives at the detector increasing resolution. To a lesser extent a reflectron also increases TOF resolution by increasing the length of the TOF flight tube. Reflectrons use two main configurations, V-optics, which makes use of a single decelerating field, or W-optics, which uses two decelerating fields that send ions back and forth, resulting in three passes through the field prior to detection. Whilst the latter option improves

instrument resolution it can also lead to a loss in sensitivity as a result of ion loss (de Hoffman and Stroobant 2007).

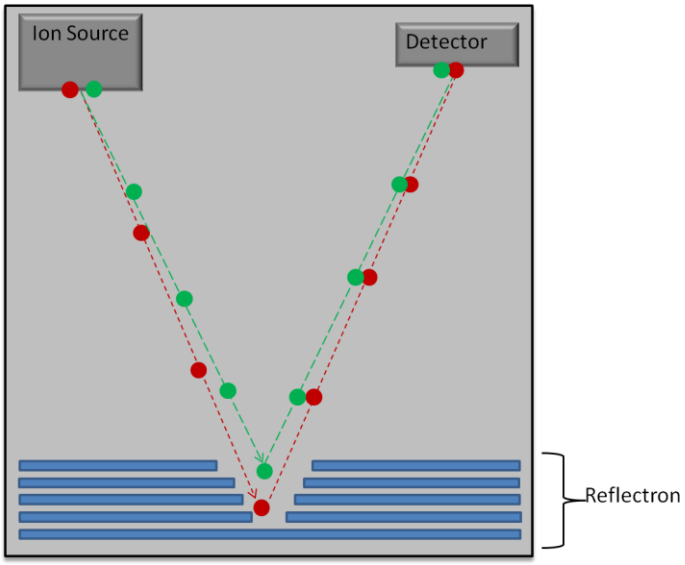


Figure 1-5 - Schematic of a time of flight mass analyser with a reflectron.
Two ions of the same m/z leave the ion source at the same time, but with different kinetic energies. The red ion has a higher kinetic energy than the green ion, and therefore penetrates deeper into the decelerating field. This serves to normalise the kinetic differences of the ions to ensure they arrive at the detector at the same time.

1.1.4 Detectors

The final element in an MS system is the detector. Ions separated by an analyser are passed into a detector, which converts the ions into a suitable signal to provide the experimental output. Ion detection is based on one of three physical properties: mass, charge or ion velocity. Of critical importance is that not only the *m/z* of an ion is detected, but also the ion abundance. Modern instruments convert ions detected into an electrical current proportional to ion abundance.

Unlike the developments in ionisation methods there has been no revolution in detector design, with advances coming as part of ongoing development to improve ion detection. As with all parts of a mass spectrometer, a number of configurations of detector are available and the choice of detector used is influenced both by the instrument and the analysis required.

The earliest detectors date back to the initial work performed by J. J Thomson in 1913. Photoplate detectors in 1913 consisted of luminescent screens which were used to visualise particles by means of a fluorescent spot (Squires 1998), although these were later replaced with ion sensitive photoplates inserted in the vacuum

regions of the MS after the analyser. Since these early days ion detection has become more sophisticated.

One of the biggest challenges associated with ion detection is MS systems is related to electrical signal strength. A typical ion beam has a current of subfemto- to nanoamphere, which requires significant electronic amplification before it can be converted to a voltage for interaction with electronic circuits.

The multichannel plate (MCP) detector is a variant of the continuous dynode electron multiplier (CDEM) detector. The MCP consists of a disk containing a series of channels each of which is a few tens of millimetres in diameter and a few millimetres in length coated in an electron-emissive surface. Ions strike the surface near a channel, resulting in a burst of electrons emitted into the channel. This sets off a cascade of electron emissions down the channel, increasing the size of the electron cloud until it is detected by an anode at the base of the channel (Koppenaal, Barinaga *et al*. 2005).

Standard electron multiplier (EM) detectors rely heavily on the velocity of an ion in order to generate the electron cloud cascade. Larger ions tend to produce lower signals on EM detectors since they have a lower ion velocity. CDEM detectors, like the MCP, produce a continuous electric field across its length. This field provides ample acceleration for the electron cloud, serving to enhance the signal even for larger ions. With the application of mass spectrometry transitioning more towards the study of biomolecules, which inherently produce large ions, this form of detector has proven invaluable for enhancing signal intensity and sensitivity.

A mass spectrometer must be able to convert the signal produced by a detector into a digital signal capable of being processed by a computer system to create an output mass spectrum. Conversion is usually performed by one of two systems; an analogue-to-digital converter (ADC) or a time-to-digital converter (TDC). ADC's take the amplified analogue signal produced by a MS detector, filter out the background noise and convert to a digital output before transferring it to memory. The information content produced from an ADC is proportional to every data point across a single MS spectrum and therefore the ADC sampling frequency is determined by the scanning speed of the MS instrument being used. TDC's make use of a discriminator to register the analogue electron pulse produced by the detector

once it overcomes a certain threshold. The discriminator converts the signal into signal pulse which persists for the duration of the electron pulse. The TDC measures the duration of the signal pulse and converts it into a histogram to be stored in the memory prior to transfer to the computer system. TDC's are time counting devices and therefore make no differentiation between the number of ions striking the detector simultaneously from the same flight cycle. If two ions strike the detector sequentially the second ion is not registered as the system is unable to register ions for a short interval after each ion event. This is known as dead time. As a consequence, TDC's suffer from limited dynamic range, with a deleterious effect on mass accuracy for intense ions (de Hoffman and Stroobant 2007).

1.1.5 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) refers to any method which involves two stages of mass analysis using a dissociation process or chemical reaction between the analysers to change the mass or the charge of the ion. Depending on instrument configuration it is possible to carry out multiple selection and fragmentation events in a process known as MS^n, where n is the number of generations of ions studied. MS/MS can be performed either in space or in time, and the method used is dependent on instrument configuration.

MS/MS in space uses an initial mass analyser to select and isolate a precursor ion, which is then subjected to fragmentation. The fragments are then separated using a second mass analyser prior to detection. Typical instrument configurations for MS/MS in space are triple quadrupoles (QqQ), aligned time of flight time of flight (TOF-TOF) or quadrupole time of flight (QTOF) instruments. Located between the mass analysers is a collision cell suitable for fragmenting the ions (Figure 1-6).

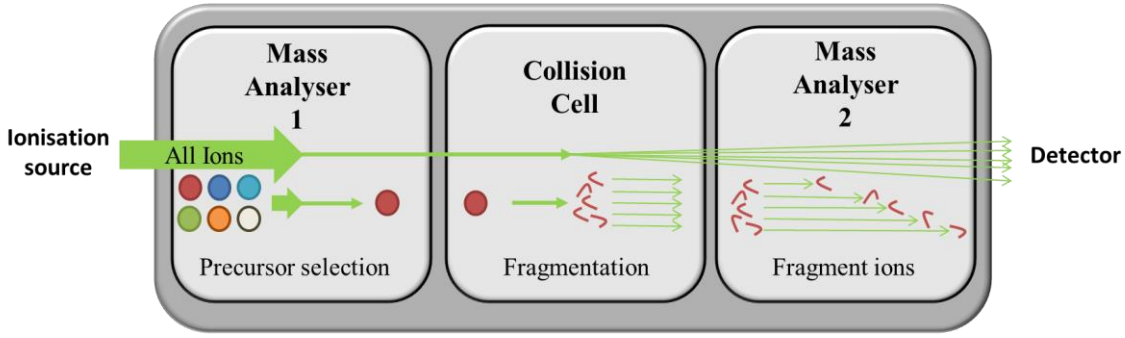


Figure 1-6 - Schematic of the arrangement of mass analyses in a tandem MS instrument.
A full ion packet is introduced into mass analyser 1, which is used to select only a single *m/z* ion; all other ions are filtered out and lost. The selected ion is then fragmented within the collision cell. These fragment ions are analysed by mass analyser 2 prior to detection.

Whilst there are a number of potential methods for ion activation and fragmentation the most widely applied method is collision induced dissociation (CID). Pioneered in the 1960's by McLafferty and Jennings (McLafferty and Bryce 1967, Jennings 1968) CID achieves fragmentation via collisions between the precursor ion and an inert buffer gas (usually He, N_2 or Ar). In the collision cell the translational energy of the ion is converted into internal vibrational energy which leads to ion fragmentation. By accelerating the precursor ions as they enter the chamber fragmentation may be increased and secondary fragmentation can also be induced (Seidler, Zinn *et al*. 2010). This form of experiment has a number of potential applications for biological mass spectrometry, which are discussed later.

1.1.6 Ion Mobility Mass Spectrometry

Ion mobility spectrometry (IMS) is a technique that separates ions based on their velocity when travelling through a certain medium. The first reported use of ion mobility was in 1898 by John Zeleny (Zeleny 1898), although the underlying principles behind IMS were later described by Langevin in 1903. Based on this initial work it was understood that ions could be separated based on their velocity through an inert gas with a superimposed electric field. Ions passing through the field have a specific, constant velocity determined by their physical properties (Uetrecht, Rose *et al*. 2010); their size/charge ratio as determined by mass, charge and shape (Kanu, Dwivedi *et al*. 2008). IMS is a rapid technique that has high sensitivity and selectivity and for these reasons IMS has been utilised for the analysis of drugs, explosives and small molecules for many years (Mesleh, Hunter *et al*. 1996, Guharay, Dwivedi *et al*. 2008).

Most simple IMS systems consist of a single drift cell. In this instrumental set-up ions are introduced into the drift tube in packets. The ions traverse the cell, which is filled with a neutral inert gas (He, Ar or N_2), under the influence of a weak, uniform electric field. As the ion packets diffuses through the drift cell larger ions suffer more collisions with the buffer gas compared to the smaller ions, their movement through the cell is retarded relative to the smaller ions resulting in a longer drift time, see Figure 1-7 (Clemmer and Jarrold 1997).

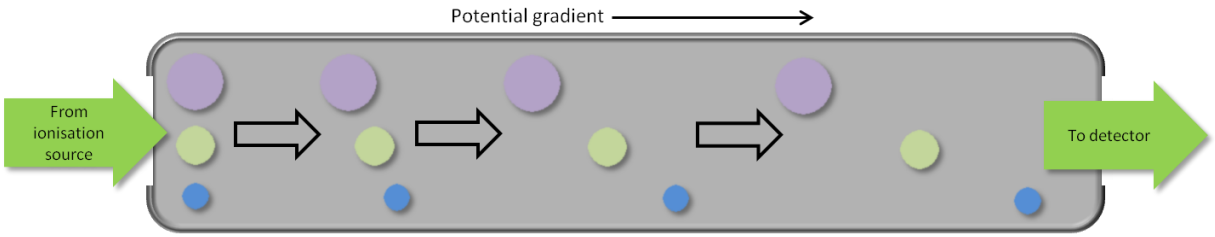


Figure 1-7 - Schematic of a drift time ion mobility cell.
Ions traverse the cell under the influence of a weak electric field. The cell is filled with an inert buffer gas, usually N_2 or He, which acts to slow the ions movement through the cell. Smaller ions, those shown in blue and green, collide less with the buffer gas and therefore traverse the cell faster than larger ions, shown in purple.

The use of IMS in tandem with MS, known as ion mobility mass spectrometry (IM-MS) dates back to the 1960s. Over the years IMS has been coupled to most MS detectors, starting with magnetic sector instruments, quadrupoles and TOF's and, more recently, moving on to ion traps and FTICRs (Uetrecht, Rose *et al*. 2010). The use of a coupled IMS-MS system creates an extremely powerful analytical tool. IMS separates molecules based on their size/charge ratio and MS can be used to determine *m/z*; when used in tandem it becomes possible to investigate an ions shape, more specifically its rotationally averaged collision cross section.

IM-MS has a number of potential applications. Whilst initial experiments focussed on the properties of atoms and simple molecules it has subsequently been used as a method of separating complex ion mixtures in proteomics (Valentine, Plasencia *et al*. 2006) and glycomics (Clowers, Dwivedi *et al*. 2005). IM-MS has had the most significant impact, however, in the field of structural biochemistry for the analysis of protein structure and protein complex assembly. The Jarrold, Clemmer and Bowers groups have published pioneering work in this field (Clemmer and Jarrold 1997, Hudgins, Ratner *et al*. 1998, Wyttenbach, Kemper *et al*. 2001)

Initial publications were based on data acquired using in-house built drift cell IM-MS instruments. The first available commercial instrument incorporating ion mobility mass spectrometry was the Synapt G1 HDMS system in 2006, followed by the Synapt G2 HDMS system in 2009. Since then a number of IM-MS platforms have become available including a commercial drift cell instrument from Agilent (the Agilent 6560 Q-TOF system) and an optional FAIMS upgrade available for instruments from Thermo Scientific.

Drift cell IM-MS (DCIMS) utilises a convention drift cell, the mode of operation of which has been previously described. Drift cell instruments have high IMS sensitivity and resolution. The relationship between an ions drift time, mass (m), charge (z), rotationally averaged cross section (Ω) and interaction with buffer gas under the effects of an electric field of known strength are well understood. A DCIMS instrument measures drift time, mass and charge, and since all other instrumental conditions remain constant, it is possible to directly calculate Ω for ions analysed using this method using Equation 1-2 (Clemmer and Jarrold 1997, Pringle, Giles *et al*. 2007) where N is the background gas density, ze the ionic charge, μ the reduced mass of the ion-neutral pair, k_b is Boltzmann's Constant, T the gas temperature and K_0 is the reduced mobility, the measured mobility corrected to 273.2K and 760 Torr.

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

Equation 1-2

High-field asymmetric waveform ion mobility spectrometry (FAIMS) exploits the differential mobility of gas phase ions at high electric field strengths. At low field strengths, such as in DCIMS ion mobility is independent of field, whereas at high field strengths mobility becomes a function of field. In this experiment ions are allowed to pass through parallel electrodes at ambient temperature and pressure. An asymmetric voltage pulse is applied to one electrode to draw ions towards it, whilst the other electrode has a continuous voltage applied to it to compensate for the drift towards the former electrode. This is known as the correction voltage. Ions require different correction voltage strengths to correct their drift towards either plate, allowing for ion mobility separation (Barnett, Ells *et al*. 1999). The theory of FAIMS

is currently poorly understood, and the high field strengths can induce structural changes making it an unsuitable method for estimating protein structural dimensions (Uetrecht, Rose *et al*. 2010).

1.1.7 Travelling Wave Ion Mobility

The first and second generation Synapt HDMS systems make use of travelling wave ion mobility (TWIMS) as the IMS method of separation. The IMS cell, shown in Figure 1-8, is located between the trap and the transfer cells.

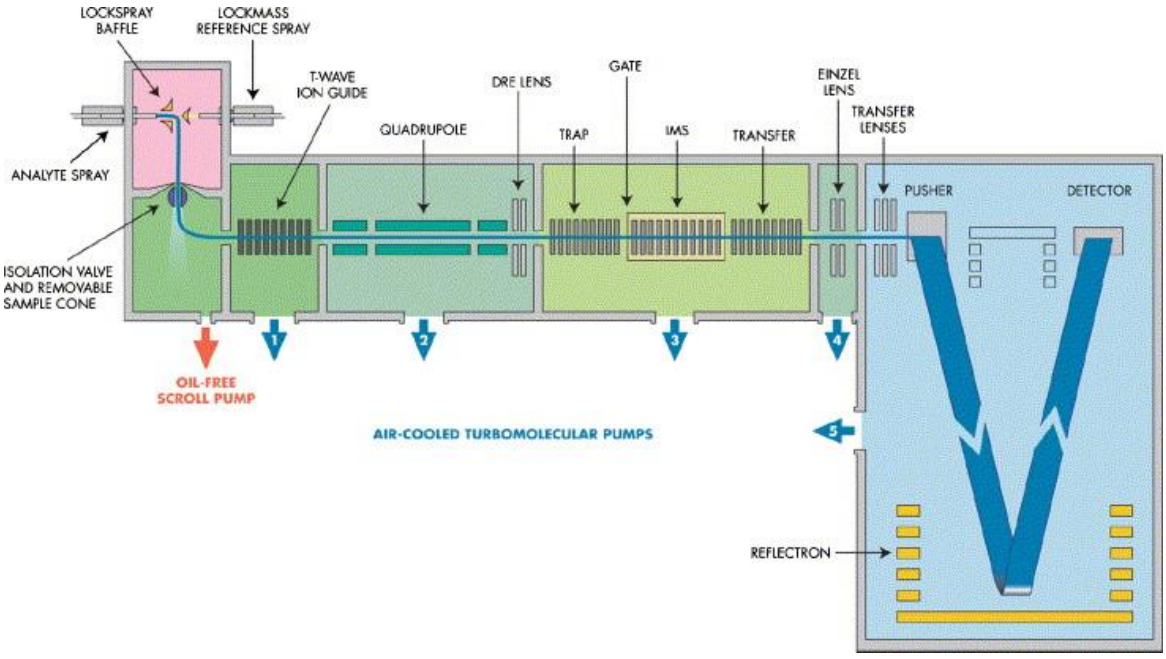


Figure 1-8 – The schematic of the Synapt G1 HDMS system, the first generation commercial mass spectrometer with an integrated T-wave ion mobility cell.
Ions produced by ESI in the source move through the quadrupole into the SRIG region of the instrumnet, containing two fragmentation cells (Trap and Transfer) either side of the IMS cell. Follow the SRIG region ions are analysed by TOF before detection.

TWIMS instruments operate under similar principles to DCIMS instruments. The difference between TWIMS and DCIMS lies in the use of the stacked ring ion guide (SRIG) to generate a voltage wave that propels ions along the cell.

Ions are introduced into the IMS cell in packets. Each ring in the ion guide has a RF field applied and each pair of rings has positive and single negative RF voltages applied. The RF field produced serves to radially confine the ions, preventing radial diffusion and reducing scattering effects. A transient DC voltage is applied to pairs of stacked rings in sequence along the length of the SRIG to create a DC electrical wave. Ions are repulsed by this potential wave and are thus propelled through the

SRIG. As in a drift cell ions with larger Ω values undergo more collisions with the buffer gas and, as a result, suffer drag which can cause them to slip behind the wave front, as depicted in Figure 1-9 (Giles, Pringle *et al*. 2004, Thalassinos, Slade *et al*. 2004, Pringle, Giles *et al*. 2007), and thus take longer to traverse the cell. This method of ion mobility separation has improved speed of data aquisition and sensitivity relative to many conventional DCIMS instruments (Scarff, Thalassinos *et al*. 2008)

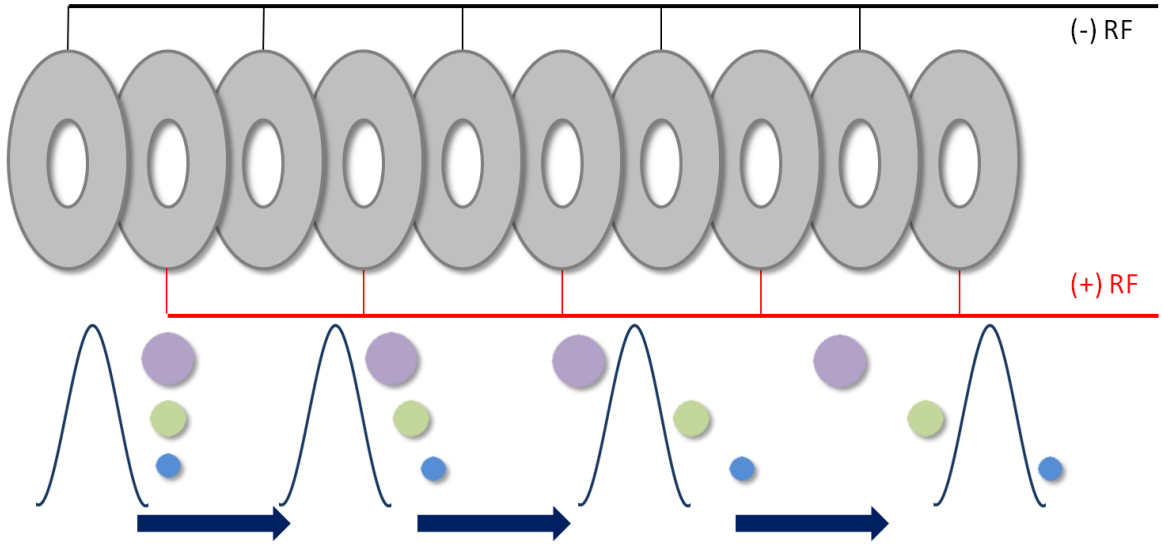


Figure 1-9 - Schematic illustrating ion mobility through a travelling wave cell.
As with drift cell IMS, ions are introduced into T-wave cell as a single packet. DC voltages applied to pairs of rings create an electrical potential wave, which pushes ions through the cell. Larger ions collide more with the inert buffer gas and are slowed in their movement, until they fall behind the DC potential wave. These ions are then carried along on the next DC wave until the either fall behind the potential wave agian, or they exit the mobility cell.

Separation using a T-wave is reliant on many of the same factors shown in Equation 1-2, but also requires the effects of both wave height, determined by DC voltage, and wave velocity to be considered. It is not currently possible to calculate absolute Ω values from TWIMS data in the same way as DCIMS data, since the properties of the electric field are varied across the SRIG (Lanucara 2014). It is possible, however, to obtain experimental estimates of Ω values by using reference compounds of known CCS to calibrate the data (Scarff, Thalassinos *et al*. 2008).

1.1.8 Calibration of Ion Mobility data

The framework for calibrating TWIMS MS was initially developed by Wildgoose, and has been refined and developed by a number of groups (Ruotolo, Benesch *et al*. 2008, Scarff, Thalassinos *et al*. 2008, Williams and Scrivens 2008, Smith, Knapman

et al. 2009, Thalassinos, Grabenauer *et al*. 2009, Michaelevski, Kirshenbaum *et al*. 2010). The method described below, and used in this thesis, has been taken from Michaelevski *et al* (Michaelevski, Kirshenbaum *et al*. 2010).

Arrival time, measured as scan number (*n*) relates to the push of the TOF; there are 200 TOF acceleration pushes registered between the release of each ion packet into the mobility cell, resulting in 200 spectra per ion packet. Multiplying the pusher time, in milliseconds, by the scan number provides an arrival time (Equation 1-3)

$$t_d = n \times pusher\ time$$

Equation 1-3

This value is then corrected for *m/z* dependant flight (t'_d), which is the time between the transfer cell exit and the TOF pusher. This is achieved by using the enhanced duty cycle (EDC) coefficient (*c*), which is instrument specific, but has a value typically between 1.4 and 1.6 (Michaelevski, Kirshenbaum *et al*. 2010) (Equation 1-4).

$$t'^{d} = t_d - \frac{c\sqrt{m/z}}{1000}$$

Equation 1-4

Previous calibrations have corrected for m/z independent flight (time spent in the transfer cell after the IMS cell). It has been observed, however, that m/z independent flight had little effect on the calculated cross sections and this step has subsequently been omitted for simplicity.

Following correction of drift time of the analyte, each calibrant cross section (Ω) is then corrected for ion charge state (*z*) ion mass (*m*) and mass of the buffer gas (m_g), Equation 1-5. Typically the buffer gas is either N_2 or He.

$$\Omega' = \frac{\Omega}{z\sqrt{\frac{1}{m} + \frac{1}{m_g}}}$$

Equation 1-5

From the calculated values, a plot of Ω' vs t'_d for the calibration standards is fitted to either a power ($y=Ax^N$) or linear ($y = Ax + N$) series fit, ensuring the R^2 value is ≥0.99. Both A and N are then used for the calculation of estimated cross section using the power fit, or linear equations.

$$CCS = A\, {t'_d}^N z \sqrt{\frac{1}{m} + \frac{1}{m_g}}$$

Equation 1-6

$$CCS = [A\, t'_d + B\,]\, z \sqrt{\frac{1}{m} + \frac{1}{m_g}}$$

Equation 1-7

As with any analytical technique it is important to ensure that appropriate calibrations standards are used for the desired analyte. The Clemmer group published a series of measured cross sections of denatured ubiquitin, cytochrome c and lysozyme to be used as calibration standards (Shelimov, Clemmer *et al*. 1997, Valentine, Anderson *et al*. 1997, Valentine, Counterman *et al*. 1997), whilst some groups used other denatured proteins, such as myoglobin (Scarff, Patel *et al*. 2009) as their calibration standard of choice. Using these denatured standards it is possible to obtain accurate CCS estimates, assuming both the protein and calibration standards are examined under the same experimental conditions (Leary, Schenauer *et al*. 2009).

Whilst these denatured standards have been shown to provide good estimates for many systems, native proteins can have significantly smaller mobilities, and in the case of complexes much larger masses, than these unfolded protein calibrants cover. Using these standards for estimating intact proteins, and large protein complex, Ω values leads to CCS values highly dependent on instrument settings with a larger, difficult to estimate, errors. Recent reports have helped rectify this problem by providing a number of optional native protein calibration standards, ranging from cytochrome c (12 kDa) to the GroEL complex (801 kDa), for use in TWIMS experiments (Salbo, Bush *et al*. 2012). These reports not only outline appropriate calibration standards, but also highlight the importance of choosing the correct type

of calibrant that closely brackets an appropriate mobility window and mass range for the analyte ions, and optimising the instrument parameters for these standards. By ensuring these conditions are met it is possible to reduce the instrument dependence of the Ω-values recorded.

1.1.9 Computational modelling of CCS

In addition to comparing Ω values obtained using TWIMS with those obtained by DCIMS it is also possible to calculate Ω vales from published X-ray crystal structures. Scarff *et al* showed that Ω values estimated for a series of protein standards using TWIMS correlated well with those calculated from protein data bank (PDB) files from both X-ray and NMR structures (Scarff, Thalassinos *et al*. 2008).

There are four main methods of calculating Ω values from published PDB data; the projection approximation (PA); exact hard sphere scattering (EHSS) method; trajectory method (TM); and the more recently introduced projection superposition approximation (PSA) method.

The PA method was first described by Mack in 1925 (Mack 1925), and was adapted for the use in modern CCS estimation by the Bowers group (von Helden, Hsu *et al*. 1993).

The PA method replaces an ion cross section with its projection; the 'shadow' the ion would cast. A number of projections are generated from different angles around the ion, and an average is used to generate the cross section. The PA method is the least computationally intensive method making it the fastest method available. It does not, however, account for interactions between the ion and the buffer gas. Whilst this method is suitable for small molecules it has been shown to consistently underestimate the Ω value for larger peptides and proteins (Jarrold 1998).

The TM approach is the most thorough modelling method. It takes into consideration many of the interactions between the buffer gas and the ions, and their respective velocities and scattering angles. By calculating scattering angles this approach assesses the full three-dimensional structure of an ion, including any concave and convex regions of the ions surface. Whilst the TM calculation is by far the most robust, it is also the most computationally intensive; the algorithm can take several

weeks on a standard computer to compute structures that other methods take only a few days to compute (Bleiholder, Wyttenbach *et al*. 2011)

The EHSS method may be considered as a middle ground between the PA and the TM; the EHSS treats the molecule as a single hard sphere for the calculation of interactions with buffer gas. This approach does not take into consideration the subtle differences caused by an ions surface shape (concave/ convex surface elements) like the TM method, but does not completely ignore buffer gas effects like the PA method. The result is a method that gives slightly more robust calculations than the PA with a fraction of the computational time required for the TM method (Mesleh, Hunter *et al*. 1996, Bleiholder, Wyttenbach *et al*. 2011).

The PSA method was developed by the Bowers group to provide an improved method for calculating theoretical Ω values for biomolecules, taking into consideration size and shape effects like the TM method but with a reduced computational demand. The PSA method builds on the PA method – an averaged projection area is calculated which is then corrected for any deviation from a perfect sphere using a shape factor. The result is a calculation that provides theoretical Ω values similar to the TM method but with a computational demand comparable to the PA method; for PDB entry 1IHM (a spherical virus capsid) the PSA method takes 663 minutes to compute Ω, compared to 5 minutes for the PA, 1401 minutes for the EHSS and 47 days (67680 minutes) for the TM (Bleiholder, Wyttenbach *et al*. 2011).

1.1.10 Comparing experimental and computational CCS measurements

In order to validate the experimental Ω values obtained experimentally it is possible to compare the CCS obtained from computational methods with those obtained experimentally. A number of previous studies have shown that it is possible to obtain CCS estimates from T-wave instruments that agree well with those calculated from X-ray crystal structures (Ruotolo, Benesch *et al*. 2008, Bush, Hall *et al*. 2010, Salbo, Bush *et al*. 2012), if experimental parameters are carefully controlled and a suitable calibration curve is produced. It is important when performing this comparison to ensure that computationally derived CCS values and experimental CCS values use of the same buffer gas, N_2 or He, as this can lead to differences in the drift time, and

thus CCS measurements (Bush, Hall *et al*. 2010, Jurneczko, Kalapothakis *et al*. 2012).

This is particularly problematic when using TWIMS, as TWIMS makes use of N_2 as the buffer gas for IMS separation, whilst all of the computational methods for estimating CCS use He. A modified version of the MOBCAL program, which performs PA, EHSS and TM methods, configured to use N_2 as the buffer gas has been developed by Iian Campuzano, but it has been optimised for small molecules and drug-like compounds and would not be suitable for protein analysis (Campuzano, Bush *et al*. 2012).

1.2 Proteins; structure and function

Proteins, along with lipids, carbohydrates and nucleotides, are biological macromolecules that are essential to all life in earth. Proteins have a variety of roles within a biological system including catalysis, molecular transport and storage, coordinated motion, mechanical support, immune protection, cellular signalling, growth and differentiation; proteins are involved in virtually every biological process across nature. The key to this diverse functionality lies in in protein structure (Stryer 2011).

1.2.1 Protein Structure

Proteins are polymer molecules comprised of amino acids capable of folding into complex three-dimensional shapes. Protein structure can be broken down into four categories: primary, secondary, tertiary and quaternary.

Primary structure refers to the sequence of the amino acids of a protein. There are twenty types of unmodified amino acid molecule, each with a conserved core and unique R group (see Figure 1-10).

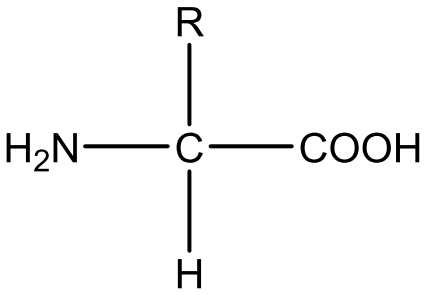


Figure 1-10 – The generic depiction of an amino acid. Differences in R group give rise to the different amino acids.

Secondary structure refers to the spatial arrangement of the amino acids in regions close in sequence. There are three major secondary structure elements; α-helical and β-sheet are regular structural features formed by amino acids linked by hydrogen bonding between the amide and carboxyl groups along the peptide backbone. The third structural element is a random coil; a disordered structure featuring few hydrogen bonds.

Tertiary structure refers to the spatial arrangement of amino acids close in space but not in sequence – it refers to the three-dimensional folding of the primary and secondary structure. Tertiary structure also includes consideration of any disulphide bonds a protein contains to stabilise the three dimensional structure.

In many proteins multiple polypeptide chains, each with their own primary, secondary and tertiary structure, come together to form the protein complex. Each polypeptide chain may be referred to as a subunit, each of which comes together to form a functional protein, or protein complex. Quaternary structure refers to both the spatial arrangement and the interactions of these subunits (Stryer 2011).

Ultimately it is thought that a proteins primary structure determines how a protein folds into its higher order structure. This was shown by Christian Anfinsen in the 1950's (Anfinsen, Redfield *et al*. 1954, Stryer 2011) and the structure, in turn, determines a proteins function. Changes in primary structure can result in variation in the manner in which a protein folds. Many inherited diseases are caused as a direct result of a DNA mutation which alters the primary sequence of a protein, thus altering its three dimensional shape and changing its function. An example of this may be found in inherited haemoglobin disorders. It is now becoming apparent that misfolded proteins are capable of not only causing, but transmitting disease, such as

prion proteins that are thought to be implicated in bovine spongiform encephalopathy (Stryer 2011).

In addition to protein folding, post-translational modifications (PTMs) are an integral part of a proteins structure, and play an important role protein function. PTMs take many forms including glycosylation, phosphorylation, ubiquitination, lipidation and proteolytic cleavage. PTMs can to either to activate or deactivate protein activity, determine the cellular localisation of a protein, target the protein for degredation or modify the protein structure (Mann and Jensen 2003).

1.2.2 Methods of studying protein structure

With protein structure tied closely not only with function, but also with disease, it has become vital to study proteins in order to gain an understanding of not only how they fold, but also how this relates to their function.

Proteins, however, are not static structures in solution. They exist as an ensemble of structures, by virtue of their ability to flex and move, and it is this inherent flexibility that allows many proteins to perform their functions. Changes in the ensemble may be as a result of response to external stimuli, such as cooperative binding, or as a result of structural changes due to mutation. It is therefore important not only to study changes in protein structure but also changes in structural dynamics. There are a number of tools available for the study of both protein structure and dynamics.

Small angle scattering (SAXS) is a solution phase technique suitable for studying protein structure. SAXS measures the interference pattern of scattered X-rays to provide information on molecular shape. Data is low resolution and therefore is only really suitable for studying quaternary and tertiary structure and overall protein shape (Kaltashow and Eyles 2005).

Circular dichroism (CD) utilises the differential absorbance of polarised light by chiral molecules to probe protein shape. Specifically, CD is capable of investigating secondary structural elements due to the chirality induced by side-chain/ disulphide bridge orientation in certain structural elements (Kaltashow and Eyles 2005).

Vibrational spectroscopy is another low resolution technique suitable for the quantification of secondary structure. It exploits the fact that molecules absorb

specific frequencies, typically in the infra-red region, that resonate with their molecular bonds (Schweitzer-Stenner 2006).

Whilst these techniques are all capable of studying protein structure the main methods of structural elucidation are X-ray diffraction and Nuclear Magnetic Resonance, with those techniques mentioned above providing complementary information to these high resolution techniques (Schweitzer-Stenner 2006).

X-ray crystallography is technique capable of providing high resolution (2 - 4Å) information regarding a proteins three-dimensional structure. Briefly, crystals are grown from a supersaturated protein solution. Crystalline proteins adopt a conserved structure which is repeated to form a lattice. X-ray beams are made to diffract off these crystals. The diffraction pattern is determined by the atomic positions of the atoms, and provides electron density for each repeat unit, or protein molecule, in the crystal. The structure of the protein is then calculated from the electron density (Keith 2008). Data quality is unaffected by protein size making it ideal for the study of larger biomolecules. It is not, however, without its limitations, foremost being the difficulty in obtaining a representative diffracting protein crystal. Optimal conditions for growing suitable crystals are unpredictable; screening a number of different methods is required for each molecule making the method time consuming for routine analysis (Berkowitz, Engen *et al*. 2012). Optimum conditions typically use non-physiological buffers and temperatures that do not adversely affect the protein backbone but can alter side chain orientation (Garman 2014). The structure obtained by this method is an average of all the protein conformations present within the lattice at any given time. Due to the requirement to form a crystalline lattice there is a bias in the structural behaviour of a molecule towards a more ordered conformation (Garman 2014) and it is possible that the crystal lattice structure is not representative of the solution phase structure (Schweitzer-Stenner 2006). Molecules with intrinsically disordered regions are often difficult to crystallise due to their inherent flexibility, such as Antibody molecules complete with linker regions (Saphire, Stanfield *et al*. 2002). Despite these limitations X-ray diffraction remains the major method used to study protein structure, as evidenced by the inclusion of over 82,000 protein structures obtained using X-ray crystallography within the Protein Data bank (PDB) at the time of writing (PDB 2014).

Nuclear magnetic resonance (NMR) has, for many years, been a major method for the study of protein structure and dynamics, transient molecular structures, and weak interactions. NMR, like X-ray crystallography, is a high resolution technique capable of atomic resolution data (Wang, Zhang *et al*. 2014). NMR is based on the principle that atomic nuclei have spin energy levels which split when under the influence of a magnetic field. Certain biologically significant atoms (^{1}H, ^{13}C and ^{15}N) with spin values of ½ can occupy two spin states, either aligned with or against the magnetic field. Application of a radio frequency field causes the atoms to change spin alignment. Atoms are then allowed to relax to the ground state, producing a signal, which is measured. Resonance frequency is determined by atomic environment and the signals provide structural information on atoms linked both chemically, through a number of bonds, and through space (Kaltashow and Eyles 2005). NMR suffers from a number of experimental limitations. NMR instruments typically have poor sensitivity compared with complimentary techniques, requiring milligrams of sample to obtain relevant data. NMR is also limited by analyte size; sensitivity and resolution are proportional to molecular weight, with molecules of >30 kDa suffering from low sensitivity and poor resolution (Berkowitz, Engen *et al*. 2012, Wang, Zhang *et al*. 2014). There is a need to enrich the isotopic concentration of ^{15}N and ^{13}C and recombinant methods of synthesis are commonly carried out. NMR remains a powerful technique for the study of protein structure, with a number advances in instrumentation and methods attempting to overcome these limitations (Kanelis, Forman-Kay *et al*. 2001, Wang, Zhang *et al*. 2014) .

Hydrogen exchange, or hydrogen deuterium exchange (HDX) as it is now commonly known, is a method for studying protein structural dynamics that has historically been linked to NMR, but is becoming increasingly associated with mass spectrometry.

HDX is based on the concept that labile hydrogen atoms (^{1}H) on a protein may exchange with those in solution. Replacing normal protonated buffers with deuterated buffers, containing >95% deuterium, ensures that all ^{1}H atoms are systematically exchanged with deuterium. The kinetics of deuterium incorporation may then be measured (Wales and Engen 2006).

Whilst all hydrogen atoms have the potential to exchange, HDX studies on protein dynamics primarily focus on the backbone amide hydrogens, since more labile hydrogen atoms exchange too quickly to be observed (Katta and Chait 1991). The chemistry and kinetics of amide hydrogen exchange is now well understood and, for unfolded proteins, or polypeptides, it is possible to calculate the intrinsic chemical exchange rates (k_2, see Equation 1-10) for these backbone amide protons based on pH, temperature and neighbouring side chains (Englander 2006). Proton exchange may be described using the proton transfer theory (Equation 1-8) (Englander and Kallenbach 1983) and is catalysed by the strong acid hydronium (H_3O^+) ion or the strong base hydroxide (OH^-) ion, or their deuterated equivalents.

$$AH + B \rightleftharpoons (AH \ldots B \rightleftharpoons A \ldots HB) \rightleftharpoons A + HB$$

Equation 1-8

Chemical exchange rate is highly dependent on pH and temperature, which are rigorously controlled in HDX experiments (Englander 2006).

The study of folded proteins also takes into consideration both pH and temperature. The chemical environment of the backbone amide, determined by adjacent side chains, becomes less of a concern in these studies relative to the solvent accessibility of the amide and the protection that amide hydrogen has by virtue of hydrogen bonding. In a folded protein it is common for many amides such as those on the surface of the protein to exchange rapidly, whilst those buried within the hydrophobic core exchange more slowly (Wales and Engen 2006). Exchange rates between the same amide hydrogen can differ by as much as 10^8 between folded and unfolded protein states (Englander and Kallenbach 1983) as a result of these effects. Most amide hydrogens, given enough time, will participate in HX. Even those amides buried within the hydrophobic core are capable of exchange as a result of local mobility that opens the region up to solvent (Kim, Fuchs *et al*. 1993). The kinetics of exchange, k_{ex}, of any amide on a protein can be described using Equation 1-9

$$k_{ex} = k_f + k_u$$

Equation 1-9

Where k_{ex} is the sum of the kinetics of exchange of the folded (k_f) and unfolded (k_u) forms of the protein (Kim, Fuchs *et al*. 1993). k_f dominates for those amides not participating in H-bonding located towards the surface of the protein, and can further be described by using Equation 1-10.

$$k_f = \beta k_2$$

Equation 1-10

where β is the probability factor of exchange (between 0-1), which is a function of a number of factors including solvent accessibility and hydrogen bonding environment, and k_2 is the intrinsic chemical exchange rate for an unstructured peptide (as described previously). β values closer to 1 indicate a higher probability of water and catalyst presence near an exchanging amide at any given time.

k_u dominates for amides located further from the surface; the kinetics behind their exchange is more complex and requires a partial, or complete, unfold of the protein for exchange, described in Equation 1-11

$$F_H \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} U_H \underset{D_2O}{\overset{k_2}{\rightarrow}} U_D \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} F_D$$

Equation 1-11

where F and U refer to the folded and unfolded forms of the protein, respectively, and subscript H and D refer to the protonated and deuterated forms of the protein, respectively. The relationship between the rate constants k_1, k_{-1} gives rise to the different k_u values. In instances where $k_2 >> k_{-1}$ a single unfolding event is sufficient for complete amide exchange, thus

$$k_u = k_1$$

Equation 1-12

in what is known as EX1 kinetics. Under physiological conditions it far more common, however, for $k_{-1} >> k_2$. In these instances the kinetics of exchange are too slow to happen in a single event; the probability of exchange occurring is exceedingly low and therefore a number or transitions between folded and unfolded states need to occur in order to observe exchange. In these instances k_u is described as

$$k_u = \frac{k_1}{k_{-1}} k_2 = k_{unf} k_2$$

Equation 1-13

where k_{unf} is the constant used to describe the unfolding process (Wales and Engen 2006).

Whilst the effects of solvent accessibility and H-bonding are evident, it is not possible to differentiate between the two factors using HDX techniques since both factors occur in parallel to one another (Wales and Engen 2006). Whilst it was initially believed that H-bonding between amides would prevent exchange it is now understood that protons participating in H-bonding undergo a transient H-bond separation to open the amide to solvent catalysis (Englander 2006). At pH 7 all proton exchange is base-catalysed; OH^- acts as the catalyst and H_2O the source of the new proton (or OD^- and D_2O in the case of deuterated buffer)

Amide hydrogens have been described as occupying four types of mobility domains based on their kinetics; the slow exchange core represents the most protected environment and contains the amides that exchange slowest; surface amides that exchange rapidly; flexible buried regions, such as loop regions, capable of high mobility that exhibit rapid exchange, yet not as rapid as surface amides; and secondary structural regions with exchange values between those of the core amides and the flexible, buried amides (Kim, Fuchs *et al*. 1993).

All of these early kinetic and fundamental studies were carried out using NMR as the method of study. Whilst ^1H has spin ½, and therefore is observable by NMR, ^2H is invisible to NMR, and thus measuring the loss of ^1H signal gives rise to information regarding protein dynamics and kinetics (Kaltashow and Eyles 2005).

1.3 Biological mass spectrometry

The advances in technology which were discussed in sections **1.1 Mass spectrometry** have made it possible to utilise mass spectrometry for the study of biomolecules. Over the last 30 years the field has grown to encompass most aspects of biological science including the study of DNA, carbohydrates, metabolites and proteomes and protein structure (Konermann, Vahidi *et al*. 2013).

1.3.1 Structural mass spectrometry

Mass spectrometry has unique features amongst the methods described since it can be used to answer a number of structural questions; from protein sequence information and identification of post-translational modifications to probing conformation and dynamics.

Primary sequence information

Primary sequence information can be obtained by a using tandem MS approach as described in section **1.1.5 Tandem Mass Spectrometry**. Typically, peptides generated from the protein after proteolytic digestion are analysed within the instrument.

Individual peptides are isolated and subjected to CID; fragmentation occurs along the amide backbone generating characteristic ions depicted in Figure 1-11 (Papayannopoulos 1995) based upon the nomenclature described by Roepstorff (Roepstorff and Fohlman 1984). Depending on charge retention the ions are labelled on the N-terminal side of cleavage (b-ions) or the C-terminal side (y-ions). CID may be performed using Q-TOF, Triple Quad or ion trap instruments typically generate b and y ions due to the low energy (10-100eV) used to induce fragmentation

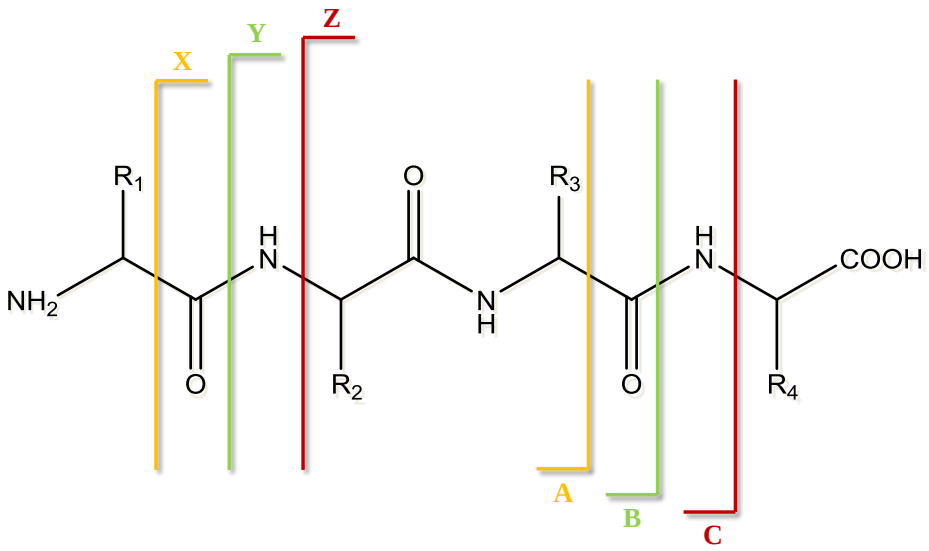


Figure 1-11 - The nomenclature of a CID spectra. Nomenclature is determined by both where the fragmentation occurs, and where the charge is retained. Imaged adapted from (Roepstorff and Fohlman 1984).

A typical peptide CID spectrum contains a number of peaks each corresponding to different peptide fragmentations. Identification of the mass loss between the peaks may be sufficient for identification of primary sequence information, which can either be performed manually or computationally.

Protein Conformation and Dynamics

There are a number of MS-based experiments capable of probing higher order protein structure that utilise native-MS, IMS-MS and HDX coupled to MS (HDX-MS).

Native MS

Native MS is the simplest tool for the study of protein structure since it requires no additional instrumentation; it is merely the analysis of a protein, or complex, in its native conformational state. As mentioned in section **1.1.2 Ionisation Sources** nESI is the softest ionisation method and native MS hinges on the low energy transition of molecules from solution-phase to gas-phase. During the ionisation process a protein is taken from a solvated solution phase environment and transferred to a desolvated, gas phase environment. It was initially believed that a protein would not preserve it's higher order structure during ionisation due to the lack of important solvent interactions in the gas phase (Wolynes 1995).

There is increasing evidence that over the timescale of a normal MS experiment a protein is able to retain a great deal of the non-covalent interactions it has within solution. The evolving structure of a gas-phase protein summarised by Breuker and McLaffetry is shown in Figure 1-12 (Breuker and McLafferty 2008). Brueker and McLafferty suggest that within a few picoseconds of desolvation the hydrophobic side chains collapse onto the protein backbone and self-solvate (Steinberg, Elber *et al*. 2008), and that it is not until a number of milliseconds following self-solvation that non-covalent interactions are lost. More recent MD simulations provided by van der Spoel indicate that a biomolecule in the gas phase is capable of forming and increased number of hydrogen bonds relative to a solution environment. Crucial hydrogen bonds between secondary structural elements are preserved in the gas phase whilst additional bonds form between the desolvated side chains which would normally form hydrogen bonds with the solvent. This data suggests that the

structures observed in the gas phase may be kinetically trapped in near solution-phase conditions (van der Spoel, Marklund *et al*. 2011). Further evidence for the preservation of a solution-like structure has been provided by ion-mobility mass spectrometry based experiments. The Bowers group have shown that Ubiqutin studied by drift cell IMS-MS has a similar gas phase CCS compared to its solution CSS. The same studied also indicated that Ubiquitin retained the same conformation in the gas phase for >100 ms following desolvation (Wyttenbach and Bowers 2011).

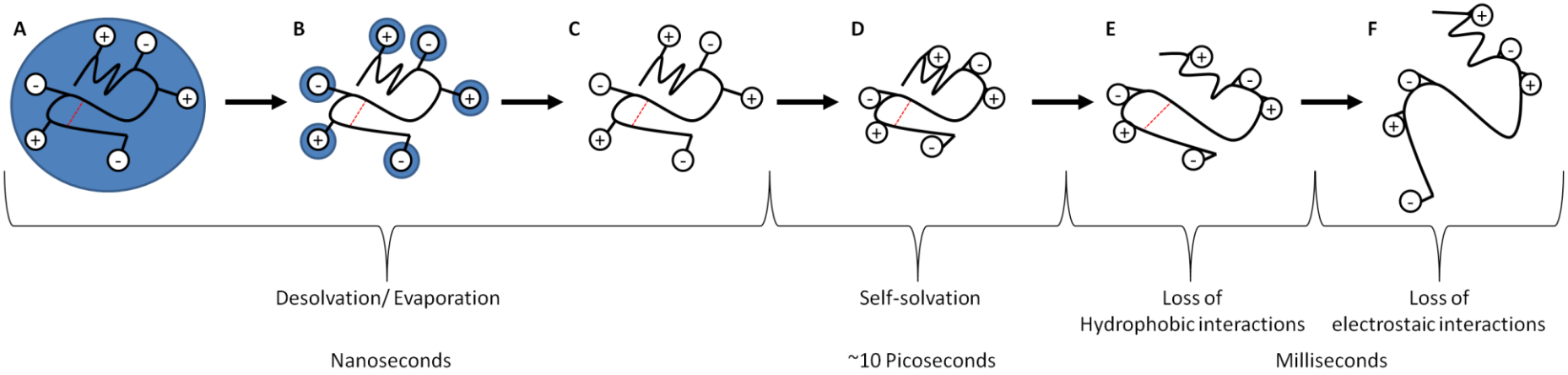


Figure 1-12 - The evolution of gas phase protein structure over time.
A - The solvent droplet containing solution phase structure protein is produced via nano-ESI. Protein is surrounded by water which (B) begins to evaporate as the ion traverses the source region. Water molecules are lost from the bulk of the protein, remaining clustered to the exposed basic sites until (C) eventually all solvent molecules are lost, leaving only a gas phase protein ion in a solution phase structure. All of this occurs within the first few nanoseconds of ionisation. Following desolvation (D) the charged side chains self-solvate, folding onto the protein backbone. This occurs within picoseconds of desolvation. Over the next few milliseconds the protein undergoes successive loss of hydrophobic (E) and electrostatic (F) interactions, forming a stable gas phase structure. Structures A to D can be retained in the gas phase for up to 100 ms (Wyttenbach and Bowers 2011). Figure adapted from (Breuker and McLafferty 2008).

The preservation of these non-covalent interactions are highly dependent on experimental conditions and instrument settings (Skinner, Breuker *et al*. 2013). Optimisation of these parameters is crucial and the effects of these conditions have been explored and summarised by a number of groups.

Small molecule and proteomics applications of MS make use of volatile, organic solvents, such as methanol and acetonitrile, commonly containing 0.1-2% organic acid. These buffers are unsuitable for native MS as they denature all but the most stable proteins. To ensure a protein complex remains in a native state a solution of ammonium acetate, or ammonium bicarbonate, with a pH range of 6 – 8 is typically used for analysis; both ammonium and acetic acid have high enough volatility to allow for the formation of ESI droplets and adequate desolvation (Hernandez and Robinson 2007).

Native protein ions typically have high *m/z* by virtue of their large molecular weight and low charge state. As a result, these ions are less affected by the normal RF methods for radially confining ions during transmission through an MS instrument (Benesch, Ruotolo *et al*. 2007). Ion transmission can be improved by increasing the pressure within the interface of the MS system (Tahallah, Pinkse *et al*. 2001) with different pressures proving optimum for different biological molecules (Douglas and French 1992). This observation was initially attributed to collisional cooling; the increased gas pressure inducing a number of low energy collisions that each served to leach energy from the ion, reducing the internal kinetic energy of the ion and thus preserve the non-covalent interactions. A more recent study by Krutchinsky *et al* has suggested that the transition between the atmospheric pressure of the source and the rough vacuum of the instrument results in the radial expansion of the cloud of ions produced by nESI. The rapid axial and radial acceleration of the ions can cause a change in the ion trajectory, resulting in fewer ions transitioning into the mass analyser (Krutchinsky, Chernushevich *et al*. 1998). Increasing the gas pressures induces a collisional dampening effect whereby ion-neutral collisions reduce the axial and radial kinetic energy of the ions and help better focus the ion beam (Benesch, Ruotolo *et al*. 2007).

Whilst collisional dampening has a positive effect on ion transmission the ion-neutral collisions convert the axial and radial kinetic energy from ion motion into

internal ion energy, effectively increasing the internal ion temperature (Benesch, Ruotolo *et al*. 2007), which can lead to ion activation and fragmentation (Konijnenberg, Butterer *et al*. 2013). Careful control of the ion temperature is vital for any native MS experiment, as over excitation of the ion can lead to fragmentation, whereas over cooling of an ion can lead to ions forming 'frozen' clusters with solvent vapour still within the source (Konijnenberg, Butterer *et al*. 2013). Ion temperature is carefully controlled by a series of acceleration voltages and gas collision, such as those described above.

Evaporative cooling is another major factor contributing to ion internal temperature. Tang and Kebarle observed that the initial temperature of an ESI droplet is always lower than the ambient environmental temperature due to the effects of solvent evaporation (Kebarle and Tang 1993, Tang and Kebarle 1993). MD simulations by Caleman and van der Spoel suggest that solvent evaporation can reduce the droplet temperature dramatically over a short period of time; Caleman and van der Spoeal report a temperature loss of 65 K over 500 ns for droplets starting at an initial temperature of 275 K (Caleman and van der Spoel 2006, Caleman and van der Spoel 2007). Increased starting temperature has also been shown to increase the rate at which the temperature drops (Caleman and van der Spoel 2006).

Native MS experiments are an inherent compromise between ion cooling and excitation. Many publications advocate the use of low source temperatures and low voltages to preserve the non-covalent interactions of a protein/ protein complex (Hernandez and Robinson 2007, Ruotolo, Benesch *et al*. 2008, Kirshenbaum, Michaelevski *et al*. 2010, Salbo, Bush *et al*. 2012), using careful optimisation of the acceleration voltages to dissociate any solvent adducts via CID. Whilst it is possible to obtain well resolved peaks for native MS spectra it should be noted that native MS experiments are not well suited for the observation of accurate mass, and that residual hydration can actually further stabilise the protein structure (Patriksson, Marklund *et al*. 2007, Steinberg, Breuker *et al*. 2007)

Assuming all required experimental parameters are met, it is possible to gain a great deal of information from native MS experiments. The charge state distribution of a native protein can provide insight into the structure and conformation of a protein (Konijnenberg, Butterer *et al*. 2013). Native proteins have been observed to have a

narrower charge state distribution at higher m/z than unfolded proteins (Katta and Chait, 1993); folded proteins have fewer exposed basic sites to protonate as a result of their folded structure. Folding of a protein under different experimental conditions can be tracked by observing changes in charge state distribution, such as the thermal unfolding of heat shock proteins (Benesch, Sobott *et al*. 2003).

Information regarding the stoichiometry and subunit composition of protein complexes has also been obtained by MS using a number of experimental approaches. Using carefully controlled solution conditions it is possible to partially destabilise a complex in solution, whilst retaining non-covalent interactions. The high electric field gradient associated with ionisation, coupled with the stability reduction lost due to removing solvent hydrogen bonding, may result in the complex undergoing dissociation during this process, allowing the observation of subunits in their native-like states (Rostom and Robinson, 1999).

Collisional Activation

Collisional activation, the process of using increased collision voltages to induce dissociation, can also be used to study protein complex assembly. Whilst a number of activation methods can be used, the most documented method has been CID (Benesch, Ruotolo *et al*. 2007). CID activation of protein complexes follows one of two mechanisms; either the complex undergoes symmetric dissociation into equally sized subassemblies or, more commonly, the complex undergoes asymmetric dissociation, losing a single subunit to form a monomer and a "stripped oligomer" (the complex minus one monomer) (Benesch, Aquilina *et al*. 2006). Interestingly, in these instances the charge distribution on dissociation also appears to be asymmetric, with the departing monomer taking many additional charges from the main complex despite its significantly smaller size (Benesch, Ruotolo *et al*. 2007)Collisional activation using the skimmer cone can also lead to collisional activation, and follows a similar mechanism of dissociation as CID activation (Benesch, Ruotolo *et al*. 2007)

Ion Mobility Mass Spectrometry

The underlying principles of IMS-MS, including estimating collision cross sections for native proteins and complexes, has already been discussed in sections **1.1.7**

Travelling Wave Ion Mobility and **1.1.8 Calibration of Ion Mobility data**. Typically IMS-MS is used as a tool to investigate protein conformation at low resolution. A protein is assumed to retain solution phase conformation(s) upon transfer to the gas phase for the timescale of the mobility experiment (ms). IMS-MS can be used to separate conformations based on Ω and m/z, providing information on the conformation ensemble of a protein, represented by arrival time distribution (ATD). Where a protein adopts multiple conformations, these will manifest as a single charge distribution on the m/z scale, but multiple arrival times on the drift time scale.

In the same way that changes in charge distribution are suggestive of unfolding, changes in a molecules ATD can be used to study the effect of solvent, thermal/collisional heating, mutagenesis or ligand binding on the protein conformation.

Hydrogen Deuterium Exchange Mass Spectrometry

As outlined in section **1.2.2 Methods of studying protein structure** HDX is a solution phase technique, which has been used to study changes in protein conformation and dynamics. Much of the early work on HDX was performed using NMR, however since the work by Katta and Chait (Katta and Chait 1991) in the early 1990's HDX has been increasingly coupled to MS. HDX-MS brings together both hydrogen deuterium exchange and mass spectrometry to create an analytical technique sensitive to minor changes in protein conformation, with the speed and broad sample applicability of mass spectrometry (Jaswal 2013). Recent advances have seen the development of automated HDX-MS techniques (Waters 2014) which have seen HDX-MS employed by a number of groups. HDX-MS is typically used to study the effects of ligand binding, protein-protein interactions and proteins with intrinsic disorder (Balasubramaniam and Komives 2013, Jaswal 2013). Whilst observing changes in deuterium uptake at a whole protein level can provide information on conformational change, the crucial step in any HDX-MS experiment is the localisation of the regions of deuterium uptake. This is achieved by means of proteolytic digest of a labelled protein, followed by LC-MS detection of the labelled peptides (Pan, Han *et al*. 2012) Since the LC-MS system operates using protonated buffer a great deal of work has gone into optimising the instrument conditions to

limit the degree of proton exchange that occurs during proteolysis/chromatographic separation. HDX kinetics are well understood, and it has been shown that rates of exchange are lowest at pH 2.5 and 0° C (Englander, 2006). Following the labelling step, samples are quenched to these conditions to reduce any additional exchange with proteated buffer, and to denature the protein prior to proteolysis. Since quench conditions require low pH for reduced enzyme kinetics, Pepsin is used as the proteolytic enzyme of choice since it retains its activity in acidic conditions (Smith, Deng *et al*. 1997, Wales and Engen 2006, Zhang, Chien *et al*. 2010, Konermann, Pan *et al*. 2011) Identification of a number of redundant peptides helps improve the spatial resolution, and supplemental fragmentation in the mass spectrometer can improve spatial resolution up to single amino acid exchange data (Rand, Zehl *et al*. 2009, Landgraf, Chalmers *et al*. 2012). Following the acquisition of peptide/ amino acid resolution exchange data, exchange information can be mapped onto a protein crystal structure to visualise regions of exchange/ change.

1.4 Aims and Objectives

The aim of this project was to use MS based techniques to investigate the structural and dynamics differences in a number of different proteins. The principal method used was native ion mobility mass spectrometry; a technique capable of probing disruptions in the proteins overall conformation. Three protein systems were investigated; prolyl oligopeptidase (Chapter 3) in collaboration with Dr Dean Rea and Dr Vilmos Fulop at the University of Warwick, haemoglobin (Chapter 4) and antibody Fc domains (Chapter 5) as part of a CASE funded research project with MedImmune.

The main aims of the research project were:

1) To use ion mobility mass spectrometry to probe the conformation of prolyl oligopeptidase and other related family enzyme to provide information regarding the mechanisms of substrate binding, and to study changes in the conformation as a result of ligand binding.

2) To use recent improvements in sample preparation and ion mobility mass spectrometry to further our understanding of the structural differences between haemoglobin variants known to cause disease

3) To use a combination of ion mobility mass spectrometry and hydrogen deuterium exchange mass spectrometry to study the conformational differences between four IgG Fc domains engineered for different functions.

1.5 Conference Papers

Edgeworth, M., Phillips, J., Lowe, D., Higazi, D. R., Scrivens, J. H. Rationalizing differences in thermodynamic stability of immunoglobulins using shape selective mass spectrometry and hydrogen/deuterium-exchange. *Proc. 62ndASMS Conf. on Mass Spectrometry and Allied topics, Baltimore, USA*, 2014

Edgeworth, M., Mitchel, L., Phillips, J., Higazi, D. R., Scrivens, J. H. Investigating Structure-Function/Stability Relationships of Therapeutic Antibodies using Shape Selective Mass Spectrometry, *Proc. 61st ASMS Conf. on Mass Spectrometry and Allied Topics, Minneapolis, USA,* 2013

Edgeworth, M., Rea, D., Fulop, V., Scrivens, J. H., Investigating the Structural Dynamics of Prolyl Oligopeptidase Family Enzymes by means of Shape Selective Mass Spectrometry. *Proc. 60th ASMS Conf. on Mass Spectrometry and Allied topics, Vancouver, Canada*, 2012

1.6 Oral Presentations

Edgeworth, M., Mensah-Dika, Y., Radi, K., Lee, N. O., Scrivens, J.H. Shape selective studies of haemoglobin. *Ion Mobility Special Interest Group, Manchester*, 2014

Edgeworth, M., Scrivens. J. H., Shape selective structural studies of monoclonal antibodies. *BMSS32ndAnnual Meeting, Cardiff*, 2011 - Oral Presentation

1.4 References

Anfinsen, C. B., *et al*. (1954). "Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease." J Biol Chem **207**(1): 201-210.

Balasubramaniam, D. and E. A. Komives (2013). "Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(6): 1202-1209.

Banerjee, S. and S. Mazumdar (2012). "Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte." International Journal of Analytical Chemistry **2012**: 40.

Barnett, D. A., *et al*. (1999). "Separation of leucine and isoleucine by electrospray ionization–high field asymmetric waveform ion mobility spectrometry–mass spectrometry." Journal of the American Society for Mass Spectrometry **10**(12): 1279-1284.

Benesch, J. L., *et al*. (2006). "Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies." Chem Biol **13**(6): 597-605.

Benesch, J. L. P., *et al*. (2007). "Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics." Chemical Reviews **107**(8): 3544-3567.

Benesch, J. L. P., *et al*. (2003). "Thermal Dissociation of Multimeric Protein Complexes by Using Nanoelectrospray Mass Spectrometry." Analytical Chemistry **75**(10): 2208-2214.

Berkowitz, S. A., *et al*. (2012). "Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars." Nat Rev Drug Discov **11**(7): 527-540.

Bleiholder, C., *et al*. (2011). "A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method." International Journal of Mass Spectrometry **308**(1): 1-10.

Breuker, K. and F. W. McLafferty (2008). "Stepwise evolution of protein native structure with electrospray into the gas phase, 10−12 to 102 s." Proceedings of the National Academy of Sciences **105**(47): 18145-18152.

Bush, M. F., *et al*. (2010). "Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology." Analytical Chemistry **82**(22): 9557-9565.

Caleman, C. and D. van der Spoel (2006). "Temperature and structural changes of water clusters in vacuum due to evaporation." J Chem Phys **125**(15): 154508.

Caleman, C. and D. van der Spoel (2007). "Evaporation from water clusters containing singly charged ions." Physical Chemistry Chemical Physics **9**(37): 5105-5111.

Campuzano, I., *et al*. (2012). "Structural Characterization of Drug-like Compounds by Ion Mobility Mass Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross Sections." Analytical Chemistry **84**(2): 1026-1033.

Clemmer, D. E. and M. F. Jarrold (1997). "Ion Mobility Measurements and their Applications to Clusters and Biomolecules." Journal of Mass Spectrometry **32**(6): 577-592.

Clowers, B., *et al*. (2005). "Separation of sodiated isobaric disaccharides and trisaccharides using electrospray ionization-atmospheric pressure ion mobility-time of flight mass spectrometry." Journal of the American Society for Mass Spectrometry **16**(5): 660-669.

Cristoni, S. and L. R. Bernardi (2003). "Development of new methodologies for the mass spectrometry study of bioorganic macromolecules." Mass Spectrometry Reviews **22**(6): 369-406.

de Hoffman, E. and V. Stroobant (2007). Mass Spectrometry: Principles and Applications, Wiley.

Dole, M., *et al*. (1968). "Molecular Beams of Macroions." The Journal of Chemical Physics **49**(5): 2240-2249.

Douglas, D. J. and J. B. French (1992). "Collisional focusing effects in radio frequency quadrupoles." Journal of the American Society for Mass Spectrometry **3**(4): 398-408.

Englander, S. W. (2006). "Hydrogen Exchange and Mass Spectrometry: A Historical Perspective." Journal of the American Society for Mass Spectrometry **17**(11): 1481-1489.

Englander, S. W. and N. R. Kallenbach (1983). "Hydrogen exchange and structural dynamics of proteins and nucleic acids." Quarterly Reviews of Biophysics **16**(04): 521-655.

Feng, X., *et al*. (2008). "Mass spectrometry in systems biology: An overview." Mass Spectrometry Reviews **27**(6): 635-660.

Fenn, J. B., *et al*. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**: 64-71.

Garman, E. F. (2014). "Developments in X-ray Crystallographic Structure Determination of Biological Macromolecules." Science **343**(6175): 1102-1108.

Giles, K., *et al*. (2004). "Applications of a travelling wave-based radio-frequency-only stacked ring ion guide." Rapid Communications in Mass Spectrometry **18**(20): 2401-2414.

Guharay, S. K., *et al*. (2008). "Ion Mobility Spectrometry: Ion Source Development and Applications in Physical and Biological Sciences." Plasma Science, IEEE Transactions on **36**(4): 1458-1470.

Guilhaus, M., *et al*. (2000). "Orthogonal acceleration time-of-flight mass spectrometry." Mass Spectrometry Reviews **19**(2): 65-107.

Heck, A. J. R. and R. H. H. van den Heuvel (2004). "Investigation of intact protein complexes by mass spectrometry." Mass Spectrometry Reviews **23**(5): 368-389.

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

Hudgins, R. R., *et al*. (1998). "Design of Helices That Are Stable in Vacuo." Journal of the American Chemical Society **120**(49): 12974-12975.

Iribarne, J. V. and B. A. Thomson (1976). "On the evaporation of small ions from charged droplets." The Journal of Chemical Physics **64**(6): 2287-2294.

Jarrold, M. F. (1998). "Unfolding, Refolding, and Hydration of Proteins in the Gas Phase." Accounts of Chemical Research **32**(4): 360-367.

Jaswal, S. S. (2013). "Biological insights from hydrogen exchange mass spectrometry." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(6): 1188-1201.

Jennings, K. R. (1968). "Collision-induced decompositions of aromatic molecular ions." International Journal of Mass Spectrometry and Ion Physics **1**(3): 227-235.

Jurneczko, E., *et al*. (2012). "Effects of Drift Gas on Collision Cross Sections of a Protein Standard in Linear Drift Tube and Traveling Wave Ion Mobility Mass Spectrometry." Analytical Chemistry **84**(20): 8524-8531.

Kaltashow, I. A. and S. J. Eyles (2005). Mass Spectrometry in Biophysics, John Wiley and Sons Inc.

Kanelis, V., *et al*. (2001). "Multidimensional NMR Methods for Protein Structure Determination." IUBMB Life **52**(6): 291-302.

Kanu, A. B., *et al*. (2008). "Ion mobility–mass spectrometry." Journal of Mass Spectrometry **43**(1): 1-22.

Karas, M., *et al*. (1987). "Matrix-assisted ultraviolet laser desorption of non-volatile compounds." International Journal of Mass Spectrometry and Ion Processes **78**(0): 53-68.

Karas, M. and F. Hillenkamp (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." Analytical Chemistry **60**(20): 2299-2301.

Katta, V. and B. T. Chait (1991). "Conformational-changes in proteins probed by hydrogen-exchange electrospray-ionization mass-spectrometry." Rapid Communications in Mass Spectrometry **5**(4): 214-217.

Kebarle, P. (2000). "A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry." Journal of Mass Spectrometry **35**(7): 804-817.

Kebarle, P. and L. Tang (1993). "From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry." Analytical Chemistry **65**(22): 972A-986A.

Keith, M. J. (2008). Bioinformatics.

Kim, K. S., *et al*. (1993). "Hydrogen exchange identifies native-state motional domains important in protein folding." Biochemistry **32**(37): 9600-9608.

Kirshenbaum, N., *et al*. (2010). "Analyzing Large Protein Complexes by Structural Mass Spectrometry." J Vis Exp(40): e1954.

Konermann, L., *et al*. (2012). "Unraveling the Mechanism of Electrospray Ionization." Analytical Chemistry **85**(1): 2-9.

Konermann, L., *et al*. (2011). "Hydrogen exchange mass spectrometry for studying protein structure and dynamics." Chemical Society Reviews **40**(3): 1224-1234.

Konermann, L., *et al*. (2013). "Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules." Analytical Chemistry **86**(1): 213-232.

Konijnenberg, A., *et al*. (2013). "Native ion mobility-mass spectrometry and related methods in structural biology." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(6): 1239-1256.

Koppenaal, D. W., *et al*. (2005). "MS Detectors." Analytical Chemistry **77**(21): 418 A-427 A.

Krutchinsky, A. N., *et al*. (1998). "Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer." Journal of the American Society for Mass Spectrometry **9**(6): 569-579.

Landgraf, R., *et al*. (2012). "Automated Hydrogen/Deuterium Exchange Electron Transfer Dissociation High Resolution Mass Spectrometry Measured at Single-Amide Resolution." Journal of the American Society for Mass Spectrometry **23**(2): 301-309.

Lanucara, F. H., S. W.; Gray, C. J.; Eyers, C. E. (2014). "The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics." Nature Chemistry **6**(4): 281-294.

Leary, J. A., *et al*. (2009). "Methodology for Measuring Conformation of Solvent-Disrupted Protein Subunits using T-WAVE Ion Mobility MS: An Investigation into Eukaryotic Initiation Factors." Journal of the American Society for Mass Spectrometry **20**(9): 1699-1706.

Lorenzen, K. and E. v. Duijn (2001). Native Mass Spectrometry as a Tool in Structural Biology. Current Protocols in Protein Science, John Wiley & Sons, Inc.

Mack, E. (1925). "AVERAGE CROSS-SECTIONAL AREAS OF MOLECULES BY GASEOUS DIFFUSION METHODS." Journal of the American Chemical Society **47**(10): 2468-2482.

Mamyrin, B. A. (2001). "Time-of-flight mass spectrometry (concepts, achievements, and prospects)." International Journal of Mass Spectrometry **206**(3): 251-266.

Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." Nat Biotech **21**(3): 255-261.

McLafferty, F. W. and T. A. Bryce (1967). "Metastable-ion characteristics: characterization of isomeric molecules." Chemical Communications (London)(23): 1215-1217.

Mesleh, M. F., *et al*. (1996). "Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential." The Journal of Physical Chemistry **100**(40): 16082-16086.

Michaelevski, I., *et al*. (2010). "T-wave Ion Mobility-mass Spectrometry: Basic Experimental Procedures for Protein Complex Analysis." J Vis Exp(41): e1985.

Mirgorodskaya, O. A., *et al*. (1994). "Electrospray-ionization time-of-flight mass spectrometry in protein chemistry." Analytical Chemistry **66**(1): 99-107.

Pan, J., *et al*. (2012). "Top-down hydrogen/deuterium exchange and ECD-stitched FTICR-MS for probing structural dynamics of a 29-kDa enzyme." International Journal of Mass Spectrometry **325–327**(0): 130-138.

Papayannopoulos, I. A. (1995). "The interpretation of collision-induced dissociation tandem mass spectra of peptides." Mass Spectrometry Reviews **14**(1): 49-73.

Patriksson, A., *et al*. (2007). "Protein Structures under Electrospray Conditions." Biochemistry **46**(4): 933-945.

PDB, R. (2014). "PDB Current Holdings Breakdown." Retrieved 28th April, 2014, from http://www.rcsb.org/pdb/statistics/holdings.do.

Pringle, S. D., *et al*. (2007). "An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument." International Journal of Mass Spectrometry **261**(1): 1-12.

Rand, K. D., *et al*. (2009). "Protein Hydrogen Exchange Measured at Single-Residue Resolution by Electron Transfer Dissociation Mass Spectrometry." Analytical Chemistry **81**(14): 5577-5584.

Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.

Ruotolo, B. T., *et al*. (2008). "Ion mobility-mass spectrometry analysis of large protein complexes." Nat. Protocols **3**(7): 1139-1152.

Salbo, R., *et al*. (2012). "Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers." Rapid Communications in Mass Spectrometry **26**(10): 1181-1193.

Saphire, E. O., *et al*. (2002). "Contrasting IgG structures reveal extreme asymmetry and flexibility." Journal of Molecular Biology **319**(1): 9-18.

Scarff, C., *et al*. (2009). "Probing hemoglobin structure by means of traveling-wave ion mobility mass spectrometry." Journal of the American Society for Mass Spectrometry **20**(4): 625-631.

Scarff, C. A., *et al*. (2008). "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 Communications in Mass Spectrometry **22**(20): 3297-3304.

Schweitzer-Stenner, R. (2006). "Advances in vibrational spectroscopy as a sensitive probe of peptide and protein structure: A critical review." Vibrational Spectroscopy **42**(1): 98-117.

Seidler, J., *et al*. (2010). "De novo sequencing of peptides by MS/MS." PROTEOMICS **10**(4): 634-649.

Shelimov, K. B., *et al*. (1997). "Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c." Journal of the American Chemical Society **119**(9): 2240-2248.

Skinner, O., *et al*. (2013). "Charge Site Mass Spectra: Conformation-Sensitive Components of the Electron Capture Dissociation Spectrum of a Protein." Journal of the American Society for Mass Spectrometry **24**(6): 807-810.

Smith, D. L., *et al*. (1997). "Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry." J Mass Spectrom **32**(2): 135-146.

Smith, D. P., *et al*. (2009). "Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies." Eur J Mass Spectrom (Chichester, Eng) **15**(2): 113-130.

Squires, G. (1998). "Francis Aston and the mass spectrograph." Journal of the Chemical Society, Dalton Transactions(23): 3893-3900.

Steinberg, M. Z., *et al*. (2007). "The dynamics of water evaporation from partially solvated cytochrome c in the gas phase." Physical Chemistry Chemical Physics **9**(33): 4690-4697.

Steinberg, M. Z., *et al*. (2008). "Early Structural Evolution of Native Cytochrome c after Solvent Removal." Chembiochem : a European journal of chemical biology **9**(15): 2417-2423.

Stryer, L. (2011). Biochemistry.

Tahallah, N., *et al*. (2001). "The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument." Rapid Communications in Mass Spectrometry **15**(8): 596-601.

Tanaka, K., *et al*. (1988). "Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry." Rapid Communications in Mass Spectrometry(2): 151-153.

Tang, L. and P. Kebarle (1993). "Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution." Analytical Chemistry **65**(24): 3654-3668.

Thalassinos, K., *et al*. (2009). "Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry." Analytical Chemistry **81**(1): 248-254.

Thalassinos, K., *et al*. (2004). "Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer." International Journal of Mass Spectrometry **236**(1–3): 55-63.

Uetrecht, C., *et al*. (2010). "Ion mobility mass spectrometry of proteins and protein assemblies." Chemical Society Reviews **39**(5): 1633-1655.

Valentine, S. J., *et al*. (1997). "Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding." The Journal of Physical Chemistry B **101**(19): 3891-3900.

Valentine, S. J., *et al*. (1997). "Conformer-dependent proton-transfer reactions of ubiquitin ions." Journal of the American Society for Mass Spectrometry **8**(9): 954-961.

Valentine, S. J., *et al*. (2006). "Toward plasma proteome profiling with ion mobility-mass spectrometry." J Proteome Res **5**(11): 2977-2984.

van der Spoel, D., *et al*. (2011). "Proteins, Lipids, and Water in the Gas Phase." Macromolecular Bioscience **11**(1): 50-59.

von Helden, G., *et al*. (1993). "Carbon cluster cations with up to 84 atoms: structures, formation mechanism, and reactivity." The Journal of Physical Chemistry **97**(31): 8182-8192.

Wales, T. E. and J. R. Engen (2006). "Hydrogen exchange mass spectrometry for the analysis of protein dynamics." Mass Spectrometry Reviews **25**(1): 158-170.

Wang, G., *et al*. (2014). "Recent advances in protein NMR spectroscopy and their implications in protein therapeutics research." Analytical and Bioanalytical Chemistry **406**(9-10): 2279-2288.

Waters (2014). ACQUITY UPLC M-Class System with HDX Technology brochure.

Wiley, W. C. and I. H. McLaren (1955). "Time-of-Flight Mass Spectrometer with Improved Resolution." Review of Scientific Instruments **26**(12): 1150-1157.

Williams, J. P. and J. H. Scrivens (2008). "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 Communications in Mass Spectrometry **22**(2): 187-196.

Wilm, M. and M. Mann (1996). "Analytical Properties of the Nanoelectrospray Ion Source." Analytical Chemistry **68**(1): 1-8.

Wilm, M. S. and M. Mann (1994). "Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?" International Journal of Mass Spectrometry and Ion Processes **136**(2–3): 167-180.

Wolynes, P. G. (1995). "Biomolecules folding in vacuo!!!(?)." Proceedings fo the National Academy of Science **92**(7): 2426-2427.

Wong, S. F., *et al*. (1988). "Multiple charging in electrospray ionization of poly(ethylene glycols)." The Journal of Physical Chemistry **92**(2): 546-550.

Wyttenbach, T. and M. T. Bowers (2011). "Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility–Mass Spectrometry Environment." The Journal of Physical Chemistry B **115**(42): 12266-12275.

Wyttenbach, T., *et al*. (2001). "Design of a new electrospray ion mobility mass spectrometer." International Journal of Mass Spectrometry **212**(1–3): 13-23.

Zeleny, J. (1898). "VI. On the ratio of the velocities of the two ions produced in gases by Röntgen radiation; and on some related phenomena." Philosophical Magazine Series 5 **46**(278): 120-154.

Zhang, X., *et al*. (2010). "Dynamics of the β2-Adrenergic G-Protein Coupled Receptor Revealed by Hydrogen−Deuterium Exchange." Analytical Chemistry **82**(3): 1100-1108.

Chapter 2 : Materials and Methods

2.1 Materials

2.1.1 Protein samples

Prolyl oligopeptidase (Prep) and dipeptidyl peptidase IV (DPP IV) were provided by collaborators within the School of Life Sciences. Both samples were supplied at a concentration of 7.5 µM. Prep and DPP IV inhibitor, a ZPP-like inhibitor, was provided at a concentration of 100 mM in 10 mM sodium phosphate, 150 mM sodium chloride.

Samples of fresh whole blood containing heterogeneous normal (HbA), sickle (HbS) and haemoglobin C (HbC) were supplied by University Hospitals Coventry and Warwickshire NHS Trust.

Four fully human intact antibodies were provided by MedImmune (Cambridge, UK). Antibodies were expressed by Chinese hamster ovary cell lines and purified using protein A as the primary capture step, followed by established polishing steps to remove any remaining impurities. IgG-WT, IgG-TM and IgG-YTE were formulated in 50 mM sodium acetate, 100 mM sodium chloride, pH 5.5. IgG-TMYTE was formulated in 25 mM histidine, 7% sucrose, pH 6.0.

Commercially-available lyophilised haemoglobin A and haemoglobin S, cytochrome C (equine heart), serum albumin (bovine), Avidin (egg white), Alcohol Dehydrogenase (*saccharomyces cerevisiae*), β-lactoglobulin (bovine milk) and glutamate dehydrogenase (bovine liver) were purchased from Sigma Aldrich LTD (Gillingham, UK).

2.1.2 Reagents

High-grade chemical reagents, including high purity water and acetonitrile, were obtained from J.T. Baker (Phillipsburg, NJ, USA) and Sigma-Aldrich (Gillingham, UK). Sodium acetate, caesium iodide, sodium iodide, diisoproply fluorophosphate, sodium chloride, sodium phosphate monobasic, sodium phosphate dibasic, endoproteinase lys C (*Lysobacter enzymogenes*) were all supplied by Sigma-Aldrich (Gillingham, UK). Deuterated water was obtained from Cambridge Isotope Labs (Tewksbury, MA, USA).

2.2 Methods

2.2.1 Sample Preparation

Chapter 3: Preperation of Prolyl oligopeptidase for MS analysis

500 µL of Prep and DPP IV were spun at 13,000 rpm for 15 minutes in centrifugal filter units with a 10 kDa cut-off (Amicon® Ultra 0.5 mL 10 K, Millipore, Ireland) to concentrate to a final volume of 50 µL. Sample retained on filter was diluted to 500 µL using 200 mM ammonium acetate. This process of concentration and dilution was repeated a further five times (six spin steps in total) to desalt the protein.

Following incubation with ZPP-like inhibitor, the sample was subject to a further two centrifugal filtration steps, using the same conditions as above, to remove the salts and inhibitor prior to MS analysis.

Chapter 4: Preparation of whole blood samples

Whole blood samples were prepared by diluting 20 µL aliquots to a final volume of 500 µL in 10 mM ammonium acetate pH 6.8. 500µL of protein sample was spun at 13,000 rpm for 15 minutes in centrifugal filter units with a 10 kDa cut-off (Amicon® Ultra 0.5 mL 10 K, Millipore, Ireland) to concentrate to a final volume of 50 µL. Sample retained on filter was diluted to 500 µL using 10 mM ammonium acetate. This process of concentration and dilution was repeated a further five times (six spin steps in total) to desalt the protein.

Chapter 5: Preparation of IgG Fc's for MS analysis.

Proteolytic digestion of mAb to Fc constructs

In order to obtains Fc fragments the intact IgG molecules were subject to partial proteolytic cleavage using endoproteinase lys C. The protocol used was optimised based on protocols published in Lau *et al* (Lau, Pace *et al*. 2010), Masuda *et al* and (Masuda, Yamaguchi *et al*. 2000) Kleeman *et al* (Kleemann, Beierle *et al*. 2008).

IgG1 samples were buffer exchanged into phosphate buffered saline (PBS) – 10mM sodium phosphate, 150 mM NaCl, pH 7.4. *Endoproteinase lys C* was added in a 400:1 IgG: enzyme ratio. The reaction was performed at room temperature and terminated after 30 min with addition of proteinase inhibitor diisopropyl

fluorophosphate (DIFP). DIFP was made up to 1 M and added in a 1:1 ratio (by volume) with the lys C to ensure complete lys C inhibition.

Purification of Fc's from Fc/ Fab mixtures.

Purification of the Fc's were attempted using two different approaches; Protein A extraction and CaptureSelect affinity purification.

CaptureSelect affinity purification

CaptureSelect affinity purification kits contain an immobilised 13kDa llama antibody V-domain which shows binding affinity to a wide variety if IgG Fc domains, including human IgG1, as the affinity matrix.

CaptureSelect affinity spin columns were equilibrated using 10 mM sodium phosphate, 150 mM NaCl at pH 7.4. Fc/ Fab mixture was added to the pre-equilibrated columns and incubated by end-over-end mixing at room temperature for 30 minutes. Fab was eluted from the column by centrifugation. To ensure complete elution of Fab, the column was washed further three times using 10 mM sodium phosphate, 150 mM NaCl, pH 7.4. Fc's were eluted from the column using 200 mM glycine buffer, pH 2.8.

Protein A Purification

Purification was performed using a Freedom EVO Liquid handling robot (TECAN). MabSelect SuRe resin (GE Healthcare Life Sciences), an immobilised Protein A resin, prepacked in MediaScout columns (Attol) was used as the affinity matrix.

Columns were equilibrated with 50 mM sodium phosphate, 150 mM NaCl pH 6.6 prior to sample loading. Unbound Fc was eluted in three PBS washes, and protein non-specifically bound to the resin was released using a detergent wash. Fc was released from the column using 50mM sodium acetate, pH 3.6. The pH of the eluent was adjusted to pH 5.5 using 500mM sodium acetate, pH 6.6 for storage prior to analysis.

Following elution from the columns, Fc fragments were analysed using LC-MS to confirm identity.

IgG Fc [protein A] elutions obtained from endoproteinase digestion were buffer exchanged into 200 mM ammonium acetate, pH 6.8, prior to native MS analysis. 500µL of protein sample was spun at 13,000 rpm for 15 minutes in centrifugal filter units with a 10 kDa cut-off (Amicon® Ultra 0.5 mL 10 K, Millipore, Ireland) to concentrate to a final volume of 50 µL. Sample retained on filter was diluted to 500 µL. This process of concentration and dilution was repeated a further five times (six spin steps in total) to desalt the protein. All centrifugation was performed in a chilled centrifuge set at 4 ^{o}C.

Preparation of calibration standards for ion-mobility mass spectrometry

All lyophilised protein standards used for collision cross section estimation were prepared at concentrations of 1 mg/mL in 200mM ammonium acetate pH 6.8. 500µL of protein sample was spun at 13,000 rpm for 15 minutes in centrifugal filter units with a 10 kDa cut-off (Amicon® Ultra 0.5 mL 10 K, Millipore, Ireland) to concentrate to a final volume of 50 µL. Sample retained on filter was diluted to 500 µL with 200mM ammonium acetate. This process of concentration and dilution was repeated a further five times (six spin steps in total) to desalt the protein.

2.2.2 Sample analysis

Instrumentation

All native MS and IM-MS experimental data was acquired on a Synapt G1 instrument (Waters, Manchester, UK) commercially modified with a 32K quadrupole RF generator to allow selection of high m/z ions. All samples were introduced into the source region of the Synapt G1 by direct infusion using nano-ESI by means of fused silica nanospray needles (Waters, Manchester, UK). The ion mobility separation gas used in for analysis was nitrogen (N_2)

Confirmation of digestion and purification of the Fc was assessed by means of LC-MS. Separation was achieved using an ACQUITY UPLC system (Waters Corporation), with spectra acquired on a Synapt G1 HDMS system (Waters, Manchester, UK) fitted with a standard 8k quadrupole RF generator.

Hydrogen deuterium exchange was performed using a CTC PAL sample handling robot (LEAP Technologies) with online digestion and peptide separation using an

HDX Manager (Waters, Manchester, UK.). Peptides were analysed using a Synapt G2 HDMS instrument (Waters, Manchester, UK).

Calibration of TOF

In all cases, the TOF mass analyser was calibrated using 2 mg/mL caesium iodide in 50% aqueous propan-2-ol over a suitable mass acquisition range, usually 1000-10000 *m/z* for native MS experiments, and 500-5000 *m/z* for denatured MS experiments.

Chapter 3: Mass spectrometry analysis of PREP and DPP IV

The capillary voltage was varied between 1.0-1.1 kV, sampling cone voltage was set to 100V and the source temperature 40^{o}C. The Trap and Transfer acceleration energies were set at 4kV. The backing pressure was maintained between 1.8 and 2 mBar, the Trap DC bias was increased to 30kV to increase desolvation.

Ion mobility separation experimental values were optimised for POP enzymes, and kept constant. Wave velocity was set to 300 m/s, wave height was set to 10 V and IMS gas pressure was set to 0.443 mBar.

During IMS experiments cone voltage was varied between 60 V and 200 V in order to provide activation energy and investigate differences in dynamics.

Chapter 4: Mass spectrometry analysis of haemoglobin variants

For denatured mass spectrometry aliquots of HbA, HbS and HbC from whole blood were diluted 1/1000 in 50 % ACN, 0.5 % HCOOH and infused into the mass spectrometry. Data obtained for each sample were deconvoluted over an appropriate *m/z* range for the species using the MaxEnt1 processing algorithm, within the MassLynx™ (v4.1) software, to provide an estimate of molecular mass.

For native mass spectrometry, instrument acquisition conditions were based on the previously established conditions used by Scarff *et al* (Scarff, Patel *et al*. 2009). The capillary voltage was set to 1.2 kV, sampling cone voltage of 60 V, trap and transfer acceleration voltages set to 10 kV and 4 kV respectively. Source temperature was set to 37^{o}C.

The effects of source pressure on calibration and on estimated CCS/ arrival time were studied and varied between high pressure (6.6-6.8 mBar) and low pressure (3.6-3.7 mBar).

Ion mobility separation experimental values were based on those taken from Scarff *et al* (Scarff, Patel *et al*. 2009) and kept constant. Wave velocity was set to 400 m/s, wave height was set to 18 V and IMS gas pressure was set to 0.542 mBar.

Chapter 5: Mass spectrometry analysis of antibody Fc domains

Confirmation of digestion and purification of the Fc was assessed by means of LC-MS. 2 µg of protein was loaded onto a reverse-phase BEH300 C4 analytical column (2.1 x 50 mm, 1.7 µm). Elution was performed at a flow rate of 0.150 mL/min by increasing the solvent concentration from 5-42% over 18 minutes. Solvent A composition was 0.01% TFA, 0.1% FA in Water, solvent B composition was 0.01% TFA, 0.1% FA in acetonitrile..

For native MS analysis, capillary voltages were set at 1.0 kV. Source temperature was set at 30^{o}C to reduce protein unfolding in the source. Sampling cone voltage was 100 V, extraction cone voltage was set at 10 V, Trap and Transfer acceleration voltages were set at 15 V to assist with collisional cooling. The source pressure was increased to 2.6 mBar. Trap DC bias was optimised to achieve a balance between desolvation and signal intensity, and a value of 10 V selected.

Ion mobility separation values used were optimised for IgG Fc-WT and kept constant for all other mutants. The IMS wave velocity at 300 m/s and wave height was set at 10 V, IMS gas pressure set to 0.437 mBar.

Hydrogen/ deuterium exchange MS

Hydrogen/ deuterium exchange was initiated by 20-fold dilution of IgG1-Fc (10 µM, in protonated aqueous buffer, 10 mM sodium acetate, pH 5.50) into deuterated buffer (10 mM sodium acetate, pD 5.50) at 293 K. After between 10 s to 24 h incubation, hydrogen-exchange was quenched by mixing 1:1 with pre-chilled quench buffer (100 mM sodium acetate, 1.0 M TCEP, 8 M urea) to a final pH of 2.55 at 274 K. Sample was immediately digested using a pepsin-agarose column (Porozyme) and the resulting peptides separated on a C18 column (1x100 mm ACQUITY BEH 1.7

μm, Waters, Manchester, UK.) with a linear gradient of acetonitrile (3-40 %) supplemented with 0.1 % formic acid at 273.5 K. Peptides were analysed using a Synapt G2 mass spectrometer (Waters, Manchester, UK.). Peptides were identified by MSE fragmentation (2 of 3 replicates with 0.3 fragments per amino acid). Key hinge peptides were confirmed using targeted MS/MS fragmentation.

Deuterium incorporation was measured in DynamX (Waters Corp.) by subtracting t0 peptide

2.2.3 Estimation of rotationally averaged collision cross section

Calibration of the data, protein cross section estimation and theoretical cross-section calculations were performed as detailed in 1.1.8 Calibration of Ion Mobility data. Protein standards were analysed under the same experimental conditions as those used for the analyte molecules. Protein calibration standards were chosen based on the drift time through the ion mobility cell. Standards with drift times which bracketed the drift time of the analyte ionswere selected in each case.

For PREP and DPP IV avidin (egg white), alcohol dehydrogenase (*saccharomyces cerevisiae*) and glutamate dehydrogenase (bovine liver) were found to be suitable calibration standards

For Hb variants cytochrome C (equine heart) and serum albumin (bovine) were found to be suitable calibration standards

For the mAb Fc domains cytochrome C (equine heart), β-lactoglobulion (bovine milk), and alcohol dehydrogenase (*saccharomyces cerevisiae*) were used as T-wave calibrants.

2.2.4 Computational Modelling

Generating computationally derived CCS estimates using MobCal.

Four computational methods were used to create CCS models from X-ray diffraction crystal structures; PA, EHSS, TM and PSA. For the PA, EHSS and TM methods MOBCAL, created by the Jarold group, (Mesleh, Hunter *et al*. 1996, Shvartsburg and Jarrold 1996) was used. MOBCAL requires files to be converted from .PDB to .mfj file format for the software, which was achieved using in-house software. The PSA method, introduced by the Bowers group (Bleiholder, Wyttenbach *et al*. 2011)

uses a web-based form to submit structures for analysis. Data files must be submitted as .xyz coordinate files. OpenBabel was used for the conversion of .pdb files to .xyz files. All computationally derived CCS estimates from all methods were obtained using helium (He) as the buffer gas. To compare He computational CCS values with N_2 CCS estimates, He values have been corrected from PA, EHSS, TM and PSA to PA', EHSS', TM' and PSA'. Correction factor was determined by comparing the molecular weight of the analyte protein with an appropriate calibration standard protein from Bush et al, based on the observation that the differences between He and N_2 CCS values obtained from drift cell data is dependent on molecular weight (Bush, Hall *et al*. 2010).

For Chapter 3 - Conformational studies of Prolyl Oligopeptidase family enzymes, a correction factor of 9% has been used for PREP, and a correction factor of 7% has been used for DPP IV.

For Chapter 4 - : Investigating the structural differences in haemoglobin variants, a correction factor of 10% has been used for Hb tetramer, and a correction factor of 12% has been used for the Hb dimer.

Generating the crystal homology model for HDX data output

Homology models for Fc-WT, Fc-TM, Fc-YTE and Fc-TMYTE were built using Discovery Studio software from Biovia (formerly Accelrys Software Inc.). These were constructed from crystal structures of 3AVE (Fc-WT), 3C2S (Fc-TM) and 3FJT (Fc-YTE) which served as templates for the different Fcs. The Trastuzumab (Herceptin) model structure was used as the template for the hinge region (Brandt, Patapoff *et al*. 2010). Models were assembled as shown in Table 2-1. Sequence alignments are shown in Appendix A.

Model name	3AVE	32CS	3FJT	Trastuzumab
Fc-WT	✓			✓
Fc-TM		✓		✓
Fc-YTE			✓	✓
Fc-TMYE		✓	✓	✓

Table 2-1 – A table showing the different crystal structures used to create the Fc homology models.

Multiple sequence alignments were made between template and target sequences using ClustalW and are shown in the appendix. 100 homology models were generated with a random seeding of deviation from initial atomic co-ordinates. Each model was energy minimised and two scores were calculated: DOPE and PDF N-total. Models were ranked by PDF N-total and then DOPE and the top model for each mutant was selected.

HDX data fitting

Data normalisation was calculated between mutant and wild type data sets using in-house software written by Dr Jonathan Phillips (MedImmune, Cambridge, UK) in MatLab (Mathworks) and Python. Structural representations were generated with PyMol (Schrödinger LLC) and plots in Figure 5-33 - 5-37 prepared with Prism (GraphPad). Hydrogen/deuterium-exchange was represented in Figure 5-32 by calculating the mean deuteration level per amino acid, according to Equation 2-1.

$$\bar{M}_j = \frac{1}{n}\sum_1^n \frac{1}{q}\sum_0^t (m_i^t - m_i^0)$$

Equation 2-1

$\bar{M}_j$ is the mean deuteration level at amino acid j, n is the number of overlapping peptides, q is the number of exchangeable amides for peptide species i, m_i^t is the isotopic weighted midpoint at time t and m_i^0 is the midpoint at time 0 (undeuterated).

2.4 References

Brandt, J. P., *et al*. (2010). "Construction, MD Simulation, and Hydrodynamic Validation of an All-Atom Model of a Monoclonal IgG Antibody." Biophysical Journal **99**(3): 905-913.

Kleemann, G. R., *et al*. (2008). "Characterization of IgG1 Immunoglobulins and Peptide−Fc Fusion Proteins by Limited Proteolysis in Conjunction with LC−MS." Analytical Chemistry **80**(6): 2001-2009.

Lau, H., *et al*. (2010). "Investigation of degradation processes in IgG1 monoclonal antibodies by limited proteolysis coupled with weak cation-exchange HPLC." J Chromatogr B Analyt Technol Biomed Life Sci **878**(11-12): 868-876.

Masuda, K., *et al*. (2000). "Pairing of oligosaccharides in the Fc region of immunoglobulin G." FEBS Letters **473**(3): 349-357.

Scarff, C., *et al*. (2009). "Probing hemoglobin structure by means of traveling-wave ion mobility mass spectrometry." Journal of the American Society for Mass Spectrometry **20**(4): 625-631.

Chapter 3 - Conformational studies of Prolyl Oligopeptidase family enzymes

3.1 Introduction

There are a number of reasons why a protein or peptide may be digested from polypeptide to individual amino acids; many proteins are broken down as a result of digestion of food matter; some intracellular proteins are only permitted within cells for a small period of time before degradation; defective proteins or peptides, created as a result translation errors or protein misfolding, are sometimes targeted for destruction before even entering the cytosol (Stryer 2011). All proteins are digested by proteolytic enzymes known as proteases, commonly referred to as peptidases.

Approximately 2% of all proteins are peptidases (Rea and Fulop 2006). It was initially believed that these peptidases only played a role in nonspecific digestion of ingested protein matter, but it has since become apparent that peptidases can also play a pivotal role in protein post-translational modifications. Peptidases have been implicated in a number of biological processes including fertilisation, embryonic development, cell cycle regulation, immune response and apoptosis (Stryer 2011).

Peptidases, like many enzymes, may be subdivided into families based on sequence homology. Families may then be further subdivided into clans based on shared tertiary structural motifs, suggestive of a common ancestral heritage. Enzymes may also be given a tag relating to their catalytic mode of action; there are six tags relating to four mechanisms of peptidase catalytic activity (A, C, M, S for the aspartic, cysteine, metallo and serine mechanisms, respectively), a mixed catalytic type (P) and an unknown type (U) (Barrett and Rawlings 1995).

Prolyl oligopeptidase (POP) family enzymes are clan SC peptidases that share a conserved catalytic domain which adopts an α/β-hydrolase topology; a protein fold found in a number of enzymes including esterases, lipases, thioesterases and hydroxynitrile lyases (Holmquist 2000). The active site forms around a catalytic triad

compromised of an aspartate (Asp), histidine (His) and serine (Ser). This catalytic triad is also found in at least four other serine proteases, suggesting a conserved catalytic mechanism.

Prolyl oligopeptidase, formally known as prolyl endopeptidase (PREP), is a key enzyme in the POP family and has been the subject of numerous studies. It is larger than other serine peptidases; PREP is an 80 kDa protein whereas trypsin and subtilisin, both serine proteases, are only 25 – 30 kDa. PREP is only capable of digesting short amino acid sequences of approximately 30 residues in contrast with, for example, trypsin which acts on whole proteins (Rea and Fulop 2011). Selective digestion on these short peptides is determined by the size exclusion of larger peptides and proteins from the active site, (Fülöp, Böcskei *et al*. 1998) discussed in more detail below. PREP has also been implicated in a number of diseases. Whilst PREP is widely distributed across the body, it is particularly concentrated in the brain and has been linked to several psychiatric disorders, including depression, mania and bulimia as well as neurodegenerative disorders including Parkinson's and Alzheimer's (Rea and Fulop 2011). PREP has been identified as a potential drug candidate, and has been extensively characterised structurally.

The crystal structure of porcine PREP (PDB:1QFM) is shown in Figure 3-1.

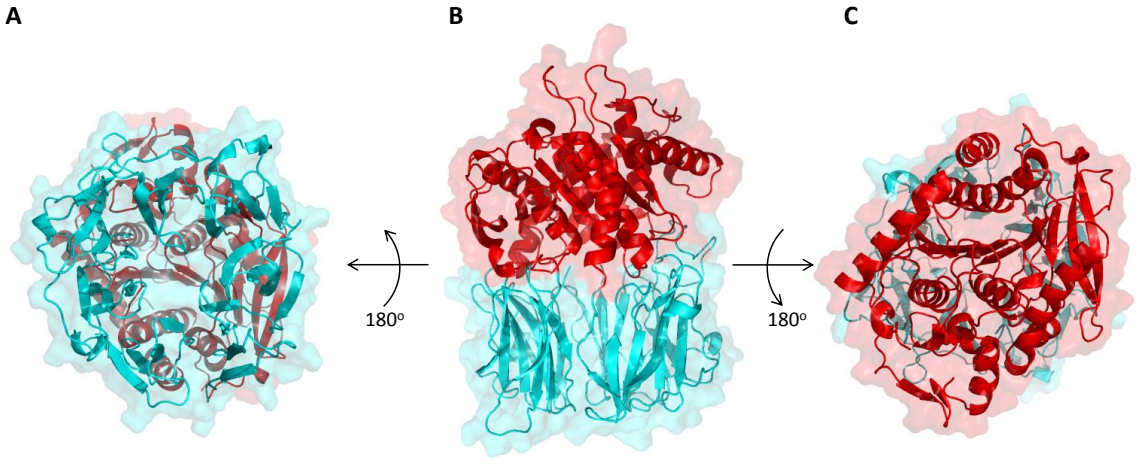


Figure 3-1 - The crystal structure of porcine Prolyl Endopeptidase (PDB 1QFM).
β-propeller domain is highlighted in blue, catalytic α/β beta-hydrolase domain is highlighted in red. The active site is located in a central cavity between the two domains. Access to the active site is therefore size restricted, with only short peptide sequences able to gain access to the active site.

PREP is an 80.7 KDa monomeric protein comprised of two discrete domains; residues 1-72 and 428-710 make up the α/β-hydrolayse catalytic domain whilst residues 73-427 make up a 7-bladed β-propeller (Rea and Fulop 2006). β-propeller domains are found in a number of proteins, and are formed by a repeating four-stranded antiparallel β-sheet motif that comes together to form the propeller. Many proteins with this motif have evolved a manner in which to close the full circle of the propeller (known as the Velcro of the propeller). In G-proteins the N- and C-terminal strands come together to form a single blade of the propeller, creating a number of H-bonds which close the circle (Fülöp and Jones 1999). PREP was the first protein discovered that exhibited an open-topology; the only interactions stabilising the first and last blade were hydrophobic interactions between blades 1 and 7 (Figure 3-2). This has been commonly referred to as open-Velcro topology. No descriptions of the specific hydrophobic interactions have been described (Fülöp and Jones 1999).

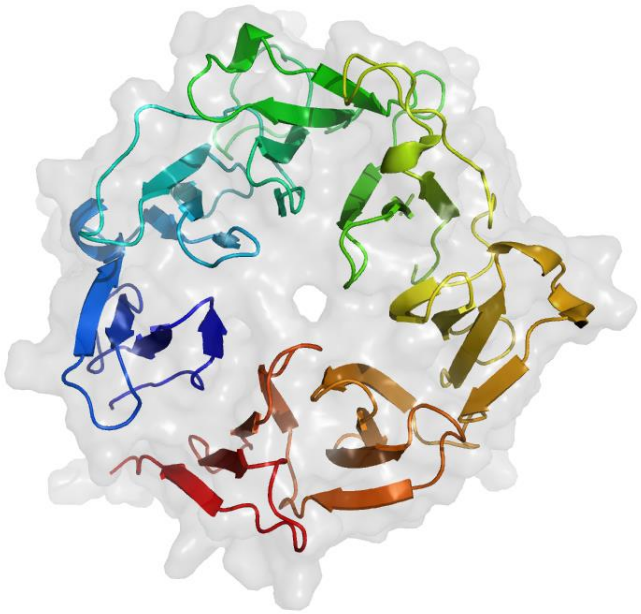


Figure 3-2 - The open topology of the PREP β-propeller. The first (N-terminal) blade, shown in blue, is held together with the seventh (C-terminal) blade, shown in red, by virtue of hydrophobic interactions only.

The active site residues lie in a central cavity located between the α/β-hydrolase and β-propeller domains. This cavity is approximately 8500 $Å^2$ in size (Rea and Fulop 2011) making it able to accommodate large peptides. Despite this, PREP's distinguishing catalytic feature is its ability to digest smaller peptides of approximately 30 residues. The presence of the β-propeller protects the active site

and acts to prevent larger, or more structured, peptides accessing the active site. There is a small opening at the base of the propeller, shown in Figure 3-2, of approximately 4 Å in diameter, and it was initially suggested that this pore, although too small to allow a small peptide (typically 6-12 Å in size) through, could widen to allow substrate passage into the active site. This was consistent with previous studies on PREP kinetics which suggested a conformational change was the rate-determining step in PREP catalysis (Polgar 1991, Polgar 1992, Szeltner, Rea *et al*. 2004). Given the unique topology of the β-propeller in PREP, it was initially believed that the separation of the first and seventh blades could widen the pore size enough to allow substrate entry (Fülöp, Böcskei *et al*. 1998). Mutagenesis experiments which introduced disulphide bridges between blades one and seven supported this theory, indicating that, under reducing conditions, the mutant enzyme was active, but it could be selectively deactivated by employing oxidising conditions (Polgar 1992).

This mechanism was commonly accepted until 2005 when a bacterial PREP was crystallised in an open conformation (Figure 3-3) (Shan, Mathews *et al*. 2005), suggesting that an alternative mechanism of substrate entry may be possible.

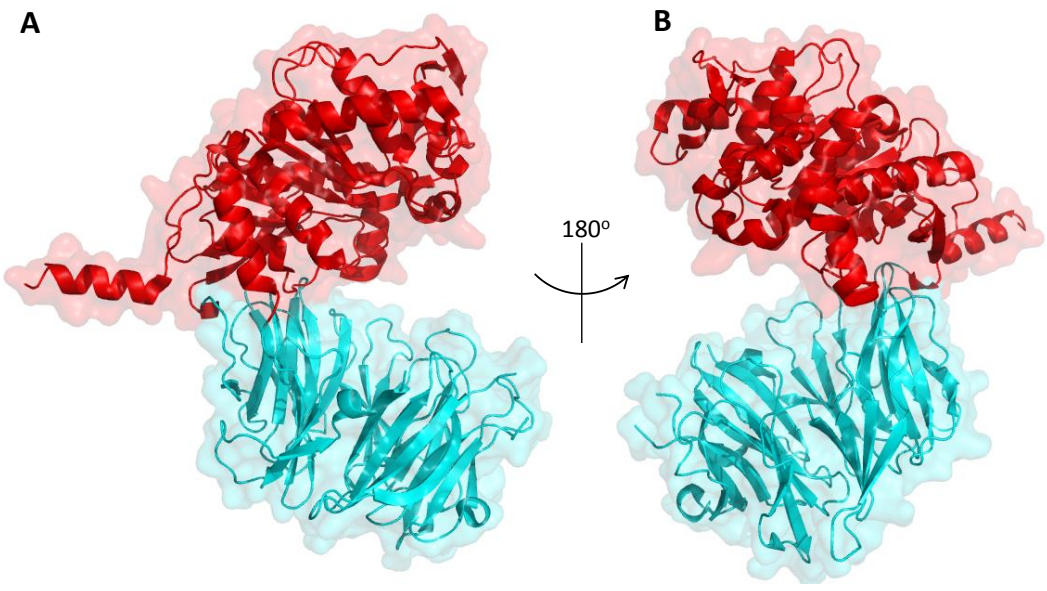


Figure 3-3 – The crystal structure of PREP in the open topology, from PDB file (1YR2)

In this mechanism, the loops which connect the two domains are thought to act as a hinge, allowing the two domains to separate for substrate entry and refold following ligand binding. Interestingly, it was observed that PREP in complex with a substrate-

homologue could not be crystallised in an open conformation, suggesting an induced conformational stability following ligand binding. In a study by Li *et al* (Li, Chen *et al*. 2010)it was shown that PREP, crystallised in an open conformation, could be made to adopt the closed conformation upon treatment of the crystals with substrate and inhibitor, strongly supporting an induced fit mechanism. Additional information provided by NMR studies suggests that both these open and closed structures exist in equilibrium in solution, supporting a mechanism based on domain separation (Kichik, Tarragó *et al*. 2011)

An alternative mechanism of substrate entry has also been suggested, based on molecular dynamic (MD) simulations. MD simulations indicated that the mechanism involving blades one and seven of the β-propeller moving apart to widen the propeller pore is not favoured given the stability of the interactions (Fuxreiter, Magyar *et al*. 2005); this was corroborated by conventional biochemical techniques, in which the β-propeller domain was purified independently of the α/β-hydrolase domain and found to be stable even when isolated (Juhász, Szeltner *et al*. 2005). MD simulations suggest that there is a potential for substrate entry via a tunnel in the inter-domain region of PREP: small loop movements in residues 1-72 of the α/β-hydrolase domain and residues 192-205 of the β-propeller domain may create a tunnel capable of allowing a peptide into the active site (Kaszuba, Rog *et al*. 2009). This mechanism was supported by electron microscopy (EM) data showing a small side opening near the regions identified by the MD calculations (Bastos, Grellier *et al*. 2005, Fuxreiter, Magyar *et al*. 2005).

The mechanism of substrate entry may be considered in the context of other POP family proteins, such as dipeptidyl peptidase IV (DPP IV). DPP IV belongs to the POP subfamily S9b (Rea and Fulop 2006). Like PREP, DPP IV has raised a degree of pharmaceutical interest due to its implications in canerogenesis (Busek, Malik *et al*. 2004). DPP IVs specific physiologic role is difficult to determine due to the number of DPP IV-like enzymes (Rea and Fulop 2006). It is believed to play a role in peptide degradation and amino acid scavenging (Tiruppathi, Ganapathy *et al*. 1990, Suzuki, Erickson *et al*. 1993), T-cell signalling (Ansorge, Buhling *et al*. 1997, Kahne, Lendeckel *et al*. 1999), and chemotaxis (Ludwig, Schiemann *et al*. 2002). DPP acts as a peptidase, removing X-Pro dipeptides from the N-terminus of peptides containing Pro as the penultimate residue (Cunningham and O'Connor 1997).

Unlike PREP, DPP IV is a homodimer (Elovson 1980) is post-translationally modified with an N-linked glycan (Aertgeerts, Ye *et al*. 2004) and exists in one of two forms; a soluble protein and an integral membrane protein (Durinx, Lambeir *et al*. 2000)). Each monomeric subunit of DPP IV is marginally larger than that of PREP; DPP IV monomers consist of 766 residues compared to 710 residues in the PREP monomer (Rea and Fulop 2006). Structurally PREP and DPP IV monomeric units show a high degree of homology. Each subunit consists of an α/β- hydrolase domain containing the catalytic triad (Ser630, Asp 708, His 740) and a β-propeller domain (Figure 3-4), with the active site located between the two domains.

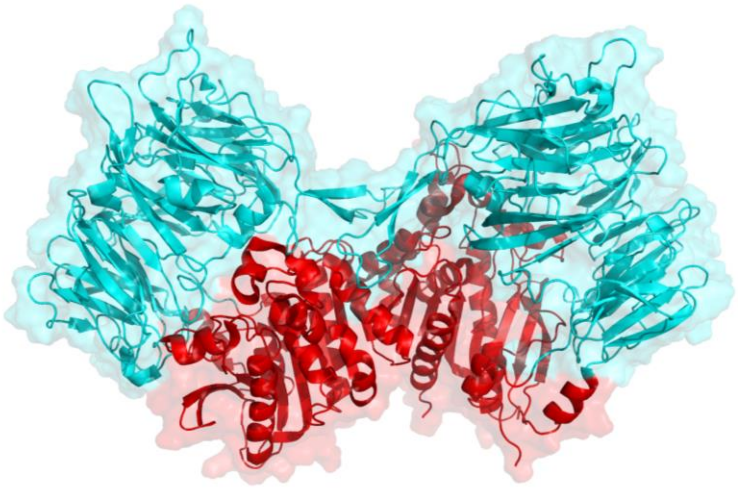


Figure 3-4 – The crystal structure of the dimeric DPP IV; the β-propeller domain is highlighted in blue, catalytic α/β beta-hydrolase domain is highlighted in red. Created from PDB: 1NU8

Despite the conserved open-Velcro topology of the DPP IV propeller domain, it contains a number of differences when compared to the PREP propeller domain. These are illustrated in Figure 3-5.

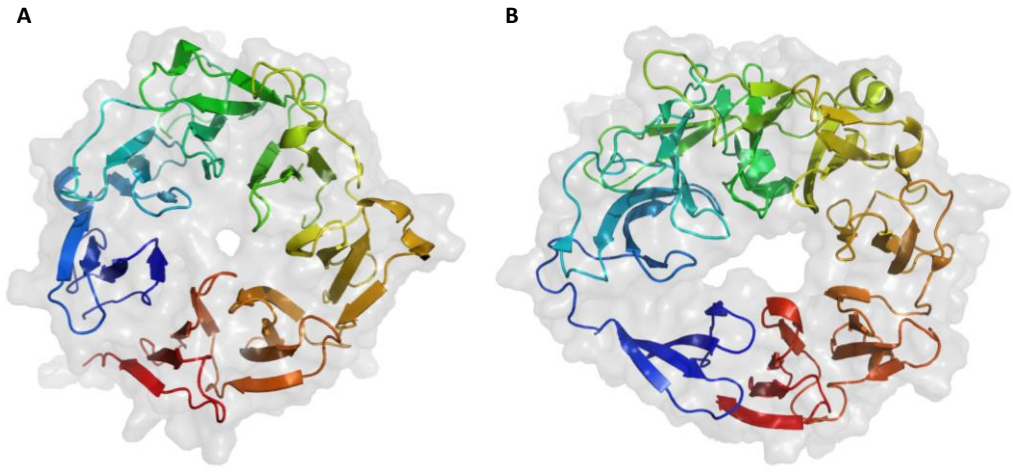


Figure 3-5 – The crystal structures of A) the 7-bladed PREP propeller and B) the 8-bladed DPP IV propeller, highlighting the structural differences between the two related proteins.

The DPP IV propeller domain consists of eight blades rather than the seven found in PREP. This extra blade results in the DPP IV adopting a more significantly ellipsoid shape compared to that seen in PREP, highlighted in Figure 3-5, which leads to a larger pore being formed in the centre (7 Å in width, 14 Å in length). DPP IV shows pronounced bending of blade 1 (dark blue in Figure 3-5) and more subtle bending of blades 2 – 4 (cyan, light green and forest green) which may lead to a larger side opening between the domains. This increase in pore size supports the hypothesis that substrate entry may occur through one of these openings; consistent with a DPP IV crystal structure showing an interaction between the side opening and a substrate (Engel, Hoffmann *et al*. 2003)There is currently no clear understanding as to through what pore substrates may enter and exist the active site.

The observation of a POP family enzyme capable of allowing substrate access to the active site through either the propeller pore, or via a side channel, suggests that the potential for this mechanism may occur in PREP, adding to the discussion surrounding the possible mechanisms of substrate entry in PREP. Whilst X-ray diffraction is an established method for studying protein structure, it is clear that additional experimental approaches are required to provide additional data. The aim of this study was to utilise TWIM-MS to probe the solution phase structure of

mammalian PREP, both isolated and in conjugation with a ligand analogue, to identify if mammalian PREP was capable of adopting both open and closed conformations in solution, and to investigate potential conformational changes induced as a result of ligand binding. DPP IV was analysed by the same methods in order to confirm that it is only capable of adopting a single, closed, conformation in solution supporting the suggested mechanistic divergence in substrate entry between PREP and DPP IV.

3.2 Results and Discussion

3.2.1 Prolyl Endopeptidase

Mass Spectrometry

Figure 3-6 shows the native-MS spectra of PREP and PREP in conjugation with ZPP-like ligand at a cone voltage of 100V.

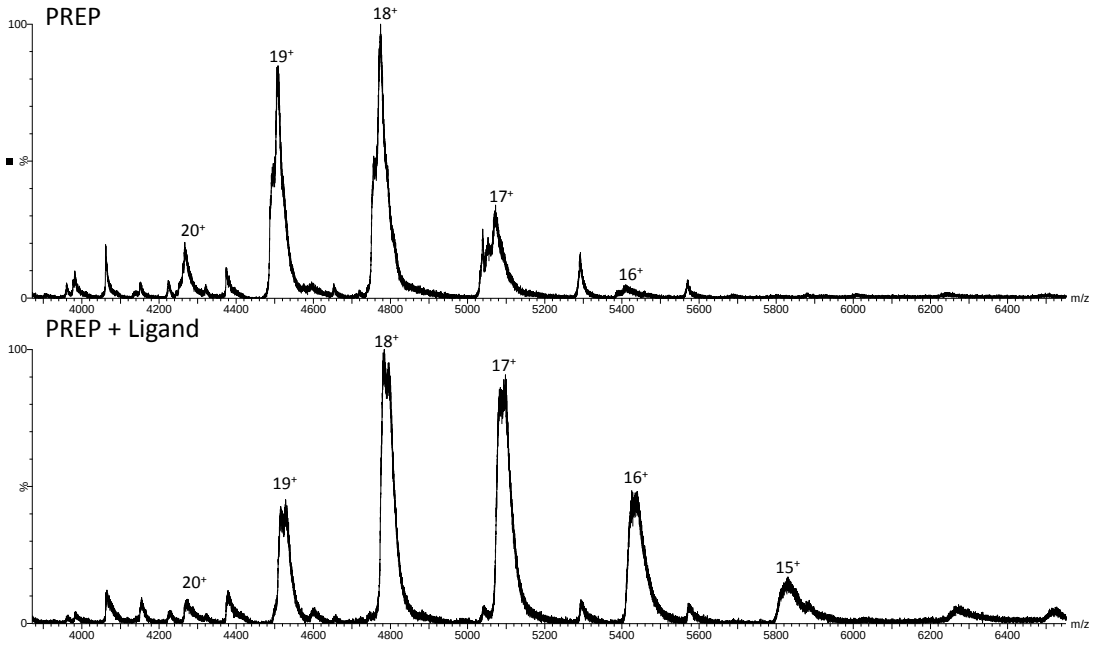


Figure 3-6 - The MS spectra of PREP and PREP with ligand (cone voltage 100 V).
Binding of the ligand by PREP causes a shift in the charge state distribution, suggesting either a change in conformation of PREP, or the blocking of a protonation site in PREP.

The spectra in Figure 3-6 show relatively broad peak widths of 25 amu and 34 amu for PREP and PREP with ligand, respectively, indicating a high level of solvent adducting on the protein. This indicates the MS settings and sample preparation would benefit from further optimisation. The inhibitor was provided in a high salt

content buffer and the observed increase in the peak width may be attributed to increased salt adducts in the ligand-bound PREP spectrum. Both spectra show evidence of two species of similar molecular weights, based on the overlapping charge distributions. For PREP, the two species have molecular weights of 85649 Da and 85924 Da whereas for PREP with ligand the species have molecular weights of 86230 Da and 86499 Da. This increase of approximately 270 Da between species one and two is suggested to be due to of incomplete desolvation of the protein.

The expected theoretical mass of PREP based on the sequence is 80769.5 Da, which differs from the observed molecular weight of 85649 Da by approximately 4800 Da, which equates to 6% of the total mass. Sobott and Robinson (Sobott and Robinson 2004) observed a mass difference of 4000 Da for GroEL. This mass increase was, however, only 0.5% of the total GroEL molecule (GroEL molecular weight is approximately 800,000 Da) which is much smaller than the 6% observed here. A possible explanation for the increase in mass may result from the proteins natural structure. PREP has been shown to have an internal cavity of approximately 8500 $Å^2$ in size (Rea and Fulop 2011). It is possible that this cavity remains solvated following elevation to the gas phase, and the retained solvent contributes to the observed mass increase. An internal cavity of 8500 $Å^2$ PREP is could potentially retain enough solvent to contribute to the 4800+ Da mass increase observed.

Following ligand binding there is an observed mass increase of 581.4 Da for the first species and 575.4 Da for the second species. The mass of the ligand is 380.44 Da. The observed mass increase in consistent with the binding of ligand, with the remaining mass proposed to be the binding of extra salts/ solvent within the binding cavity.

It has been proposed that the PREP in solution exists in equilibrium between its open and closed states, and that substrate binding shifts this equilibrium towards the closed state (Li, Chen *et al*. 2010). Figure 3-6 shows that PREP and PREP with ligand share the $20^+ \rightarrow 16^+$ charge states, suggesting that a population of the proteins share a similar conformation. Addition of ligand causes a shift in the charge state distribution of the protein from the $16^+ \rightarrow 20^+$ in the ligand-free state to the $15^+ \rightarrow 20^+$ in the ligand-bound state. It has been shown that a proteins charge state distribution can be related to its conformation (Konijnenberg, Butterer *et al*. 2013)

with higher charge states relating to a more open conformation (Katta and Chait 1993) arising from more available protonation sites. It is therefore possible that the binding of the ZPP-like inhibitor results in a conformational change that disrupts the available protonation sites. The ZPP-like inhibitor binds to the catalytic serine of PREP (Kaszuba, Rog *et al*. 2009). Serine is a polar, non-charged amino acid and therefore not a proton acceptor in solution. Inhibitor binding to this serine would not result in a change in charge distribution by removing one of the protonation sites of sites. It is possible, particularly given the poor MS resolution, that multiple copies of the ligand are binding the protein at other sites which may affect the charge distribution. Whilst this shift supports the theory that binding of ligand may induce a conformational change in the protein, further evidence is required to strengthen the theory.

Ion mobility mass spectrometry

Figure 3-7 shows the arrival time distributions for PREP protein and PREP with ligand analysed using a low cone voltage activation energy of 60 V for the charge states observed in both species (the $16^+ \rightarrow 20^+$ charges). Drift time chromatograms presented here are obtained from summation of the whole peak, including solvent adducts, since there was insufficient MS resolution to separate the peaks into separate species.

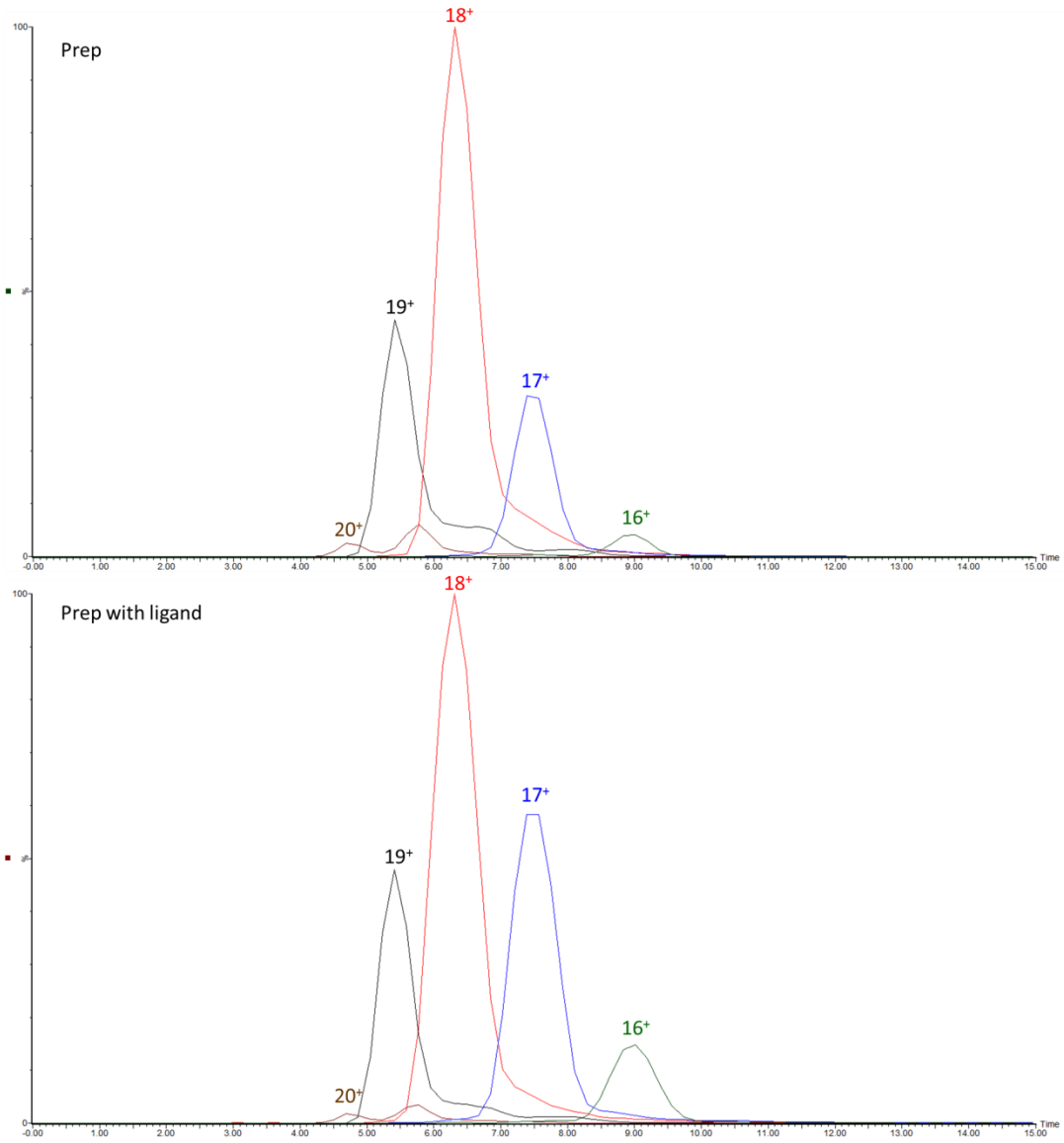


Figure 3-7 – Arrival time distributions for PREP and PREP + ligand (cone voltage 60 V). 16^+ (green), 17^+ (blue) and 18^+ (red) charge states show only a single conformation in both PREP and PREp with ligand. 19^+ (black) and 20^+ (brown) charge states show evidence of an additional extended conformation, which appears at a higher relative intensity in the PREP with ligand ATD compared to PREP.

Under low energy activation conditions both PREP, and PREP with ligand show only a single observed conformation at charge states $16^+ \rightarrow 18^+$. This observation may suggest that PREP substrate entry occurs via small loop movements, rather than via large domain separation. Lower activation energies lead to less desolvation (Sobott and Robinson 2004), and it has been observed that additional solvent adducts can stabilise native protein structures (Steinberg, Breuker *et al*. 2007, Patriksson, Marklund *et al*. 2007). It is therefore possible that the additional solvent adducts are helping to stabilise the single conformation. Figure 3-7 shows that, with increasing charge, both PREP and PREP with ligand can adopt an additional more extended

conformation consistent with a more open conformation. As with Figure 3-6, Prep shows a higher relative intensity of the 19^+ and 20^+ conformations, suggesting it may more readily unfold compared with Prep with ligand. The intensity of the second, extended conformation of the 19^+ and 20^+ charge states is also higher in Prep when compared with Prep with ligand, again suggesting that PREP with ligand may less readily unfold. This is consistent with previous observations of PREP with ligand favouring a more closed conformation compared to PREP (Li, Chen *et al*. 2010).

Samples were analysed using IMS at a higher cone voltage of 100 V to investigate these perceived stability differences, the result are shown in Figure 3-8. Drift time chromatograms presented here are obtained by summation of the whole peak, including solvent adducts, since there was insufficient MS resolution to separate the individual peaks

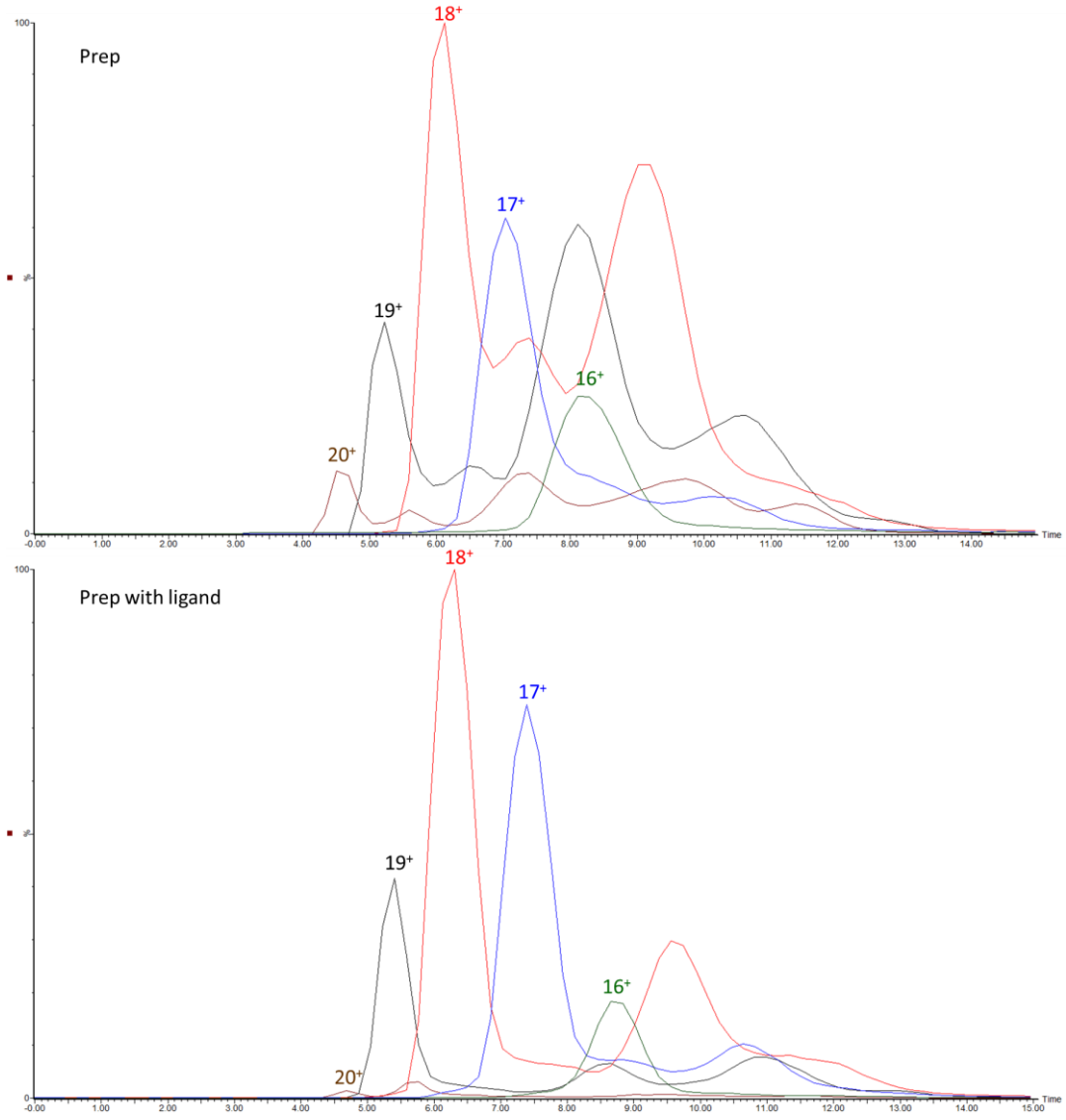


Figure 3-8 – Arrival time distributions for PREP and PREP + ligand (cone voltage 100 V).

16^+ (green) charge states show only a single conformation in both PREP and PREP with ligand. 17^+ (blue) and 18^+ (red) 19^+ (black) and 20^+ (brown) charge states show evidence of additional extended conformations, which appears at a higher relative intensity in the PREP with ligand ATD compared to PREP.

The MS spectra shown in Figure 3-6 indicates that even at this relatively elevated energy level there is still a high degree of solvent still bound to the protein. At the lowest charge state, 16^+, there appears to be only a single conformation, in agreement with the low energy data from Figure 3-7. The lowest charge state of a protein observed by mass spectrometry is thought to be most representative of the native, solution phase structure (Grandori 2003, Samalikova and Grandori 2003). Both PREP and PREP with ligand show evidence of only one conformation at both low (60V) and high (100V) cone voltages for the 16^+ charge state, suggesting that mammalian PREP may only have a single conformation in solution.

With increasing charge a more extended conformation appears in both the PREP and PREP with ligand samples for the $17^+ \rightarrow 18^+$ charge states. As with the 60 V data, this extended population is observed at an increased intensity in the PREP sample compared to PREP with ligand sample. At elevated cone voltage the trend is even more pronounced; PREP with ligand shows predominantly compact conformation with the major charge state presenting at 18^+. Prep without ligand shows and significantly more intense extended conformations for the 18^+, 19^+ and 20^+ charge states compared to Prep with ligand. Proteins are known to unfold as a factor of charge in the gas phase as a result of columbic repulsion (Thalassinos, Slade *et al*. 2004). The higher relative intensity of the extended conformations observed for the PREP sample suggests the protein is able to more readily adopt an extended conformation in the gas phase at the same change states when compared to ligated Prep. Similar to the 60 V IMS data, this suggests a stability difference between the two samples, supporting the observations made by Min Li (Li, Chen *et al*. 2010) that attachment of the ligand may induce more compact conformation.

The 17^+ and 18^+ charge states show evidence of an intermediate peak located between the compact and extended conformations, suggesting an additional conformation of PREP present in both ligand-free and ligand-bound states. This conformation could represent a partially extended closed PREP structure, consistent with a minor degree of protein unfolding, or may represent a solvent stable intermediate between the open and closed structures shown in Figure 3-1 and Figure 3-3. At the higher charge states a number of different conformations become apparent as the protein beings to unfold and form a number of stable gas phase structures.

Figure 3-6 shows two potential species of PREP which were proposed to be different solvated forms of the protein. IMS was used to study these separate species and identify if they correspond to the same, of different species, based on their drift time. Under IMS conditions the MS spectrum showed reduced resolution, making it difficult to fully separate the different species observed in the PREP sample (Figure 3-9). The selection of different section of the board MS peak, shown in Figure 3-9, allowed for the analysis of the different solvated states of PREP. By performing this analysis the different species were pseudo-resolved from broad MS peak.

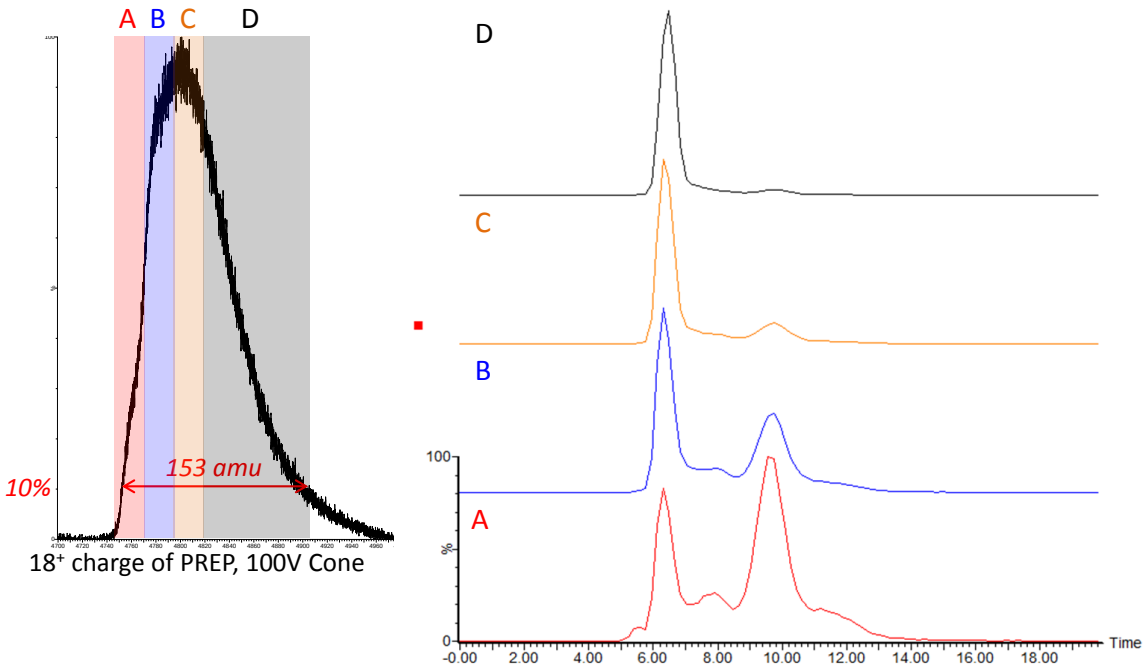


Figure 3-9 – An analysis of the different protein conformations in the broad 18^+ charge state of PREP. Sections A, B, C and D represent differently solvated forms of the same charge state. Comparison of the arrival time distributions suggests that as the amount of solvent adducts increase, the relative intensity of the extended conformation decreases.

Closer inspection of the 18^+ charge state of PREP, shown in Figure 3-9, suggests that it is made up of four distinct regions labelled A → D, where A is proposed to be related to the 85649 Da species, and B the 85924 Da species. C and D represent additional solvated regions. Comparison of the arrival time distribution presents by sections A, B, C and D allow the effects of excess solvent and salt adducts on the conformational distribution to be investigated. The lowest molecular weight species, A, seems to be capable of adopting three major conformations; a compact, an intermediate and an extended conformation. As the amount of solvent adduct increases (B and C), the intensity of the intermediate and extended conformations decrease until there is only compact conformation remaining (D). It has been observed that excess solvent can stabilise the native conformation of a protein (Patriksson, Marklund *et al.* 2007, Steinberg, Breuker *et al.* 2007). The data presented here suggests that the presence of solvent adducts can alter protein conformation, with additional solvent favouring a more compact, native conformation. The observation of only a single conformation of PREP at more highly solvated states (Figure 3-9, insert D) supports the theory that PREP in solution can only adopt a single conformation.

CCS Estimation

Estimated CCS values obtained experimentally have been compared with those calculated from PDB files. The calibration curves for the estimation of CCS from experimental data are shown in Figure 3-10, and were acquired at a 100 V cone voltage.

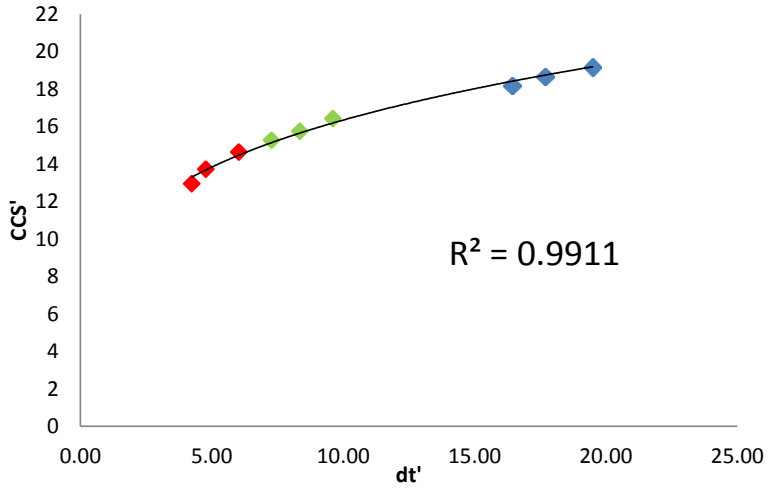


Figure 3-10 – The calibration curve for the estimation of the CCS of PREP and PREP with ligand. The data points refer to avidin (red), alcohol dehydrogenase (green) and glutamate dehydrogenase (blue) and bracket the arrival time window of both PREP and PREP with ligand samples.

The calibration curve shown in Figure 3-10 shows an R^2 value of >0.99 (Scarff, Thalassinos *et al*. 2008). PDB entries 1QFM (Figure 3-1) and 1YR2 (Figure 3-3) were used for estimation of protein cross sections for the unligated PREP samples. Collision cross sections were calculated using projection approximation and exact hard sphere scattering methods. Following CCS estimate, collision cross sections were corrected to account for the differences between N_2 and helium buffer gases. Computational CCS values were increased by 9%, based on the observed differences between He and N_2 CCS value of BSA and concanavalin A obtained experimentally in Bush *et al* (Bush, Hall *et al*. 2010).

A comparison between experimentally estimated CCS values and the computationally calculated CCS values for PREP and PREP + ligand are shown in Figure 3-11, respectively.

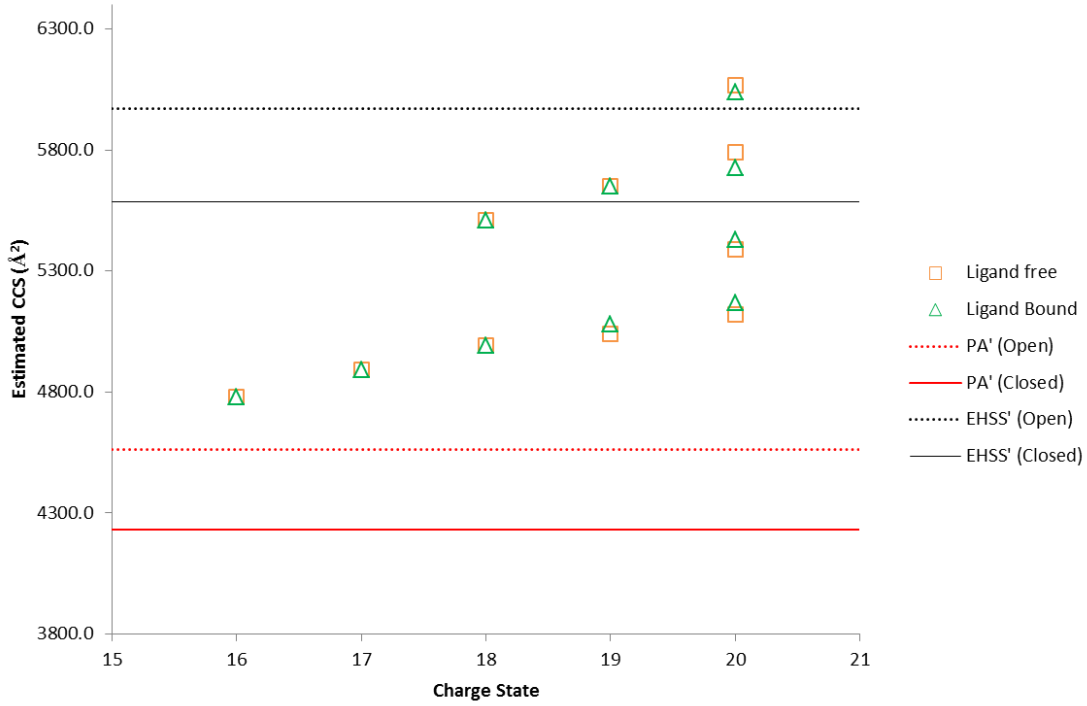


Figure 3-11 – Estimates of the CCS of PREP, and PREP with ligand determined experimentally compared to the open (double solid line) and closed (dotted line) CCS values obtained from crystal structures 1YR2 (open) and 1QFM (closed).
Ligand free and ligand bound PREP shows similar CCS estimates experimentally, illustrating the limited effect ligand binding has on conformation. The experimental difference between the extended and compact conformations evident in the 18^+ and 19^+ charge states suggests these may correspond to the open and closed conformations observed in crystal structures.

Figure 3-11 suggests that the experimental CCS measurements for the native 16^+ charge state of PREP and PREP with ligand, which is proposed to represent a closed form of PREP, is between the PA and EHSS values. The PA method is known to underestimate CCS values for proteins (Scarff, Thalassinos *et al*. 2008), and whilst the EHSS is known to not be the most robust method for CCS calculation it does provide a good compromise between computational demand and speed. The Barran research group have observed that, in a number of proteins, the lowest charge state is consistently smaller than the CCS value obtained from the crystal structures (Jurneczko and Barran 2011), consistent with the observation reported here. The strong agreement between all data points for PREP and PREP with ligand illustrate the limited effect ligand binding has on the CCS of PREP.

The theoretical CCS measurements provide information on the expected magnitude of the CCS change between Open and Closed states of PREP. Based on the PA and EHSS CCS values the difference in open PREP and closed PREP is approximately

6-8%. The 18^+ and 19^+ charge states both have two major conformations which have been proposed to represent Open and Closed forms of the protein. The two conformations of the 18^+ charge state show a difference in CCS of 9%, whilst the two conformations in the 19^+ charge state show a difference of 10%. The magnitude of the change is similar between the experimental and theoretical, and the increased size difference observed in the experimental data may be attributable to charge effects. This data suggests that the two conformations may represent open and closed conformations of PREP.

3.2.3 Dipeptidyl peptidase IV

Mass Spectrometry

Figure 3-12 shows the native MS spectra of DPP IV with and without ligand.

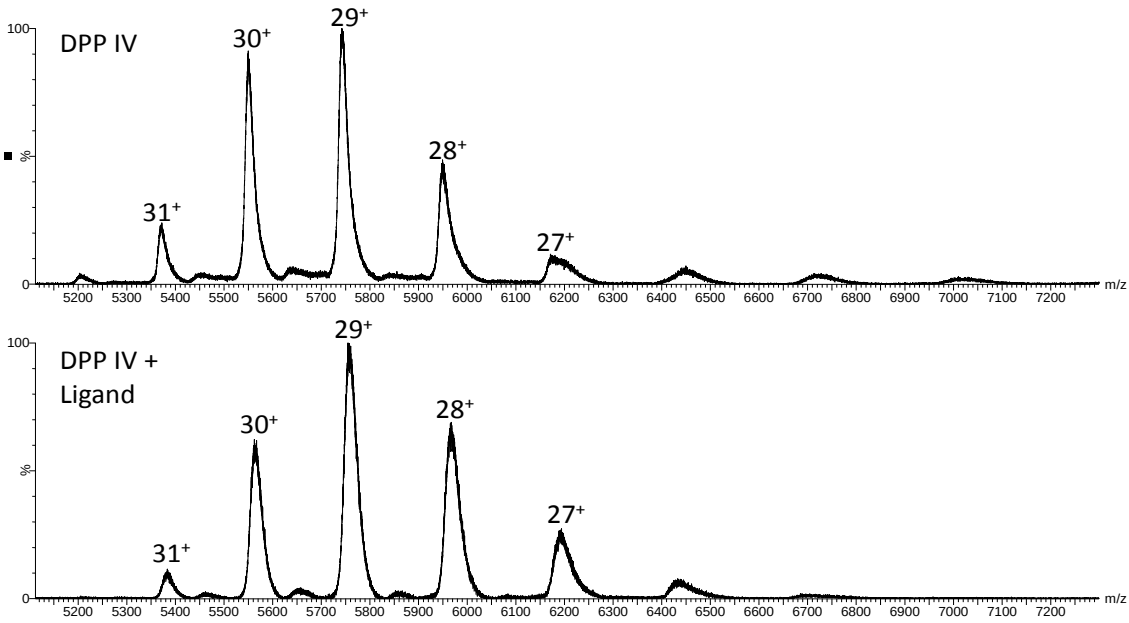


Figure 3-12 – The MS spectra of DPP IV with and without bound ligand.
Ligand binding appears to have a limited effect on the charge distribution, evident only in a reduction in the relative intensity of the 30^+ charge state, suggesting a limited effect on conformation following ligand binding.

The $27^+ \rightarrow 31^+$ charge states of DDP IV are observed at the 5200-6400 *m/z* range of the MS spectrum. Compared to PREP, DPP IV shows a shift to both a higher m/z region of the spectrum, and in the number of observed charges, consistent with the presence of a larger molecule. There was no evidence for any monomeric species in the spectrum, suggesting that non-covalent interactions are preserved upon elevation to the gas phase. The accurate mass obtained from the native-MS spectrum of DPP

IV shown in Figure 3-12 was calculated at 166,530 Da, which agrees with the theoretical mass of 166,528 Da to within 2 Da (or 0.001%). Following incubation with ligand there is an observed mass shift of approximately 660 Da, consistent with the binding of two molecules of the ZPP-like inhibitor at 280 Da each. The peak width of the DPP IV with ligand spectra increases from 21 amu to 30 amu, suggesting a possible overlap of peaks. This may indicate that, whilst the majority of the DPP IV molecules have acquired two molecules of inhibitor, there may be a subpopulation with only one molecule. Any future experiments should ensure that a higher concentration of inhibitor is used.

The accurate mass for DPP IV calculated from the native MS spectra shows a good agreement to the theoretical mass, which is in contrast to the poor agreement between the experimentally observed and theoretical masses of PREP. It has been suggested that the mass discrepancy observed with the PREP sample was a result of solvent trapped within the inner cavity of PREP. The low mass difference observed for DPP IV may suggest that DPP IV is more easily desolvated, than PREP, suggesting a possible difference in the conformation and dynamics between DPP IV and PREP.

Following ligand binding there is a minor shift in the charge distribution. This shift is less than prominent that observed in the PREP analysis as there is no change in observed charge states, but a shift in intensity. This may be due to a minor conformational change in the protein following ligand binding, or may indicate that the bound inhibitor serves to block protonation sites in the active site; given the current theories regarding ligand entry into the active site of DPP IV the latter suggestion is more likely. This is in agreement with the PREP observations of ligand binding, and serves to reinforce the hypothesis of a ligand-dependant blocking of protonation in the active site as a result of binding.

Ion mobility mass spectrometry

The drift time chromatograms for DPP IV are shown in Figure 3-13.

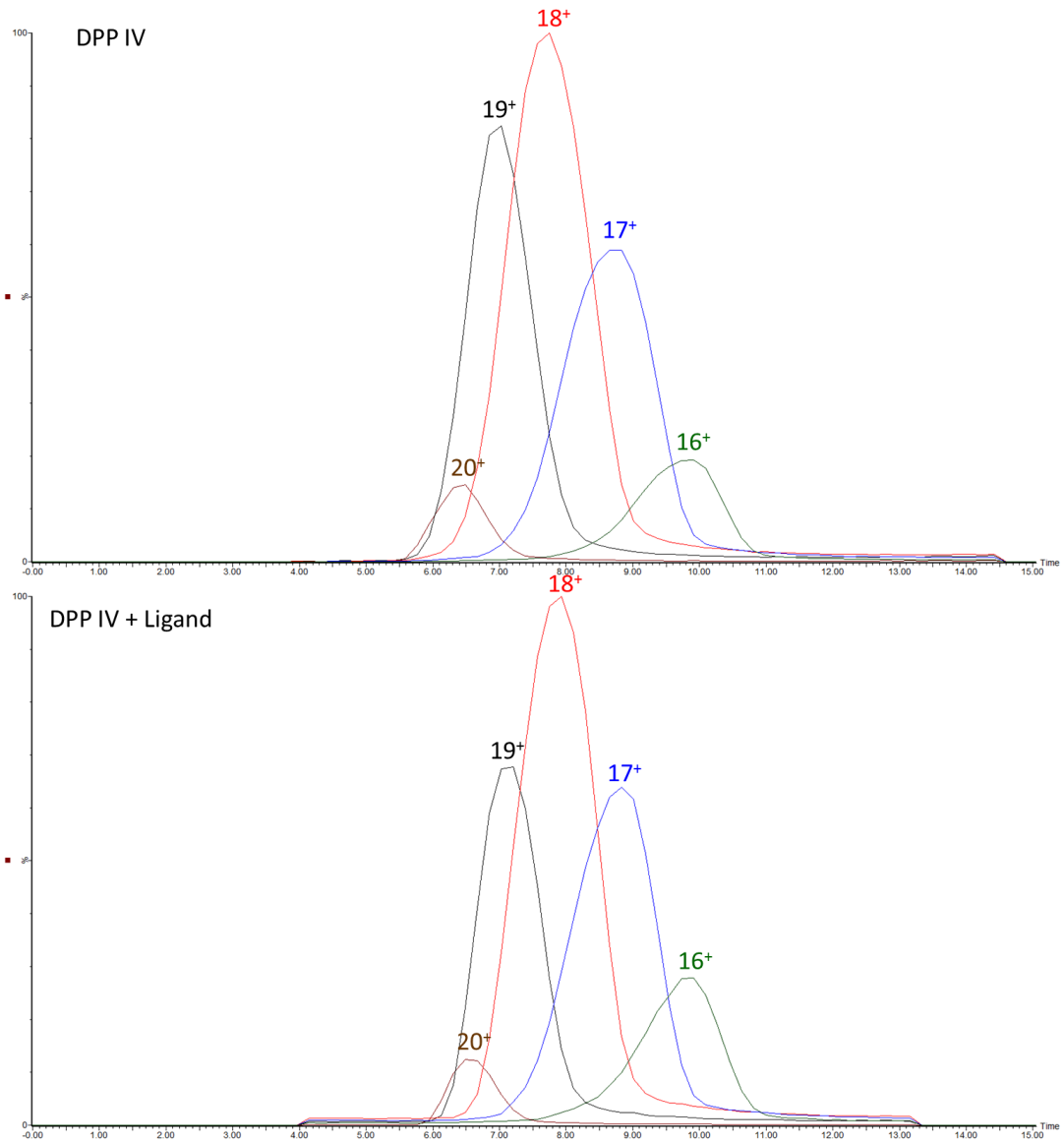


Figure 3-13 – The drift time chromatograms for DPP IV acquired at the same conditions as PREP. All charges states show only a single conformation. Ligand binding appears to have no effect on ATD, indicating ligand binding has no effect on conformation

The drift time chromatograms for DPP IV at all charge states show only a single observable conformation. This suggests that DPP IV is only capable of adopting a single conformation at each charge state, and is resistant to unfolding as a factor of charge. This is in contrast to PREP, which showed clear evidence of undergoing a conformational change to a more open state as a factor of charge at intermediate charge states, followed by larger scale unfolding at higher charge states. This data supports the hypothesis that substrate entry in DPP IV can occur through either the side pore or the β-propeller pore, but not via a domain-opening mechanism.

Following ligand binding there is no observed change in the arrival time distribution of DPP IV.

CCS Estimation

The CCS estimates from DPP IV are shown in Figure 3-14. Computationally derived CCS measurements have been corrected based upon published cross sections taken from Bush *et al* (Bush, Hall *et al*. 2010). A correction factor of 7% was used, based on the observed differences between He and N2 values of alcohol dehydrogenase.

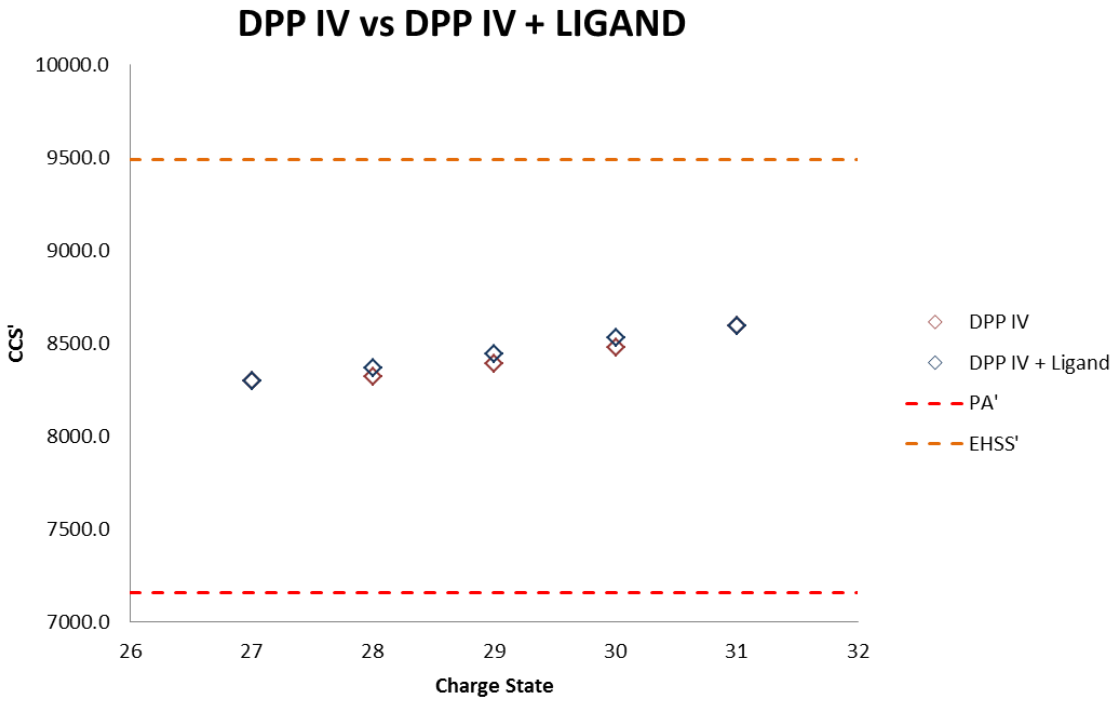


Figure 3-14 – Experimentally derived CCS estimates of DPP IV with and without ligand compared with computationally derived CCS values obtained from PDB entry 1NU8.
Both DPP IV and DPP IV with ligand show a single conformation. CCS differences suggest no change in conformation as a result of ligand binding

As observed with PREP, the DPP IV data falls between the corrected EHSS and PA calculated values. It is known that obtaining a good agreement between CCS estimated from TWIMS data and computer models is highly dependent on which calibration standards are used (Salbo, Bush *et al*. 2012). It is possible that the lower m/z calibration standard (Avidin, molecular weight 64 kDa) is distorting the CCS estimate, leading to an underestimate of experimental CCS relative to the theoretical CCS.

As with PREP, there is no significant change in the CCS of DPP IV with ligand compared to that of the ligand free sample, consistent with previous studies on DPP IV with substrate.

3.3 Conclusions

The aim of this work was to utilise MS and IMS approaches to investigate the solution phase dynamics of two POP family enzymes with a proposed difference in the mechanism of allowing substrate into the active site. Bacterial PREP has been crystalised in an open conformation in solution, suggesting a whole-domain separation as a mechanism for substrate entry. IMS data suggests that procine PREP is capable of adopting a more extended conformation when unfolding is induced as a factor of charge, and that at higher cone voltages this extended conformation is more readily adopted as a factor of charge. A comparison of the CCS values from the x-ray crystal structures of PREP in its open and closed conformation show agreement with the experimentally obtained CCS estimates, suggesting that the dominant compact conformation is the closed state whilst the more extended conformation is related to the open state.

Whilst it is clear that PREP is capable of adopting these conformations when activated by cone voltage/ charge, it is also observed that native charge states, those with the lowest levels of energy, contain only a single observed conformation in both low and high energy experiments. The observed increase in the closed conformation with additional solvent supports the suggestion that open-PREP is a non-physiological conformation induced as a result of gas phase unfolding rather than a second solution phase conformation adopted by PREP. The most likely explanation is that, following collision/ charge dependant activation, the salt bridge connecting the two domains is induced to break, allowing the domains to separate and adopt an open conformation in the gas phase.

DPP IV was suggested to undergo no large conformational change for substrate entry, which is supported by this work. DPP IV shows only a single conformation across all charge states, suggesting the molecule is far more resistant to unfolding, possibly as a virtue of size, when compared to PREP.

Ligand bound versions of both DPP IV and PREP showed little observed structural difference to their ligand-free forms. Some small differences were observed between PREP and ligated-PREP in terms of the intensity of the extended conformations, suggesting that the binding of ligand may have a small effect of the overall stability of the PREP closed conformation, which supports the finding of Min Li (Li 2010). These observed changes were minor, representing a 5-10% shift in intensities, and further work would be required to investigate stability differences induced as a result of ligand binding.

Future work

Whilst we have suggested that the discrepancy observed between the observed and theoretical PREP masses is as a result of solvent, this needs further investigation. It is possible that errors in transcription/ translation have led to a protein containing additional amino acids which could affect its structural properties. In order to identify any of these errors PREP should be analysed under denaturing conditions in order to obtain the true mass of the monomer. It is clear that native PREP conditions require further optimisation in order to improve MS resolution. Given the observations presented here, this may prove difficult since PREP seems to readily adopt an open conformation upon activation, making the balance of activation vs cooling/ focussing a difficult process.

Whilst the computational methods used here for the estimation of PREP/ DPP IV CCS have provided good agreement with experimental data, there are additional methods available. The computationally demanding TM method would provide the most robust estimate of CCS value, whilst the novel PSA method would also provide a better estimate than that of the PSA. There are currently a number of issues with the PSA method regarding molecular size and compatibility which has resulted in some erroneous calculations (Sobott 2014). Future iterations of the algorithm may provide the speed and accuracy required to allow improved comparisons between experimental and theoretical CCS values.

This work has suggested there may be a small difference in the overall stability of PREP following the binding of inhibitor/ ligand. Following optimisation of the instrument conditions, further experiments with more subtle variation between

acceleration voltages, and improved reproducibility, could be performed to confirm the experimental significance of these observations.

3.4 References

Aertgeerts, K., *et al*. (2004). "N-linked glycosylation of dipeptidyl peptidase IV (CD26): effects on enzyme activity, homodimer formation, and adenosine deaminase binding." Protein Sci **13**(1): 145-154.

Ansorge, S., *et al*. (1997). "CD26/dipeptidyl peptidase IV in lymphocyte growth regulation." Adv Exp Med Biol **421**: 127-140.

Barrett, A. J. and N. D. Rawlings (1995). "Families and Clans of Serine Peptidases." Archives of Biochemistry and Biophysics **318**(2): 247-250.

Bastos, I. M. D., *et al*. (2005). "Molecular, functional and structural properties of the prolyl oligopeptidase of Trypanosoma cruzi (POP Tc80), which is required for parasite entry into mammalian cells. ." Biochemical Journal **388**(1): 29-38.

Busek, P., *et al*. (2004). "Dipeptidyl peptidase IV activity and/or structure homologues (DASH) and their substrates in cancer." Int J Biochem Cell Biol **36**(3): 408-421.

Bush, M. F., *et al*. (2010). "Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology." Analytical Chemistry **82**(22): 9557-9565.

Cunningham, D. F. and B. O'Connor (1997). "Proline specific peptidases." Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology **1343**(2): 160-186.

Durinx, C., *et al*. (2000). "Molecular characterization of dipeptidyl peptidase activity in serum: soluble CD26/dipeptidyl peptidase IV is responsible for the release of X-Pro dipeptides." Eur J Biochem **267**(17): 5608-5613.

Elovson, J. (1980). "Biogenesis of plasma membrane glycoproteins. Purification and properties of two rat liver plasma membrane glycoproteins." J Biol Chem **255**(12): 5807-5815.

Engel, M., *et al*. (2003). "The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism." Proceedings of the National Academy of Sciences **100**(9): 5063-5068.

Fülöp, V., *et al*. (1998). "Prolyl Oligopeptidase." Cell **94**(2): 161-170.

Fülöp, V. and D. T. Jones (1999). "β Propellers: structural rigidity and functional diversity." Current Opinion in Structural Biology **9**(6): 715-721.

Fuxreiter, M., *et al*. (2005). "Flexibility of prolyl oligopeptidase: Molecular dynamics and molecular framework analysis of the potential substrate pathways." Proteins: Structure, Function, and Bioinformatics **60**(3): 504-512.

Grandori, R. (2003). "Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry." J Mass Spectrom **38**(1): 11-15.

Holmquist, M. (2000). "Alpha Beta-Hydrolase Fold Enzymes Structures, Functions and Mechanisms." Current Protein and Peptide Science **1**(2): 209-235.

Juhász, T., *et al*. (2005). "Unclosed β-Propellers Display Stable Structures: Implications for Substrate Access to the Active Site of Prolyl Oligopeptidase." Journal of Molecular Biology **346**(3): 907-917.

Jurneczko, E. and P. E. Barran (2011). "How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase." Analyst **136**(1): 20-28.

Kahne, T., *et al*. (1999). "Dipeptidyl peptidase IV: a cell surface peptidase involved in regulating T cell growth (review)." Int J Mol Med **4**(1): 3-15.

Kaszuba, K., *et al*. (2009). "Molecular dynamics study of prolyl oligopeptidase with inhibitor in binding cavity." SAR QSAR Environ Res **20**(7-8): 595-609.

Katta, V. and B. T. Chait (1993). "HYDROGEN-DEUTERIUM EXCHANGE ELECTROSPRAY-IONIZATION MASS-SPECTROMETRY - A METHOD FOR PROBING PROTEIN CONFORMATIONAL-CHANGES IN SOLUTION." Journal of the American Chemical Society **115**(14): 6317-6321.

Kichik, N., *et al*. (2011). "15N Relaxation NMR Studies of Prolyl Oligopeptidase, an 80 kDa Enzyme, Reveal a Pre-existing Equilibrium between Different Conformational States." ChemBioChem **12**(18): 2737-2739.

Konijnenberg, A., *et al*. (2013). "Native ion mobility-mass spectrometry and related methods in structural biology." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(6): 1239-1256.

Li, M., *et al*. (2010). "Induced-fit mechanism for prolyl endopeptidase." J Biol Chem **285**(28): 21487-21495.

Ludwig, A., *et al*. (2002). "Dipeptidyl peptidase IV (CD26) on T cells cleaves the CXC chemokine CXCL11 (I-TAC) and abolishes the stimulating but not the desensitizing potential of the chemokine." J Leukoc Biol **72**(1): 183-191.

Patriksson, A., *et al*. (2007). "Protein Structures under Electrospray Conditions." Biochemistry **46**(4): 933-945.

Polgar, L. (1991). "pH-dependent mechanism in the catalysis of prolyl endopeptidase from pig muscle." European Journal of Biochemistry **197**(2): 441-447.

Polgar, L. (1992). "Prolyl endopeptidase catalysis. A physical rather than a chemical step is rate-limiting." Biochem J **283 (Pt 3)**: 647-648.

Rea, D. and V. Fulop (2006). "Structure-function properties of prolyl oligopeptidase family enzymes." Cell Biochem Biophys **44**(3): 349-365.

Rea, D. and V. Fulop (2011). "Prolyl Oligopeptidase Structure and Dynamics." CNS & Neurological Disorders - Drug Targets (Formerly Current Drug Targets **10**(3): 306-310.

Salbo, R., *et al*. (2012). "Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers." Rapid Communications in Mass Spectrometry **26**(10): 1181-1193.

Samalikova, M. and R. Grandori (2003). "Protein charge-state distributions in electrospray-ionization mass spectrometry do not appear to be limited by the surface tension of the solvent." J Am Chem Soc **125**(44): 13352-13353.

Scarff, C. A., *et al*. (2008). "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 Communications in Mass Spectrometry **22**(20): 3297-3304.

Shan, L., *et al*. (2005). "Structural and mechanistic analysis of two prolyl endopeptidases: Role of interdomain dynamics in catalysis and specificity." Proceedings of the National Academy of Sciences of the United States of America **102**(10): 3599-3604.

Sobott, F. (2014). Personal Correspondance. E. M. J.

Sobott, F. and C. V. Robinson (2004). "Characterising electrosprayed biomolecules using tandem-MS--the noncovalent GroEL chaperonin assembly." International Journal of Mass Spectrometry **236**(1-3): 25-32.

Steinberg, M. Z., *et al*. (2007). "The dynamics of water evaporation from partially solvated cytochrome c in the gas phase." Physical Chemistry Chemical Physics **9**(33): 4690-4697.

Stryer, L. (2011). Biochemistry.

Suzuki, Y., *et al*. (1993). "Dietary regulation of rat intestinal angiotensin-converting enzyme and dipeptidyl peptidase IV." Am J Physiol **264**(6 Pt 1): G1153-1159.

Szeltner, Z., *et al*. (2004). "Concerted Structural Changes in the Peptidase and the Propeller Domains of Prolyl Oligopeptidase are Required for Substrate Binding." Journal of Molecular Biology **340**(3): 627-637.

Thalassinos, K., *et al*. (2004). "Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer." International Journal of Mass Spectrometry **236**(1–3): 55-63.

Tiruppathi, C., *et al*. (1990). "Evidence for tripeptide-proton symport in renal brush border membrane vesicles. Studies in a novel rat strain with a genetic absence of dipeptidyl peptidase IV." J Biol Chem **265**(4): 2048-2053.

Chapter 4 : Investigating the structural differences in haemoglobin variants

4.1 Introduction

Hemoglobin is the protein in erythrocytes responsible for the transport of both oxygen molecules from the lungs to tissues, and carbon dioxide from tissues to the lungs (Thein 2011). It is the most abundant protein in mammals and, due to its high abundance and vital role in gas transport, is one of the most well characterised proteins on the planet. The three dimensional structure of haemoglobin has been

solved using X-ray crystallography, and the structural changes and kinetics associated with oxygen binding have been extensively investigated and characterised (Buxbaun 2007).

The haemoglobin molecule, shown in Figure 4-1 is a heterotetramer consisting of a two pairs of chains.

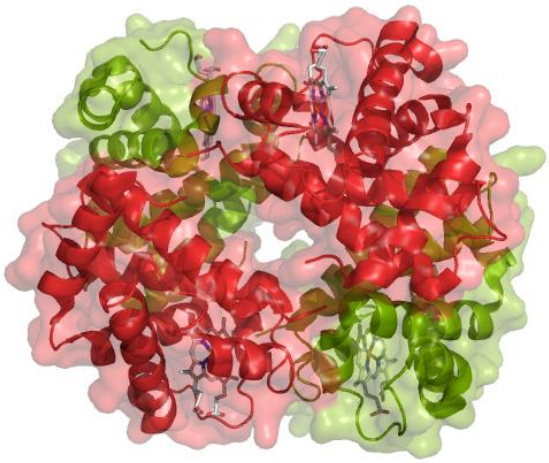


Figure 4-1 – The crystal structure of normal adult haemoglobin containing two α-chains (red) and two β-chains (green), each bound to a single haem group.

All haemoglobin molecules consist of two identical copies of both an α-like chain and a β-like chain each containing a prosthetic haem group located in a small pocket in the protein (Table 4-1). It is the haem group that affords oxygen binding function to the molecule (Buxbaun 2007). During human embryo development different haemoglobin tetramers are present in the blood. These are determined by the different expression of α-like and β-like genes at each of these stages. There are two α-like globin chains, α and ζ, and four β-like globin chains, β, δ, γ and ε (Thein 2011). These different globin chain variants have different oxygen binding affinities, with those chains expressed during the embryonic stages of development, α-like ζ – globin and β-like γ- and ε-globin, having the highest affinity for oxygen. In adult humans 96-98% of all haemoglobin is haemoglobin A (HbA), consisting of two α- and two β-globin chains. A minor haemoglobin, haemoglobin A_2 (HbA_2), consisting of two α- and two δ-globin chains, is also present at minor concentrations (approximately 2% in adult erythrocytes) (Bain 2006).

Alpha Chain	VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGS AQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSH CLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
Beta Chain	VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPD AVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENF RLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

Table 4-1 – The sequences of haemoglobin α-chain and β-chain, the two major chains present in adult haemoglobin

Disorders affecting haemoglobin, known collectively as haemoglobinopathies, are some of the most common of all inherited disorders; it is estimated that ~7% of the global population are carriers for an inherited haemoglobin disorder. Haemoglobinopathies are subdivided into two main groups: thalassemias, disorders relating to the quantitative levels of one or more of the protein chains, and haemoglobin structural variants, disorders resulting from mutations to the genes that manifest as amino acid substitutions in the protein primary sequence. Whilst there are a wide range of different haemoglobin variants the vast majority have no significant effect on haemoglobin properties and cause no clinical problems (Bain 2006).

Over 1000 structurally different haemoglobin variants have been described. The most debilitating of these is the HbS variant which leads to sickle-cell anaemia. First described by Pauling in 1949 (Pauling, Itano *et al*. 1949), and later characterised by Ingram in 1956 (Ingram 1956), HbS is a β-chain variant characterised by a single amino acid substitution at the 6^{th} position of the β-chain (Glu → Val). This mutation induces a conformational rearrangement in the tetramer that facilitates molecular stacking. Deoxygenated HbS polymerises via this stacking mechanism, leading to the erythrocytes adopting cresentic disc shape as opposed to the normal biconcave disc (Murayama 1967). Fortunately, sickle-cell anaemia is only present in people homozygous for the HbS gene and is benign in heterozygous individuals.

HbC is another clinically significant structural mutant of haemoglobin. First described by Hunt and Ingram in 1958 (Hunt and Ingram 1958), HbC is β-chain variant characterised by a single amino acid substitution at the 6^{th} position of the β-

chain (Glu → Lys). As with HbS, HbC allows the formation of a condensed phase of haemoglobin (Vekilov, Feeling-Taylor *et al*.); oxygenated HbC is capable of forming intraerythrocytic crystals which contribute to the pathology of the disease (Lessin, Jensen *et al*. 1969). As with HbS, HbC is benign in individuals heterozygous for the disease but leads to disease in patients homozygous for the mutation (Lawrence, Fabry *et al*. 1991). HbC can also lead to disease in patients that are heterozygous for both HbS and HbC; in these cases patients exhibit the oxy-HbC crystals found in normal homozygous HbC patients (Nagal and Lawrence, 1991).

The use of mass spectrometry for the analysis of haemoglobin has been well documented. Clinical diagnosis of haemoglobinopathies uses a number of different techniques (Clarke 2000, Shimizu, Nakanishi *et al*. 2006) with ESI-MS and MS/MS approaches rapidly becoming one of the preferred choices due to low sample consumption and the potential for high-throughput, automated data when coupled to HPLC (Witkowska, Lubin *et al*. 1991, Clarke 2000, Wild, Green *et al*. 2001, Daniel, Turner *et al*. 2005, Shimizu, Nakanishi *et al*. 2006). Mass spectrometry has also been used to study the mechanism of assembly and disassembly of the haemoglobin tetramer (Griffith and Kaltashov 2003, Griffith and Kaltashov 2006, Boys and Konermann 2007, Scarff, Patel *et al*. 2009)

The work by Scarff *et al* made use of ion mobility mass spectrometry to study the gas phase structures of the tetramer, and its constituents, and compare them with solution phase structures elucidated from x-ray diffraction crystal structures. Since the publication of that work in 2009 a novel set of calibration standards have become available for native protein calibration standards (Bush, Hall *et al*. 2010). Developments in the field of native mass spectrometry have also refined the protocols for both protein preparation and MS analysis.

The focus of this work is to revisit the Scarff *et al* research using these new developments in native mass spectrometry, and also to make use of the previously unavailable PSA method for calculating protein cross sections, in order to study the gas phase structures of normal haemoglobin A and the mutant haemoglobin S and haemoglobin C variants.

4.2 Results and Discussion

4.2.1 Identification of Haemoglobin variants

Denatured MS spectra

Prior to any native analysis the integrity of HbA, HbS and HbC samples obtained from whole blood samples were confirmed by MS. Spectra representative of HbA, HbS and HbC analysed using ESI-TOF-MS under denaturing conditions (50% MeOH, 0.1% HCOOH) are shown in Figure 4-2. HbA is comprised of two chains; the α- chain, which has a theoretical mass of 15126.2 Da in the apo form, and the β-chain, which has a theoretical mass of 15867.2 Da in the apo form (based on amino acid sequence). The HbS mutation of glutamate → valine introduces a mass change of -30 Da (glutamate mass 147.13 Da, valine mass 117.15 Da) to the beta chain, whilst the HbC mutation of glutamate → lysine (lysine mass 146.19 Da) introduces a mass change of -1 Da. This mass shift can be observed in the spectra shown in Figure 4-2, where the β-chain peaks show a minor shift in the HbS spectra (for the 15^+ charge state expected shift is 2 m/z, observed shift is 2 m/z). This shift is present, but less prominent, in the HbC sample as a result of the smaller mass difference (15^+ charge state expected shift 0.0667 m/z, observed shift 0.141).

Important to note is that all three samples contain two major charge state distributions, supporting the view that each sample contains a homozygous Hb protein for HbS, HbC or HbA.

Both the observed mass differences and the presence of only two major chains are highlighted in the deconvoluted mass spectra Figure 4-3. Deconvolution of the mass spectra to true mass scale indicates that HbA, HbS and HbC all contain only two major species All masses show agreement to within 2 Da of theoretical masses. For HbS, this is a sufficient mass accuracy for identification of the α- and β- chains. For HbC this is not, as the error is larger than the expected change in mass as a result of mutation. The α- chain in all patient samples should have the same mass and can therefore be used as an internal standard. α-chain masses in all cases were corrected to 15126.2 Da, and the same correction factor applied to the β-chain. Using the α-chain in this manner it was possible to identify the and 1 Da mass shift between HbA and HbC, supporting the view that the sample is homozygous HbC

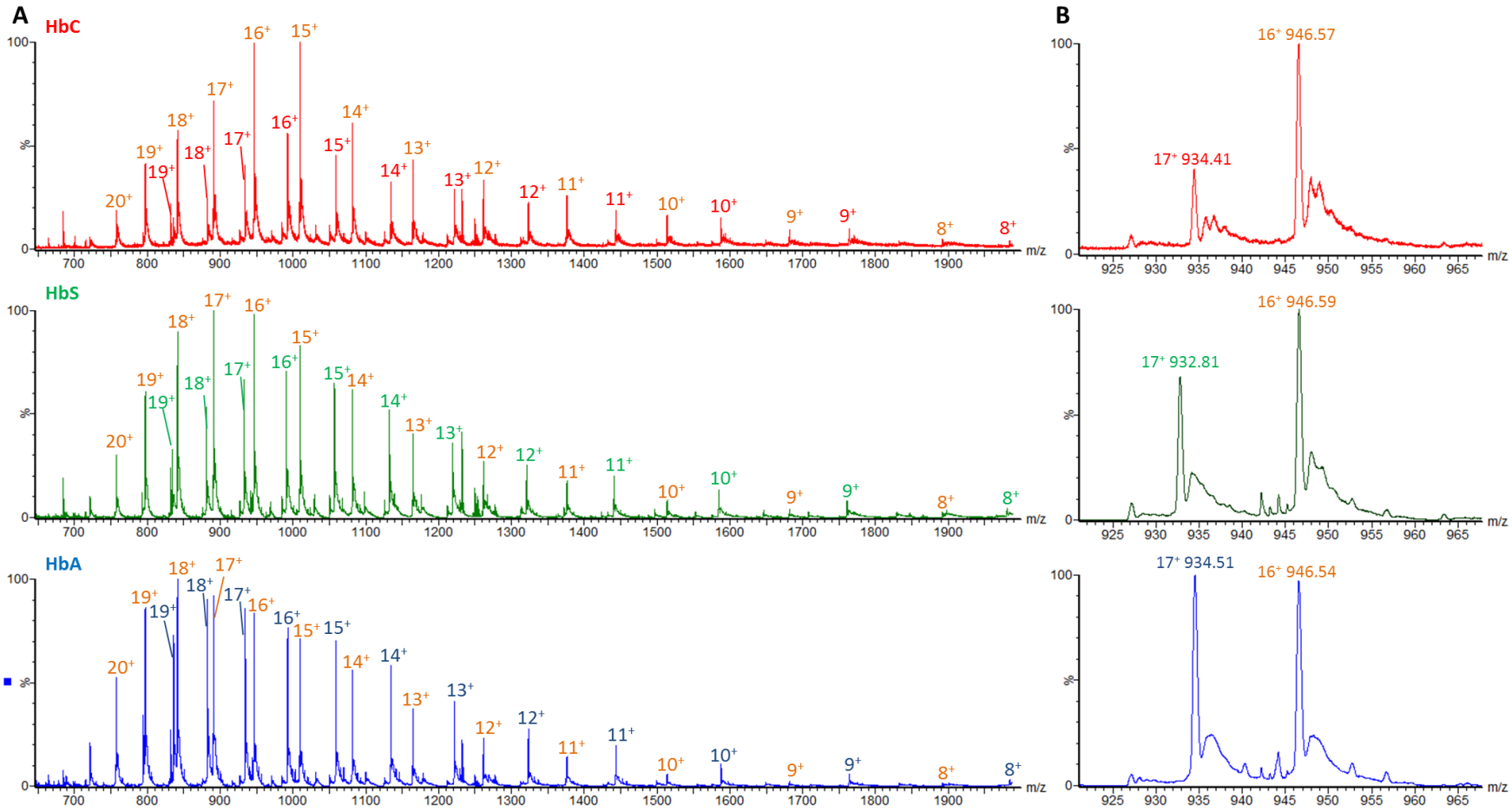


Figure 4-2 – A - The ESI-MS spectra for HbA (blue), HbS (green) and HbC (red) in denaturing conditions.
All three species contain two charge state distributions; the α-chain, labelled in orange, which is consistent between the three samples and the β-chain, which is labelled in blue for HbA, green for HbS and red for HbC. B – a magnification of the 16^+ charge state of the α- chain and the 17^+ charge state of the β-chain illustrates the m/z differences within the β-chain molecular weight for HbA and HbS and HbC chains.

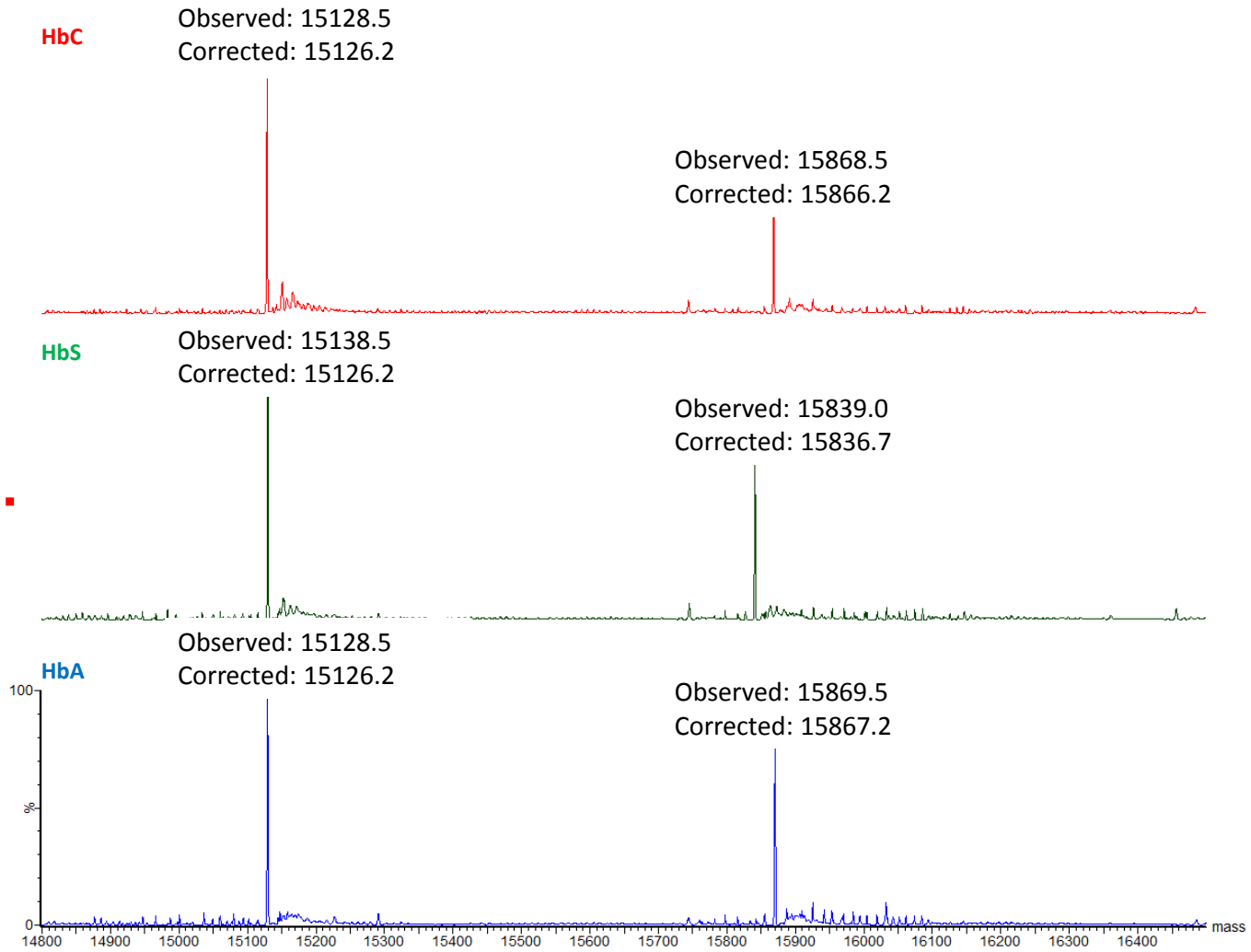


Figure 4-3 – The deconvoluted MS spectra for HbA (blue), HbS (green) and HbC (red), obtained from the data shown in Figure 4-2. All samples show a conserved peak, relating to the α-chain and a variable peak, relating to the β-chain. Peaks are labelled with the experimentally observed masses, and the corrected masses based on the mass correction of the α-chain.

Native MS spectra

Spectra representative of HbA, HbS and HbC analysed by means of ESI-TOF-MS under native-like conditions are shown in Figure 4-4. The data for the tetramer region were deconvoluted to give masses of 64,455.5 Da for HbA, 64,395.4 Da for HbS and 64,452.3 Da for HbC. These agree well with the theoretical masses of 64,453.2 Da and 64,393.4 Da for HbA and HbS (Ofori-Acquah, Green *et al*. 2001) and 64,451.2 for HbC. Figure 4-4 shows peak broadening, consistent with solvent adducts and incomplete desolvation. The effect of residual solvation of a protein has been explored and has been shown to stabilise the native complex (Steinberg, Breuker *et al*. 2007, Patriksson, Marklund *et al*. 2007). Accurate mass is less important than the assessment of structure and conformation in these experiments, since the identities of the samples have already been confirmed by denatured MS, and therefore optimisation of conditions for improved resolution was not carried out.

The spectra in Figure 4-4 show the presence of the Hb tetramer, $\alpha_2\beta_2$ complete with four prosthetic haem groups, the Hb dimer, αβ with two haem groups, and the apo- and holo- forms of the momoneric α- and β- chain. Consistent with previous studies, there is no evidence of a trimeric species or higher oligomeric species, giving confidence that the protein is being studied in its solution phase conformation.

The observed charge state distribution for HbA, HbS and HbC tetramers in this study span the $14^+ \rightarrow 18^+$ charge states, centred at the 16^+ charge state. This is a significant change to that shown in Scarff *et al*, in which major charge state distribution spanned the $15^+ \rightarrow 18^+$ species, and Ofori-Acquah *et al*, in which major charge state distribution spans the $17^+ \rightarrow 19^+$ species (Ofori-Acquah, Green *et al*. 2001, Scarff, Patel *et al*. 2009). A shift towards a lower, narrower charge state distribution is indicative of a more folded structure, suggesting a more native-like conformation. This observation may suggest the hypothesis that a desalting approach using only Amicon centrifugal filter units for buffer exchange, in preference to using ion-exchange resin beads, is a gentler method for desalting proteins.

In 2000 J. Fernandez de la Mora identified a relationship between the number of charges a protein can acquire and its shape in solution. Proteins with a globular, folded, shape are unable to acquire more charges than the Rayleigh limit of a solvent

droplet of equal size, determined by Equation 4-1 (Fernandez de la Mora 2000), where Z_R is the maximum number of charges, and *m* is the mass of the protein.

$$Z_R = 0.0778\sqrt{m}$$

Equation 4-1

For haemoglobin $Z_R = 20$, which would suggest that all previous investigations have retained a globular structure for haemoglobin. It is therefore likely that the changes in sample preparation simply lead to a better observation of the lower charge states of haemoglobin, rather than changing proteins structure.

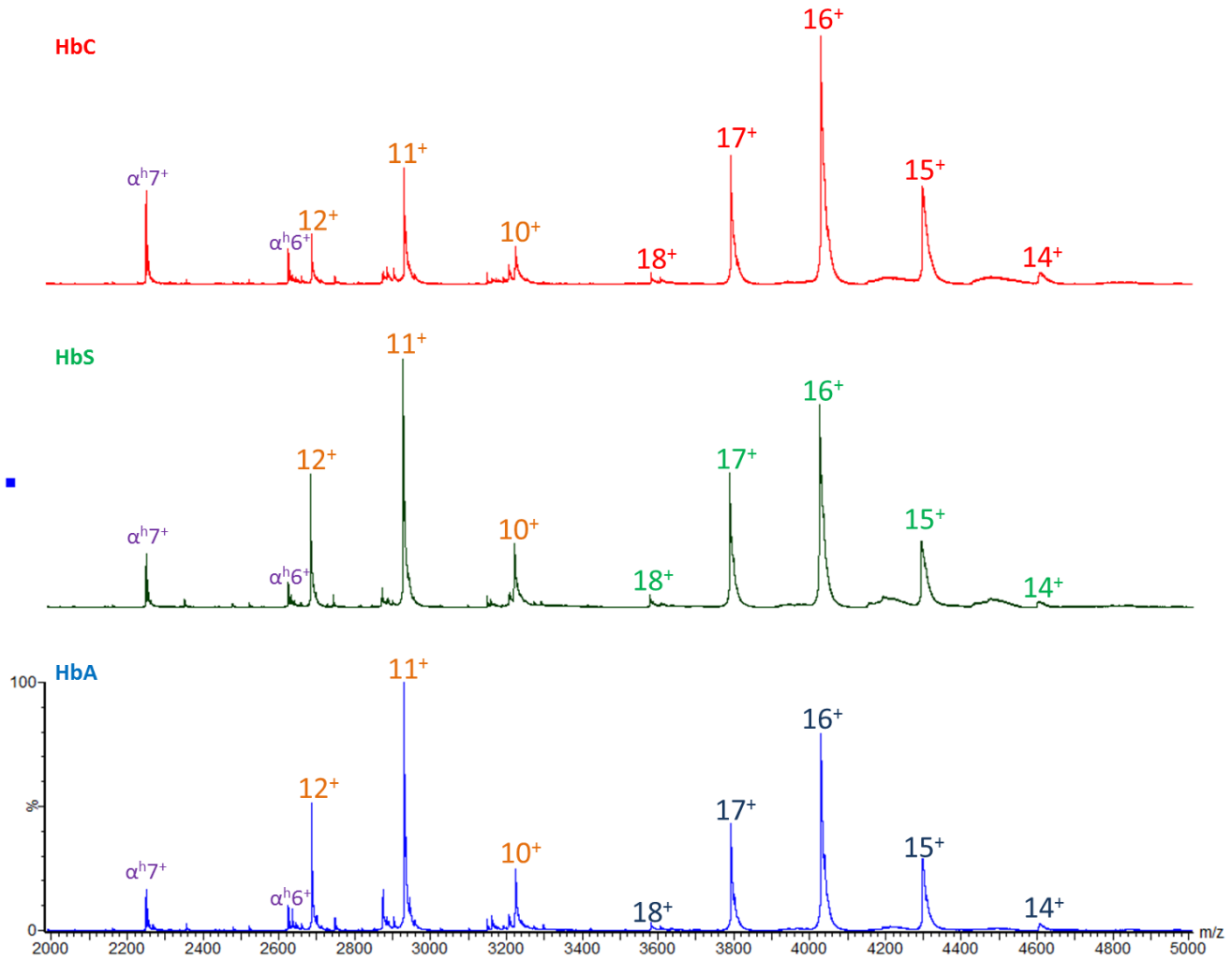


Figure 4-4 - The ESI-MS spectra for HbA (blue), HbS (green) and HbC (red) in native conditions.
All three samples show evidence of monomer peaks, labelled in purple, dimer peaks, labelled in orange, and tetramer peaks labelled in blue (HbA), green (HbS) and Red (HbC).

All three samples also show evidence of a low intensity species distinct from the Hb tetramer charge state present at 3900-4600 *m/z*. Deconvolution of these low intensity peaks suggests they are the 15^{+}-17^{+} charge states of an approximately 66.5 kDa species. It is unlikely that this species is formed as a result of fragmentation/ adduct formation of the Hb tetramer given the observed molecular weight. All samples used in this analysis were obtained from whole blood, and therefore it is possible that this charge series corresponds to human serum albumin (HSA). HSA is the most abundant protein in human plasma, and has a molecular weight of 66.4 kDa, although this may be higher dependant on post-translational modifications (Putnam 2012). The observed charge state distribution of 15^{+}-17^{+} is also consistent with the observed charge state distribution of bovine serum albumin (Bush, Hall *et al*. 2010), the bovine homologue of HSA. Deconvolution of the denatured Hb data (Figure 4-2) over a wider mass range showed no evidence for the presence of HSA, however this may be due to high intensity Hb α- and β- chain signals masking the ions.

4.2.2 Investigating the effect of source temperature on protein structure

With native-like MS experiments careful optimisation of the instrument parameters is important for retention of protein structure. The intent of this study was to revisit data previously published by Scarff *et al* (Scarff, Patel *et al*. 2009) in light of recent improvements in to native MS experiments. Previous publications on native haemoglobin used elevated source temperatures which could lead to partial denaturation. The effect of temperature of the source on the structure of HbA and HbS were investigated using a combination of native MS and IMS-MS. The results are shown in Figure 4-5.

Native MS

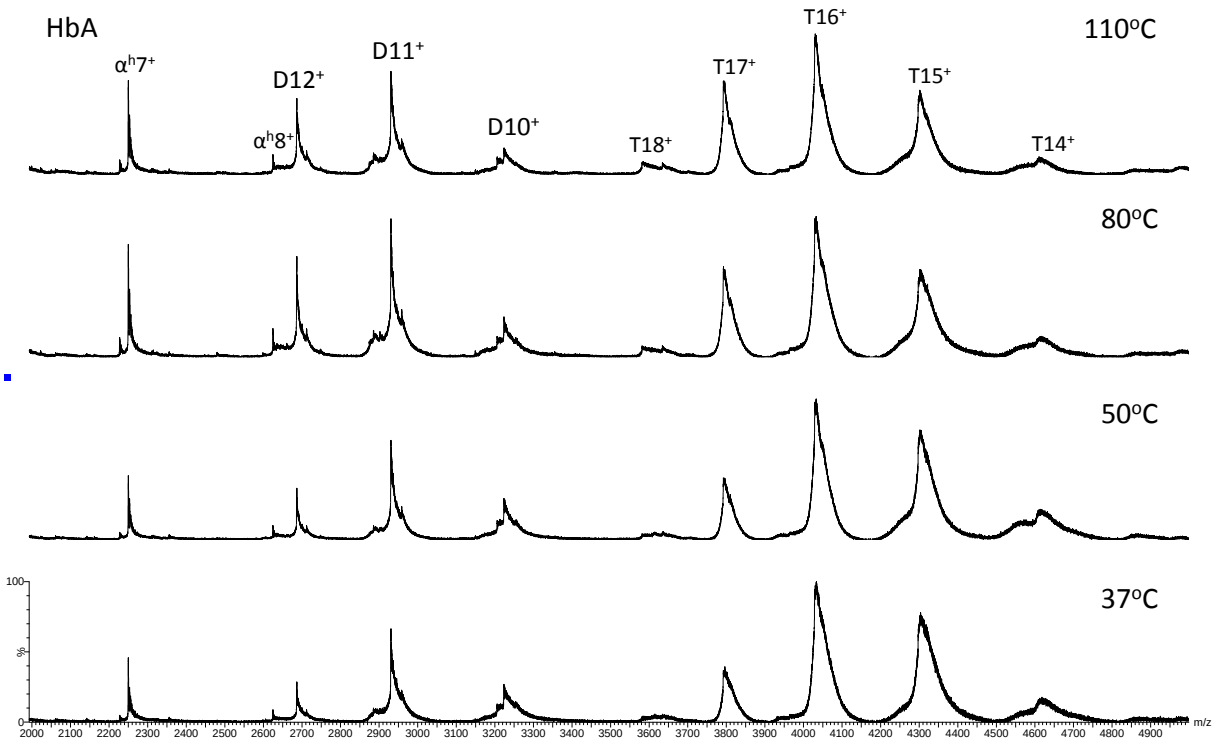


Figure 4-5 – Native MS spectra of HbA illustrating the effect of varying source temperature on the MS spectrum. Increasing the source temperature appears to increase the relative intensity of the monomer and dimer compared to the tetramer, suggesting dissociation of the tetramer oligomers

Figure 4-5 shows an increased peak width at half maximum (PWHM) for all samples compared with spectra shown in Figure 4-3. This peak width increase arises as a result of increased pressures throughout the system induced by operating with IMS separation on and a source pressure of 6.8 mBar (consistent with Scarff *et al* (Scarff, Patel *et al*. 2009)). It can be observed, however, that as the temperature increases the PWHM narrows. This improved resolution comes from improved desolvation induced at elevated temperatures. Figure 4-5 also suggests that at increased temperatures the relative intensity of the α^{holo} monomer 7^+ charge state increases, most prominently at 80^{o}C and 100^{o}C. This observation could be explained as by dissociation of the dimer. At 80^{o}C there is, however, also an increase in the relative intensity of the dimer suggesting that at elevated source temperatures there is either better transmission of the dimer ions, or disruption of the tetramer to generate more monomer/ dimer ions.

There also appears to be a minor increase in relative intensity of the T^{18+} charge state with an increase in temperature, coinciding with a minor decrease in the intensity of the T^{14+} charge state. This is also observable in the dimer; as temperature increases so does the relative intensity of D^{12+} compared to D^{10+}. As a protein unfolds more basic sites become exposed, leading to an increase in the number of sites available for protonation on the molecule. A previous study by Justin Benesch (Benesch, Sobott *et al*. 2003) explored the thermal unfolding of a protein by using a heated capillary to unfold a protein in solution prior to electrospray analysis. It was observed that with increasing capillary temperature *Ta*HSP16.9, a dodecamer, thermally dissociated into monomeric and dimeric subunits. It was also observed that for monomeric lysozyme, increasing the capillary temperature led to a wider charge state distribution. The data presented here shows a similar trend, supporting the theory that increasing the source temperature adversely affects the folded state of the protein. It should be noted, however, that heating a sample using a thermally-controlled capillary results in a higher degree of dissociation of the complex compared to varying the temperature of the source region, suggesting that source temperature has a limited effect on structure.

Figure 4-6 shows the effect of source temperature on an HbS sample.

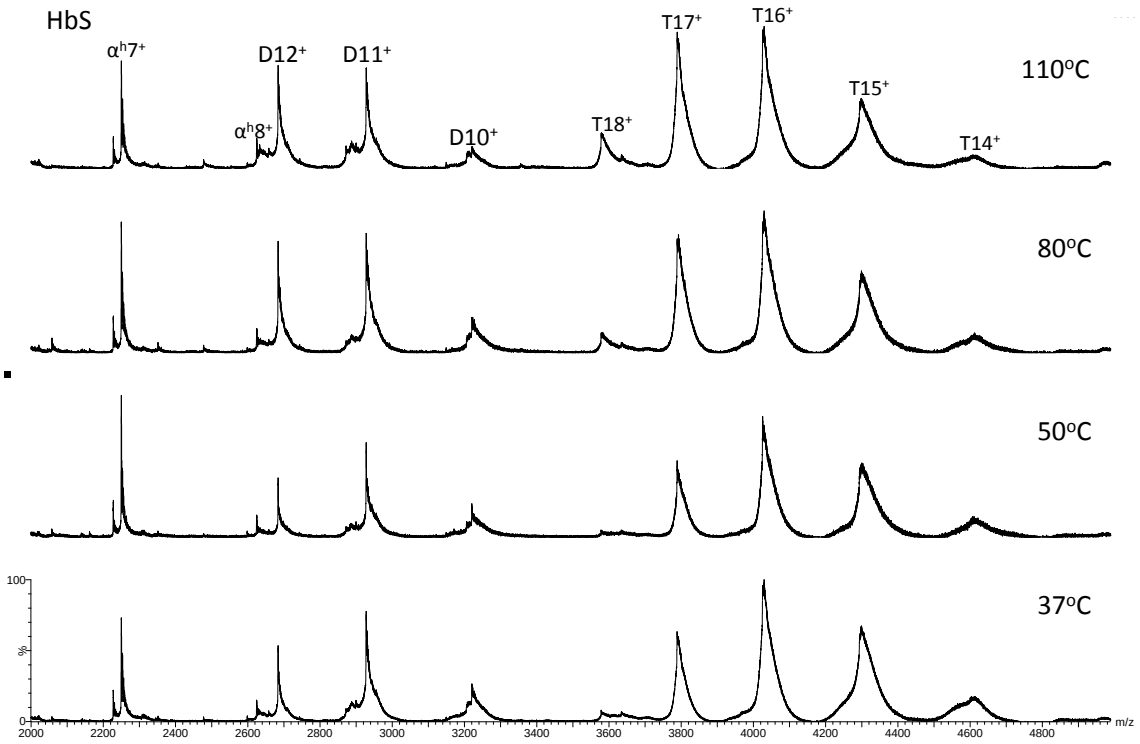


Figure 4-6 - Native MS spectra of HbS illustrating the effect of varying source temperature on the MS spectrum. Increasing the source temperature appears to increase the relative intensity of the monomer and dimer compared to the tetramer, suggesting dissociation of the tetramer oligomers

Figure 4-6 shows a similar trend to that shown in Figure 4-5. With increasing temperature the intensity of monomer and dimer peaks increases, suggesting a temperature-dependant dissociation of the tetramer. There is also an observed shift in charge state distribution, for both the tetramer and the dimer, towards more extended conformations, consistent with unfolding of both species. The increase in intensity of the T^{18+} ion is more pronounced in the HbS sample than for HbA, suggesting a possible difference in thermal stability.

It is likely that the observed low mass peak shouldering observed in both Figure 4-6 and Figure 4-7 is as a result of the peak broadening of the proposed HSA peaks via solvent adduct formation.

Ion mobility mass spectrometry

Figure 4-7 shows the arrival time distributions of HbA charge states 14^+ to 17^+. The ATD for the 18^+ charge state is not included since it had a poor signal to noise ratio under these experimental conditions.

The 14^+ charge state, which is proposed to be representative of the native structureshows that with increasing temperature there is no major change in the drift time, suggesting that the protein undergoes no large conformational changes as a result of temperature. There is evidence, however, of an increase in the ATD for the 14^+ charge state with increasing temperature, with broadening occurring beyond 50^oC. Increased peak broadening is consistent with an increase in the number of conformations a protein is able to adopt, suggesting an increased flexibility with the molecule, consistent with a higher degree conformational freedom.

For the 37^oC data, increasing charge leads to a decrease in arrival time, as a more highly charged molecule traverses the mobility cell faster, as well as increased tailing of the peak until a shoulder appears at 17^+. Peak shouldering is consistent with the emergence of an additional conformation as a result of partial unfolding of the protein.

Looking at both sets of the data together, increasing temperature leads to a minor increase in drift time, followed by a decrease in drift time across all charge states. These observations suggest an expansion followed by a relaxation of the protein in the gas phase. Interestingly, the decrease in drift time coincides with an increase in the proteins ATD suggesting that the relaxation that occurs may facilitate the protein adopting a wider range of conformations. Increasing both source temperature and charge leads to peak broadening and shouldering to an increased amount when considered together. The effects of both temperature and charge have an additive effect of destabilising the protein to the point that, at the 17^+ charge state at 50^oC, 80^oC and 110^oC HbA has two distinct peaks in its ATD consistent with two distinct conformations.

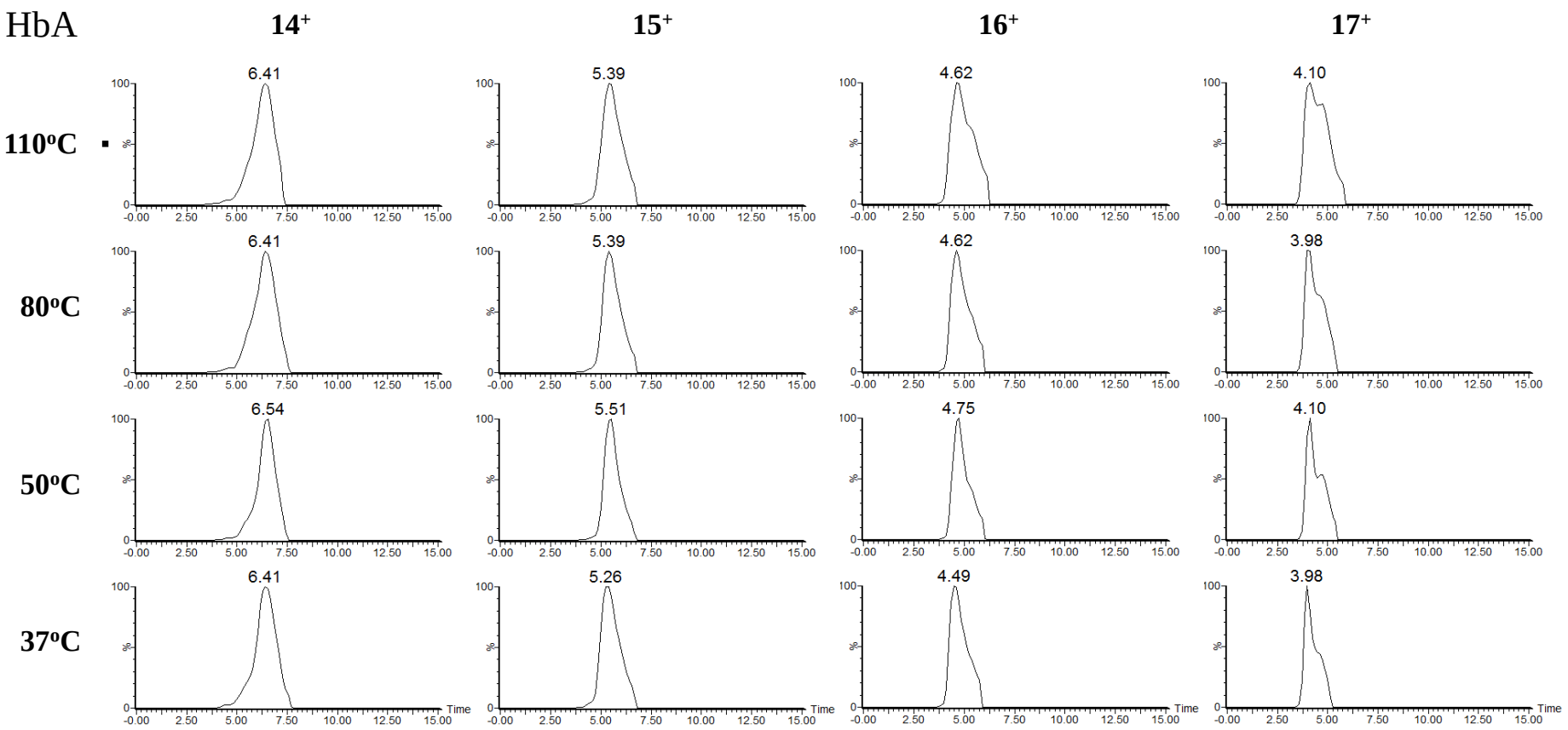


Figure 4-7 – The arrival time distributions for the $14^+ \rightarrow 17^+$ charge state of HbA at four different temperatures: 37ºC, 50ºC, 80ºC and 110ºC.
Increasing the temperature appears to have a limited effect on the arrival time distributions of the 14^+ and 15^+ charge states. 16^+ and 17^+ charge states show evidence of increased peak shouldering, consistent with the appearance of an additional conformation, suggesting that elevated temperatures may have a deleterious effect on the conformation of the higher charge states.

Figure 4-8 shows the observed ATD's for HbS. There is evidence for increased peak broadening of each charge state with increased temperature consistent with an increase in the dynamics of the protein at higher temperatures. There is also evidence for a more extended conformation of HbS observable at higher charge states, suggested by the shouldering present in 16^+ and 17^+ charge states at all temperatures.

Whilst the trend shown by the HbS data is similar to that observed for HbA, there is evidence to suggest that HbS is affected to a greater degree by charge and temperature effects. Both HbS and HbA show changes in the drift time profile of the 14^+ charge state; HbA shows a minor increase, followed by a decrease. HbS, however, shows a significant decrease in drift time, consistent with a gas phase collapse of the protein. Both HbA and HbS also show increased peak broadening as a factor of temperature; HbS shows a much more severe peak broadening at FWHM compared to HBA suggesting that it is more susceptible to temperature-dependant effects. Both proteins show evidence of peak shouldering as a factor of charge, again with HbS showing a greater extended profile than HbA.

HbS

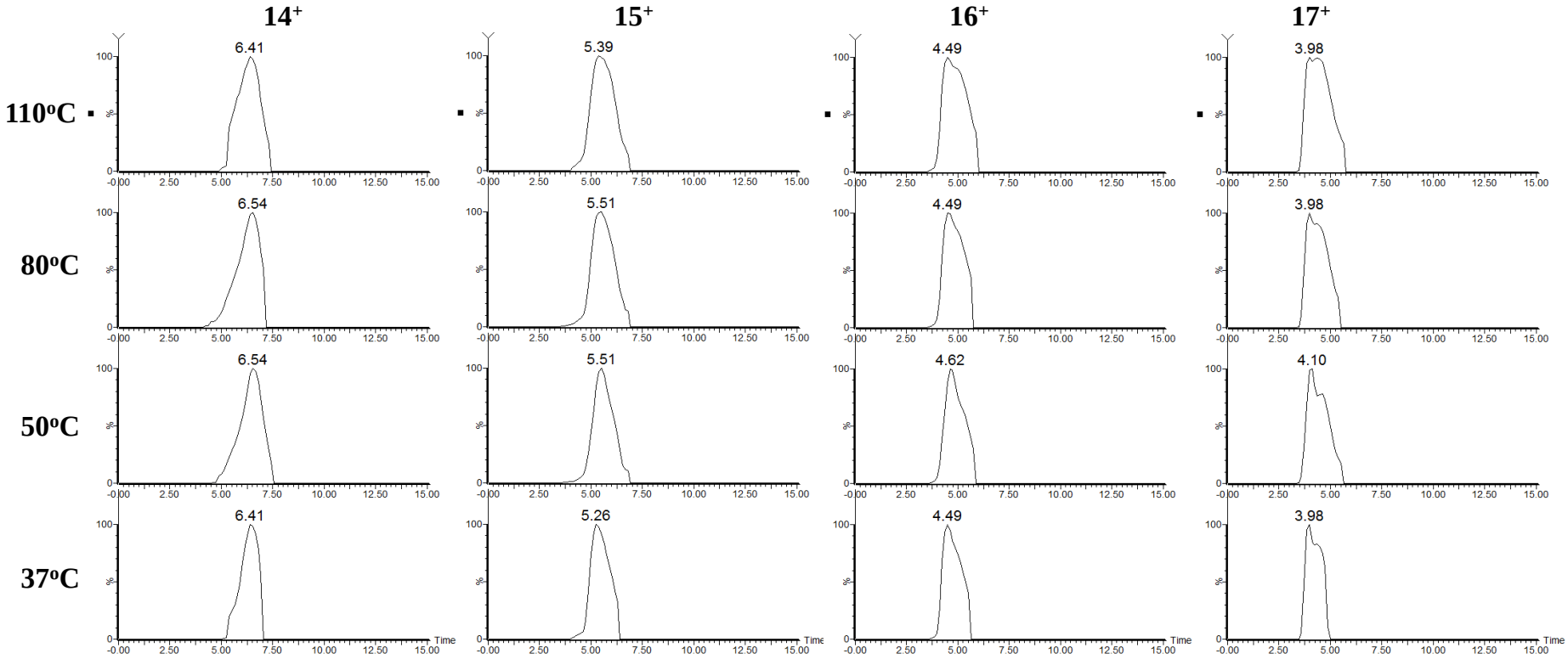


Figure 4-8 - arrival time distributions for the $14^+ \rightarrow 17^+$ charge state of HbS at four different temperatures: 37°C, 50°C, 80°C and 110°C.
Increasing the temperature appears to have a limited effect on the arrival time distributions of the 14^+ and 15^+ charge states. 16^+ and 17^+ charge states show evidence of increased pear shouldering, consistent with the appearance of an additional conformation, suggesting that elevated temperatures may have a deleterious effect on the conformation of the higher charge states.

4.2.3 Investigating the effect of backing gas pressure on protein structure

The effects on increasing the pressure in the source are discussed in detail in Section **1.3 Biological mass spectrometry**. Briefly, changing the backing pressure serves to a provide collision focusing effect on the ions which helps to improve transmission of protein complex ions through the mass spectrometer. Optimum conditions are protein dependant, and must be carefully controlled as high pressures can lead to collisional activation of a protein complex.

The backing gas pressure was reduced from 6.8 mBar to 4.7 mBar to investigate the effect of pressure on the observed MS spectra. Based on the observations outlined above, and literature suggestions (Kirshenbaum, Michaelevski *et al*. 2010, Salbo, Bush *et al*. 2012), source temperature was reduced to 37^{o}C to minimise the effects of temperature on protein conformation, dynamics and complex activation. The effects of this pressure change on MS peak shape and drift time/ CCS values are compared below.

Native MS

Figure 4-9 shows that at the reduced pressure of 4.7 mBar, there is a significant improvement in the observed peak shapes of HbA across the whole mass range; FWHM changed from 43.85 amu at 6.8 mBar to 5.42 amu at 4.7 mBar for the $T16^{+}$ ion, 12.39 amu to 1.23 amu for the $D11^{+}$ ion and 1.5 amu to 1.2 amu for the $\alpha^{h}7^{+}$ ion. This is consistent with previous observations of increased peak width at higher pressure as a result of gentler desolvation conditions throughout the instrument (Sobott and Robinson 2004, Hernandez and Robinson 2007).

Comparison of the relative intensities of the dimer and tetramer suggests that lowering the pressure from 6.8 mBar to 4.7 mBar negatively effects the transmission of Hb tetramer, $D11^{+}$ intensity is higher than $T16^{+}$ in the 4.7 mBar spectra compared to the 6.8 mBar spectra, where dimer intensity is lower than that of tetramer. It is possible, however, that the observed loss of intensity is as a result of peak broadening, and not of a loss of ion transmission. A comparison of the integrated peak area of the $T16^{+}$ peak and $D11^{+}$ peak, representative of the number of ions transmitted by the instrument, indicates a $D11^{+}$:$T16^{+}$ ratio of 1:5.2 at 6.8 mBar compared to 1:4.1 at 4.7 mBar. Whilst this does indicate a minor loss of ion

transmission with changing backing pressure, it is not as significant as first suggested based on ion intensity.

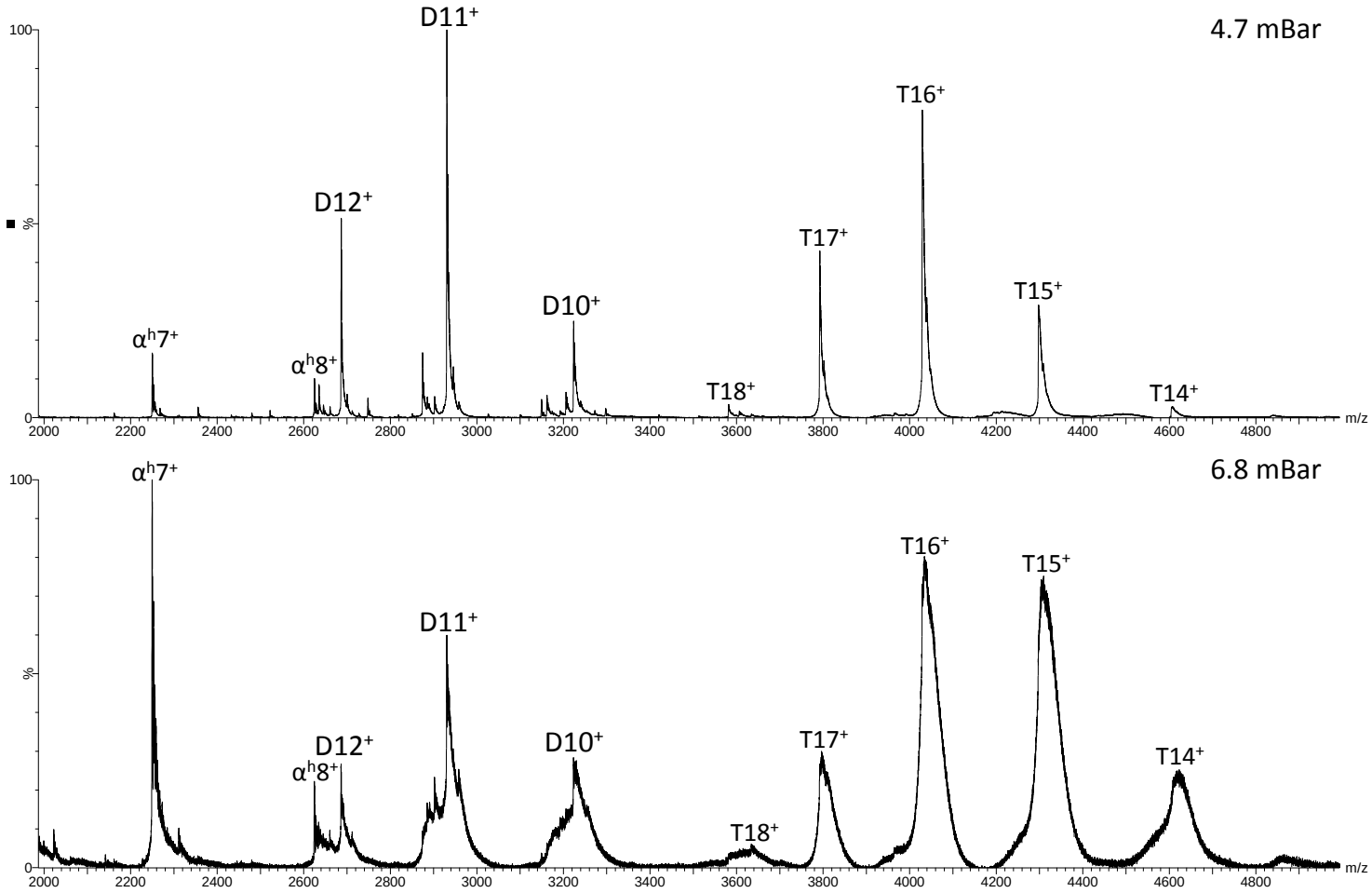


Figure 4-9 – Native MS spectra of HbA illustrating the effect of varying backing pressure on the MS spectrum between low pressure (4.7 mBar) and high pressure (6.8 mBar). At increased source pressure relative intensity of the monomer appears much higher, and relative peak width increases. This suggests a worse signal-to-noise ratio for the tetramer, and may suggest some degree of collisional activation of the haemoglobin dimer/ tetramer to produce monomer.

Figure 4-9 also indicates an increase in the relative intensity of monomer at 6.8 mBar is higher than that observed at 4.7 mBar. Increasing the gas pressure in the cell induces a number of low energy collisions which serve to reduce the axial and radial kinetic energy of the ion (Benesch, Ruotolo *et al*. 2007). The optimum gas pressure is known to be protein specific, based on a proteins physical properties, (Tahallah, Pinkse *et al*. 2001), and increasing the pressure can increasing the internal kinetic energy, or temperature, of the ion to the point that collisional activation can occur (Konijnenberg, Butterer *et al*. 2013). It is therefore likely that the increase in monomer is as a result of the activation of either the dimer or the tetramer. Interrogation of the spectra provided no evidence for the presence of an Hb trimeric species, suggesting additional monomer is formed from the activation of the dimer.

Comparison of the estimated collision cross sections

As can be seen in Figure 4-10, dropping the pressure significantly improves the quality of the CCS calibration. At lower pressures we see a shift in the corrected drift times of the calibration standards; the drift time window changes from 3.1 – 7.5 ms to 1.6 – 4.51 ms. This shift is as a result of a reduction in the collision focusing; fewer gas collisions results in ions with a higher axial kinetic energy which allows them to traverse the mobility cell faster. There is also an observed shift in the values observed for the cytochrome C ions. At higher pressures the cytochrome C ions behave in a different manner, illustrated by the negative correlation between the ions seen in Figure 4-10 B, compared with Figure 4-10 A. This suggests that, at higher pressures, the monomeric cytochrome C may be undergoing gas phase activation as a result of the increased collisions, affecting its structure.

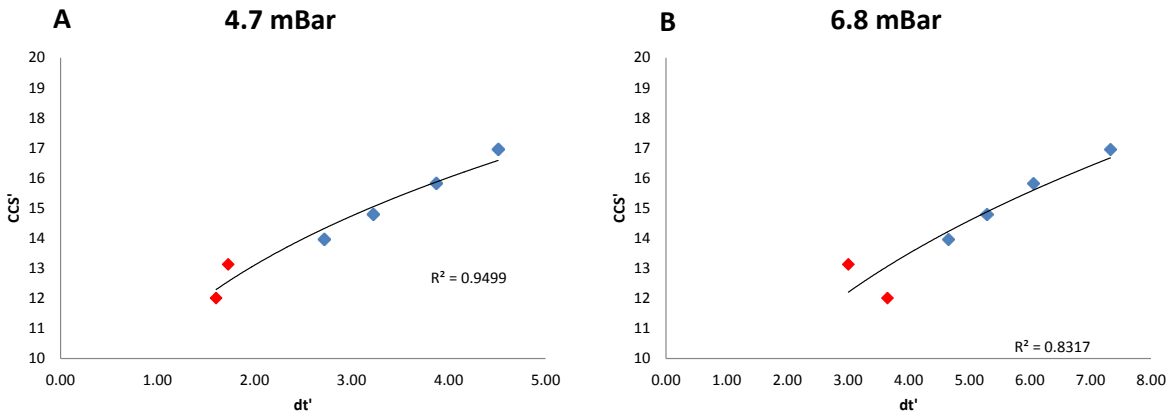


Figure 4-10 - Calibration curves obtained using cytochrome C (red) and bovine serum albumin (blue) at the A) 4.7 mBar and B) 6.8 mBar

The result of the reduced pressure sees the quality of fit of the calibration curve increase from 0.83 to 0.95, making it far more suitable for obtaining CCS estimates.

A comparison of the CCS estimates for HbA obtained using the calibration at each pressure is shown in Figure 4-11.

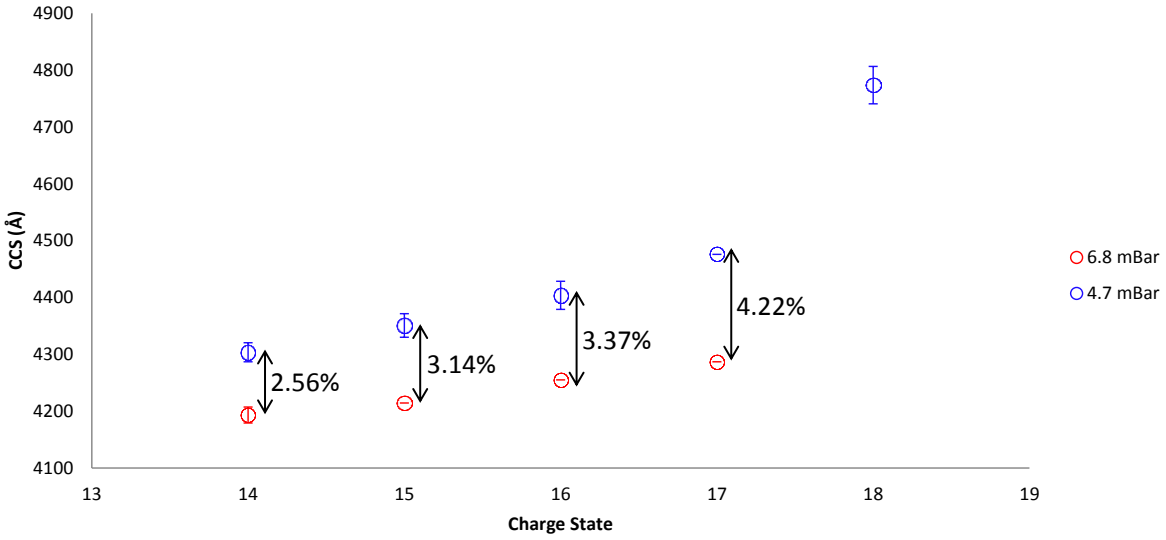


Figure 4-11 – a comparison between the CCS estimates of HbA obtained using the calibration standard shown in Figure 4-10. At 4.7 mBar the better calibration leads to a minor increase in the estimated CCS values.

The improved calibration leads to CCS estimates consistently higher for the 14^+ - 17^+ charge states of HbA. The two data sets do, however, show similar trends suggesting that there is no major change in the conformational ensemble of the proteins as a result of changing the pressure. A poor signal-to-noise ratio, as a result of improper desolvation at the higher pressures, resulted in poor data for the 18^+

charge state at 6.8 mBar. By changing the pressure regime is possible to resolve this peak and estimate a CCS value.

4.2.4 Validating the collision cross sections

Haemoglobin tetramer

A comparison between four computational methods PA, EHSS, PSA and TM (discussed in more detail in 1.1.9 Computational modelling of CCS) and the experimentally estimated CCS values (obtained using 4.7 mBar source pressure and 37^{o}C source temperature) is shown in Figure 4-12. PDB entry 4HHB was used to create the computational CCS values (crystal structure shown in Figure 4-1). The computational CCS value produced was obtained using He as the buffer gas. To make it more applicable to the experimental dataset the computational model has been corrected to represent a pseudo-N_2 value. An increase in CCS of 10% was chosen as the correction factor, based on the observed differences in the He and N2 experimental data shown in Bush *et al* (Bush, Hall *et al*. 2010). Previous groups have reported the use of molecular dynamics to energy minimise X-ray crystal structures prior to performing CCS estimations (Ruotolo, Benesch *et al*. 2008). Due to the computational demands of MD modelling, no energy minimisation has been performed on these models.

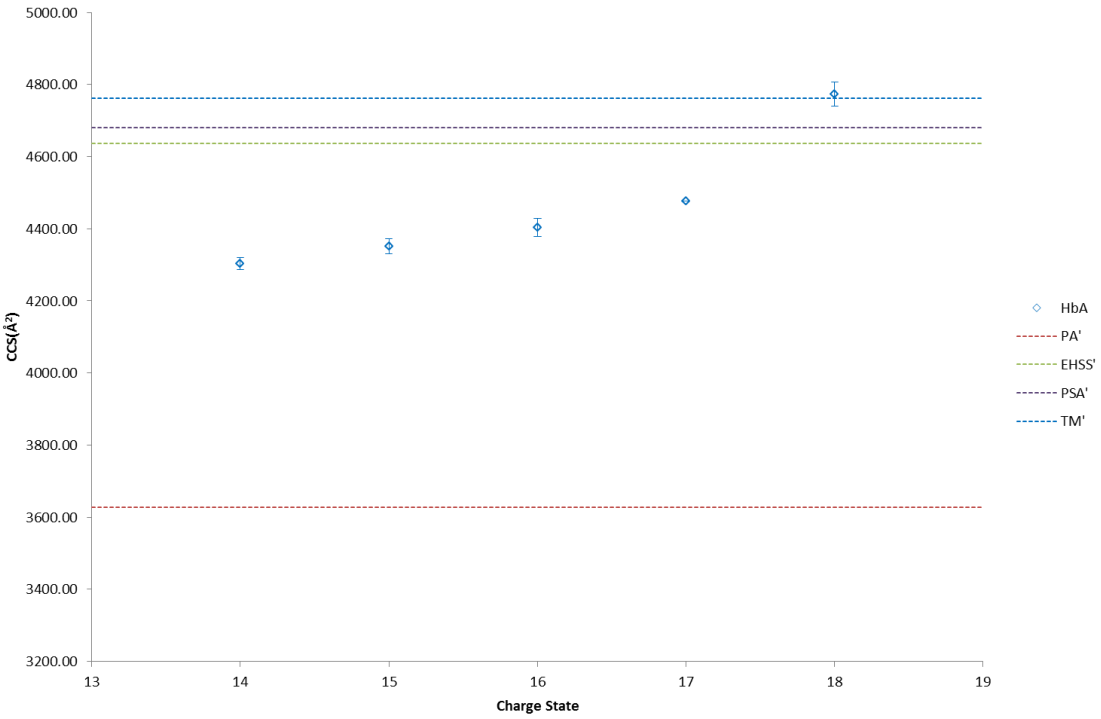


Figure 4-12 – A comparison of the CCS values for HbA estimated from experimental data and calculated from PDB entry 4HHB. PA', EHSS', TM' and PSA' computational methods all represent the CCS increased by 10% to represent pseudo-N_2 theoretical models.
The experimental CCS calculated for the 14^+ charge state is lower than the computational methods, consistent with previous reports. With increasing charge, the CCS estimate increases, consistent with protein unfolding as a result of charge.

The TM method is the most robust method of CCS estimation, and therefore represents the best computational CCS value. The data shown in Figure 4-12 shows a difference between the CCS estimate of the 14^+ ion and the TM' CCS value is 9%, indicating a significantly smaller CCS in solution. The 18^+ charge state of HbA shows good agreement with the calculated TM' CCS value (0.25%) suggesting that under these experimental conditions the 18^+ charge state represents a less compact and possibly more native-like conformation than the lower charge states.

As seen previously, the PA method underestimates cross section values for HbA. The TM method is regarded as the most robust computational method, but it is by far the most computationally demanding. For HbA tetramer the TM method took approximately 6 days to complete the calculation using a standard desktop PC of modest computational power (Intel Dual Core CPU, 2.8/2.8 GHz and 2.98 Gb RAM). The EHSS method took approximately 6 hours to complete and provided a CCS which showed a 2% difference to the experimental CCS; the PSA method took

approximately 12 hours to complete and provided a CCS which agreed to 1.1% with experimental values.

Whilst the TM method is thought to be the most robust computational method, its computational time makes it unsuitable to use if speed is a key factor in analysis. The PSA method, which has been applied here to haemoglobin for the first time, shows a good agreement with the TM method. The PSA would therefore be considered a good alternative to the TM method.

Experimental CCS values for HbS were compared with computational values obtained using all four methods. PDB entry 2HBS, Figure 4-13, was used to generate the computational models. As can be seen in Figure 4-13, PDB entry 2HBS consists of a dimer of the two tetramers.

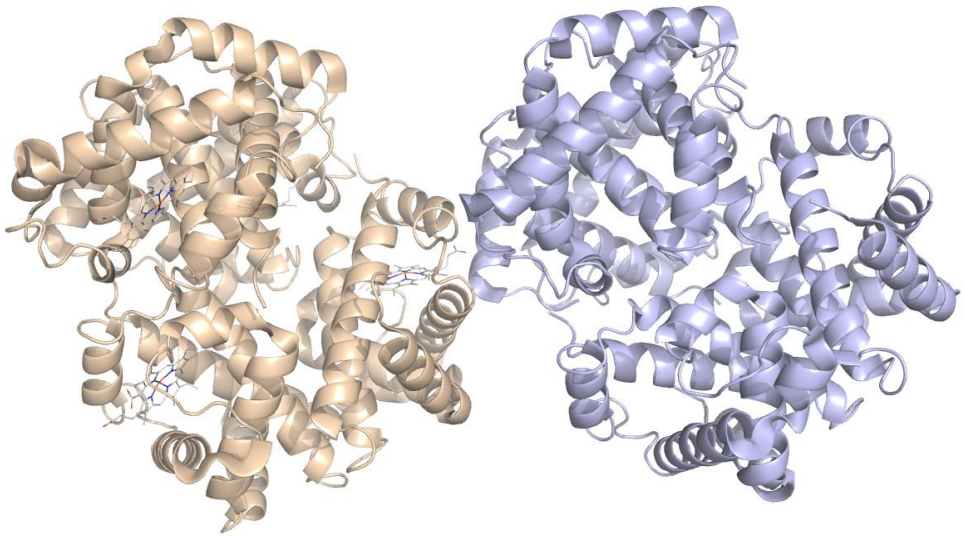


Figure 4-13 – The crystal structure of PDB entry 2HBS

Of these two tetramers, only one of the pair contains coordinates for the prosthetic heme groups. Pymol was used to remove the atomic coordinates for the tetramer containing no heme (in blue) to create a file with only a single haemoglobin molecule. These atomic coordinates were then used to create the CCS estimate. As before, no molecular modelling has been performed on the PDB files and the computational values have been corrected by +10%.

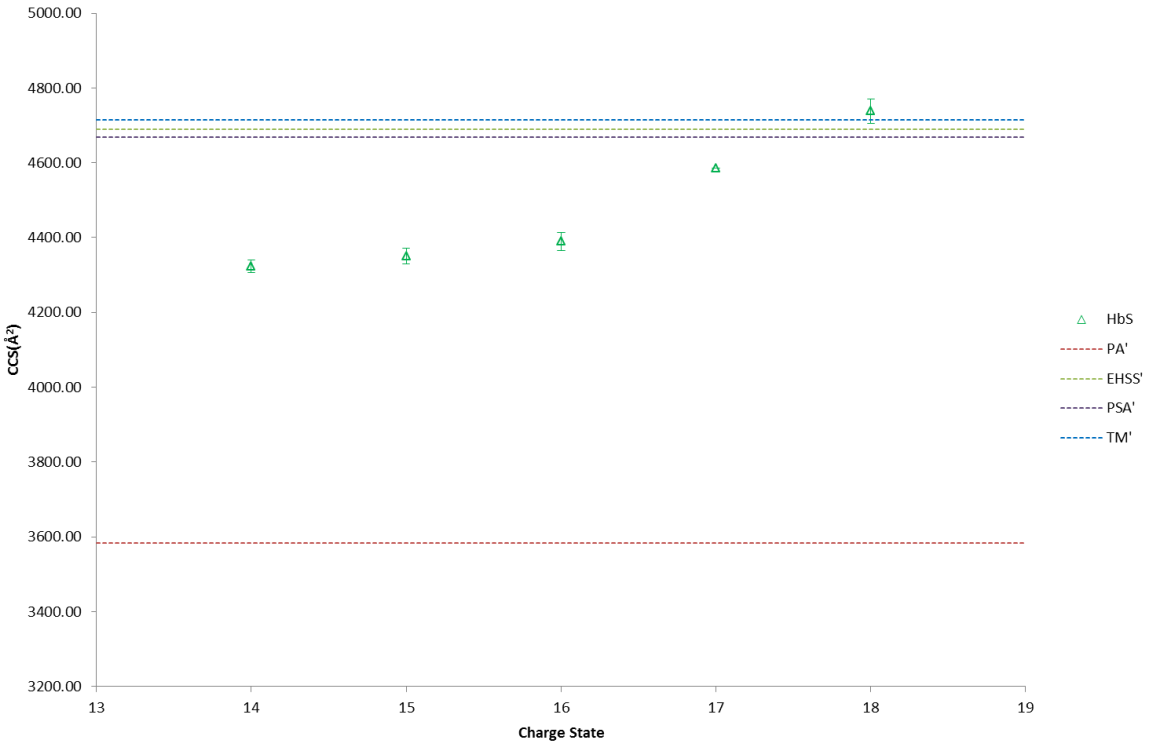


Figure 4-14 – A comparison of the CCS values for HbS estimated from experimental data and calculated from the modified PDB entry 2HBS using PA', EHSS', TM' and PSA' based on the corrected computational CCS values.
The experimental CCS calculated for the 14^+ charge state is lower than the computational methods, consistent with previous reports. With increasing charge, the CCS estimate increases, consistent with protein unfolding as a result of charge.

As with the HbA data, Figure 4-14 shows that the CCS estimate of the 14^+ charge state of HbS exhibits a large difference to that of the computational models. Once again there is a strong correlation between the 18^+ experimentally estimated CCS value and TM CCS values, with a 0.5% difference, suggesting this may represent a more native-like conformation. Once again the PSA method for HbS provided a CCS value in good agreement (with a difference of 1.81%) the TM method. Computational times for HbS for the four methods were similar to as those presented for HbA.

Haemoglobin dimer

Atomic coordinate data for the αβ-dimers were created using Pymol to remove half the atomic coordinates from the intact tetramer, leaving only a single αβ-dimer with prosthetic heme groups (see Figure 4-15). As with Hb tetramer data, CCS estimates were corrected to account for the differences between He and N_2 buffer gases. A correction factor of 12% was applied for Hb dimers. This is based on the observation

that proteins with smaller molecular weight/ Ω-values show larger differences between He and N_2 observed Ω obtained on a drift cell instrument (Bush, Hall *et al*. 2010).

Figure 4-15 – The crystal structure of the HbA αβ-dimer (from PDB: 4HHB)

The HbA dimer CCS data, Figure 4-16, shows a similar trend to the tetramer data; the lowest charge state (10^+) shows a poor agreement with the computational CCS values, differing by 15%, whereas the 12^+ charge state shows better agreement with the EHSS method (2.7%) and the TM method (1.3%). This again suggests a gas-phase compaction of the dimer, consistent with previous observations for the tetramer.

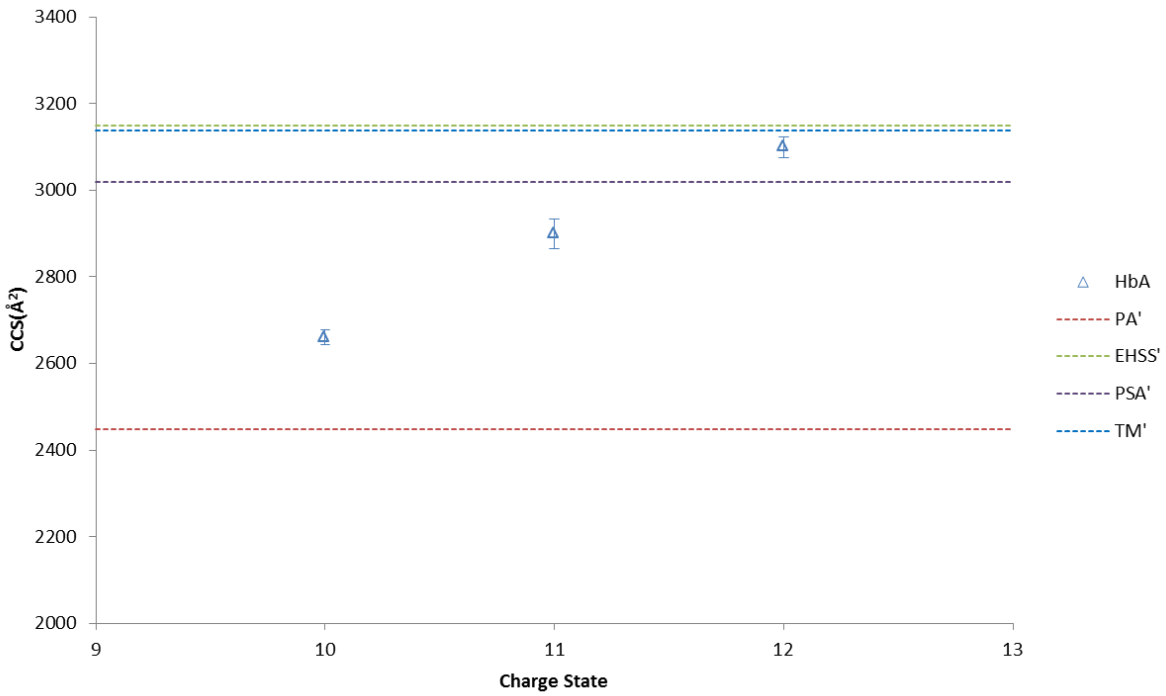


Figure 4-16 – A comparison of the CCS values for HbA dimer estimated from experimental data and calculated from the modified PDB entry 4HHB using PA', EHSS', TM' and PSA' based on the corrected computational CCS values. The experimental CCS calculated for the 10^+ charge state is lower than the computational methods, consistent with previous reports. With increasing charge, the CCS estimate increases, consistent with protein unfolding as a result of charge.

The estimated HbA dimer CCS is, however, larger than half that of the tetramer. The tetramer CCS value is 4304±17 $Å^2$ compared to 2660±17 $Å^2$ for the lowest charge states (dimer is 61% of the monomer). It may be expected that this difference would be approximately 50%, given that the dimer is half the size of the tetramer. Haemoglobin dimers have an extensive overlap when they come together to form the tetramer, shown in Figure 4-17. This overlap means the diameter of the tetramer at its largest point, the value obtained when performing any CCS measurement, would be expected to be less than the sum of the diameter of two dimers.

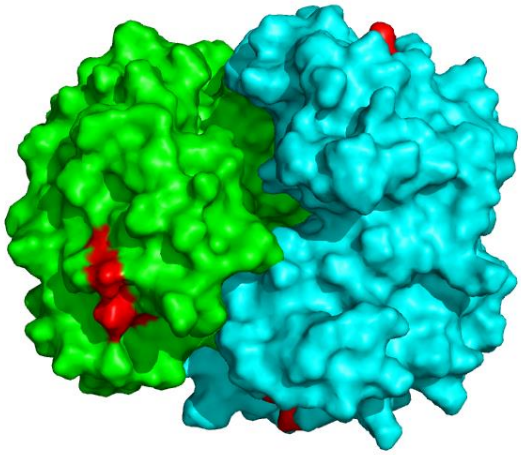


Figure 4-17 – The crystal structure on HbA (4HHB) showing the two dimers (blue and green), highlighting the overlap at the interface site. Heme cofactors are coloured in red.

For the HbS dimer the observations are similar (Figure 4-20).

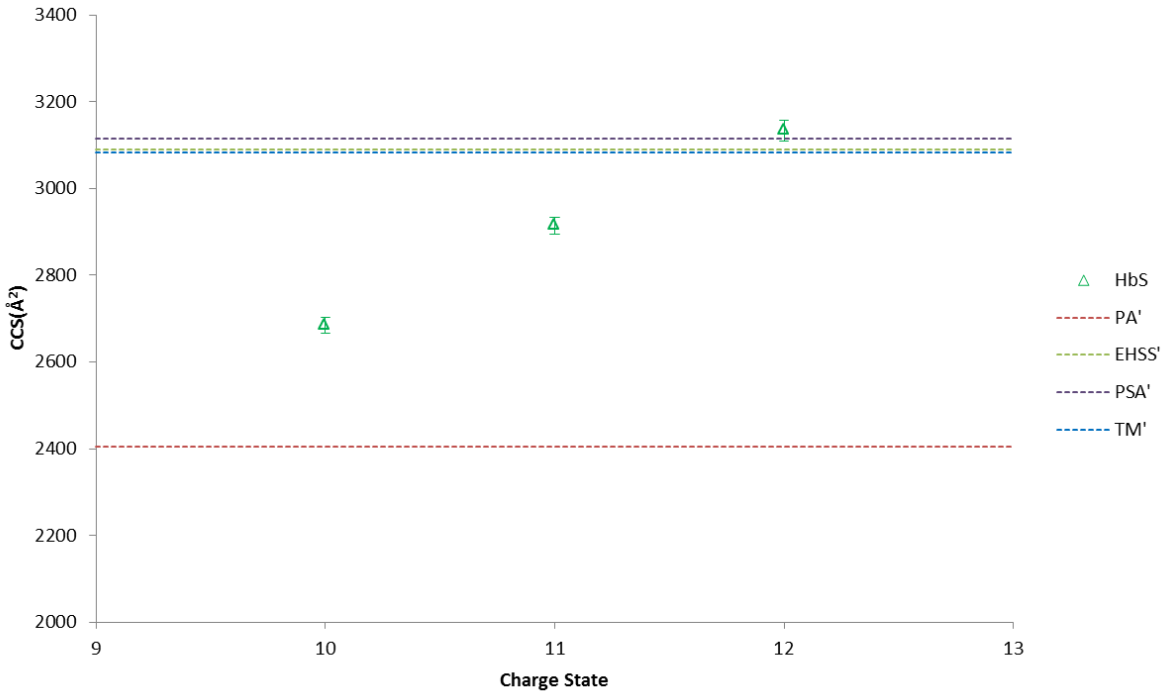


Figure 4-18 – A comparison of the CCS values for HbS dimer estimated from experimental data and calculated from the modified PDB entry 4HHB using PA', EHSS', TM' and PSA' based on the corrected computational CCS values.
The experimental CCS calculated for the 10^{+} charge state is lower than the computational methods, consistent with previous reports. With increasing charge, the CCS estimate increases, consistent with protein unfolding as a result of charge

The 10^+ charge state of HbS shows a worse agreement with computational models (approx. 12% difference between experimental and EHSS, TM and PSA) than the 12^+ charge state (approx. 1.5% difference between experimental and EHSS, TM and PSA), consistent with previous observations.

Both HbA and HbS tetramer and dimer show large differences between the lowest charge state and the computational models. The Barran group have reported that for a number of proteins the CCS of the lowest charge state is consistently lower than that of the CCS measurement obtained from crystal structures (Jurneczko and Barran 2011), which is consistent with the observations in Figure 4-13, Figure 4-14, Figure 4-16 and Figure 4-18. It has also been observed, however, that higher charge states tend to have experimental CCS values higher than CCS values obtained from crystal structures. The highest charge state observed here is in good agreement with the computational CCS values, suggesting some form of error. It has been observed that ion mobility CCS measurements can be dependent on both TWIMS conditions and molecular CCS (Bush, Hall *et al*. 2010, Michaelevski, Eisenstein *et al*. 2010). Michaelevsi *et al* suggest that optimum T-wave conditions are 300-400 m/s wave velocity and 7-10 V wave height. This data set makes used of a wave velocity of 400 m/s and a wave height of 18 V, based on the original Scarff *et al* publication (Scarff, Patel *et al*. 2009). It has previously been suggested that higher wave height voltages can lead to gas phase compaction of a protein (Michaelevski, Eisenstein *et al*. 2010). Whilst this observation has been challenged more recently (Salbo, Bush *et al*. 2012), Salbo *et al* did not perform any experiments at such high wave height voltages. It is possible that the observed discrepancy between the calculated CCS values and the experimental CCS values are as a result of a wave height-dependant effect on the CCS. It is also possible that the choice of calibration standards is having an adverse effect on the estimated CCS value. Salbo *et al* recommend the use of calibration standards that carefully bracket the masses and drift time of the analyte (Salbo, Bush *et al*. 2012). Whilst care was taken to ensure calibration standards were chosen that bracket the drift time of the tetramer, calibrants were also chosen which bracketed the dimer. Cytochrome C, the low mass calibration standard, may be distorting the CCS estimates as a result of its low mass and drift time, leading to an underestimation of the CCS.

4.2.5 A comparison on normal haemoglobin with mutant variants

Experimentally derived arrival time distributions and estimated CCS values for HBA were compared to those obtained from HbS and HbC in order to investigate any structural differences between the proteins. A comparison of the native MS spectra, shown in Figure 4-4, suggests that these proteins have a similar structure and conformation, based on the charge distribution, and therefore any changes to structure as a result of mutation would likely be subtle, such as a minor compaction/ expansion that would not disrupt the overall surface area of the protein (and thus, its ability to obtain charge from the solvent).

Figure 4-19 shows the arrival time distribution data for the three proteins at each of the 14^+ – 18^+ charge states observed on the MS spectra.

At the 14^+ charge state HbA and HbS show the similar times, suggesting a similar conformation. There is, however, a noticeable increase in the FWHM of the peak – HbS has a 0.2 ms wider ATD than HbA suggesting that, whilst the two variants adopt a similar conformation, HbS may have a higher degree of structural flexibility within that conformation. The 14^+ charge state in HbC, by comparison, has a lower drift time (4.1 ms) than either HbA or HbS (4.36 ms) suggesting a more compact conformation. HbC also has the broadest ATD (0.9 ms compared to 0.5 ms and 0.7 ms for HbA and HbS, respectively) suggesting that it may have a higher degree of conformational flexibility. The nature of the mutation may provide some insight into this increased destabilisation. In HbA the β6 amino acid is the negatively charged glutamic acid. In HbS this glutamic acid has been mutated to neutral valine, whereas in HbC the glutamic acid has been mutated to positively charged Lysine. The change from a negative side chain in HbA to a positive side chain in HbC is likely to have a significantly larger impact on protein structure than the mutation to a neutral side chain (HbS), since it will change the electrostatic interactions occurring at the local environment.

As the charge on the protein is increased all three molecules show a decrease in observable drift times through to 18^+. The observed peak broadening in HbS and HbC remains consistent with increasing charge, supporting the theory that these two variants have increased flexibility as a result of the mutations. There is an observed peak shouldering in all three samples at 15^+, suggesting the emergence of a second,

more extended, conformation. This shoulding appears more prominent in HbA by virtue of the narrower FWHM in the HbA sample.

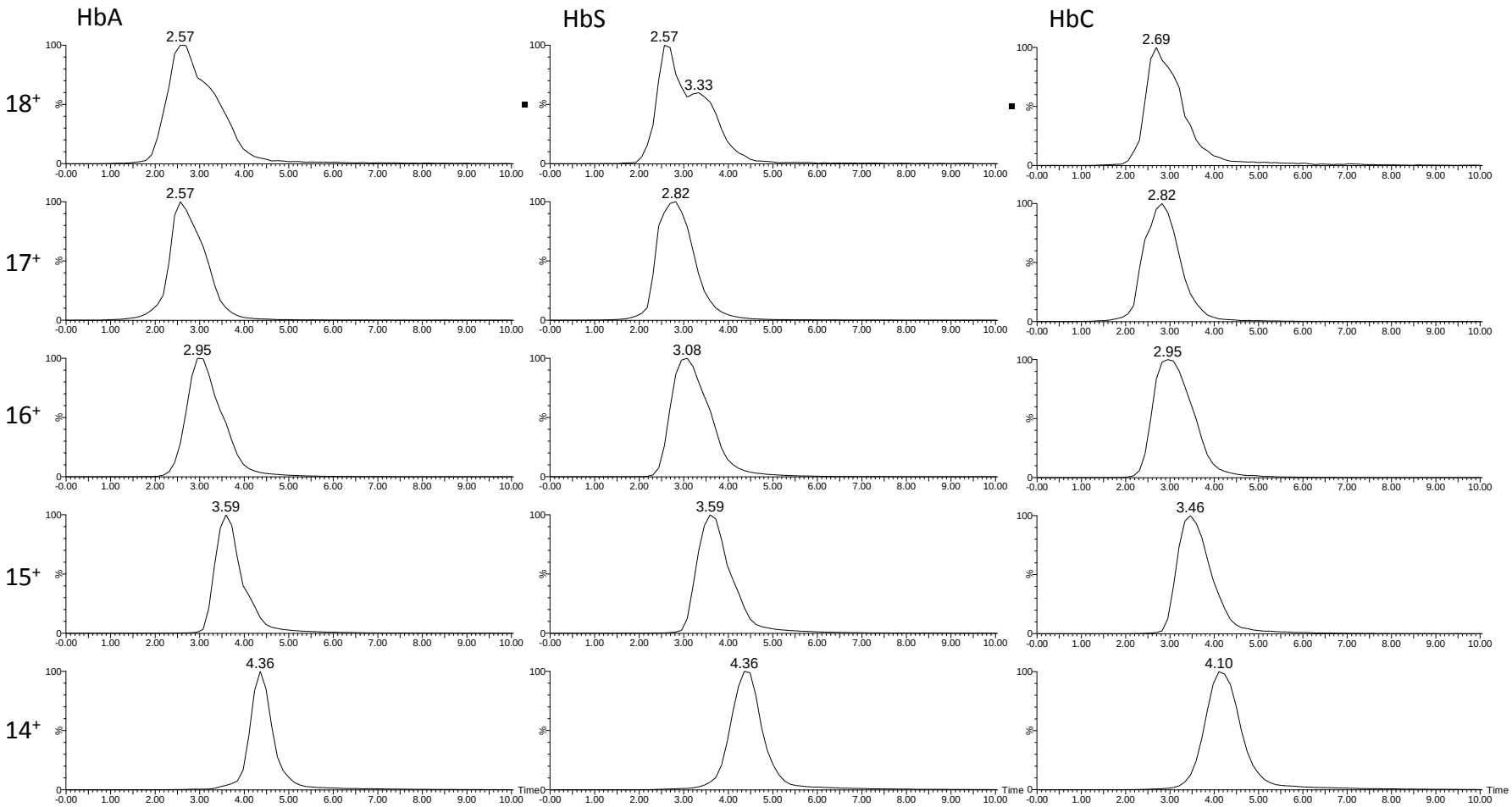


Figure 4-19 – A comparison of the arrival time distributions of HbA, HbS and HbA at the different charge states.
HbA and HbS 14^+ charge states shows simialr drift times, whilst HbC shows a a smaller drift time, suggesting HBC may adopt a more compact confromation. HbA, HbS and HbC show evidence of ATD shouldering at higher charge states, consistent with the emergence of additional, extended, conformations. The relative intenisty of the shouldering is higher in HbS and HbC samples, suggesting a potential difference in gas phase stability.

For the 17^+ charge state there is a shift in the arrival times of HbS and HbC; both proteins exhibit a larger drift time than HbA suggesting a shift to a more extended conformation in the mutants relative to HbA. For the 18^+ charge state both HbS and HBA show similar arrival times, suggesting a shared conformational ensemble, with the HbC ion having a slightly longer arrival time than the other two mutants. All three proteins, at this stage, show clear signs of a secondary conformational ensemble; in the HbS sample this is most evident by observation of a second peak with an arrival time of 3.33 ms, but it can also be seen by the prominent peak shouldering in HbA and HbC. This additional species is at a higher intensity in HbC than in HbA, but cannot be fully resolved in either sample. The presence of these extended conformations illustrates the ability of the protein to unfold and form stable, gas phase structures (Scarff, Thalassinos *et al*. 2008).

Using the calibration curve shown in Figure 4-10 arrival times were converted to CCS values to provide enable direct comparison. CCS values were calculated for the most intense drift times present in Figure 4-19 for all charge states. A comparison of HbA and HbS CCS values is shown in Figure 4-20, a comparison of HbA and HbC is shown in Figure 4-21, a comparison of HbS and HbC is shown in Figure 4-22.

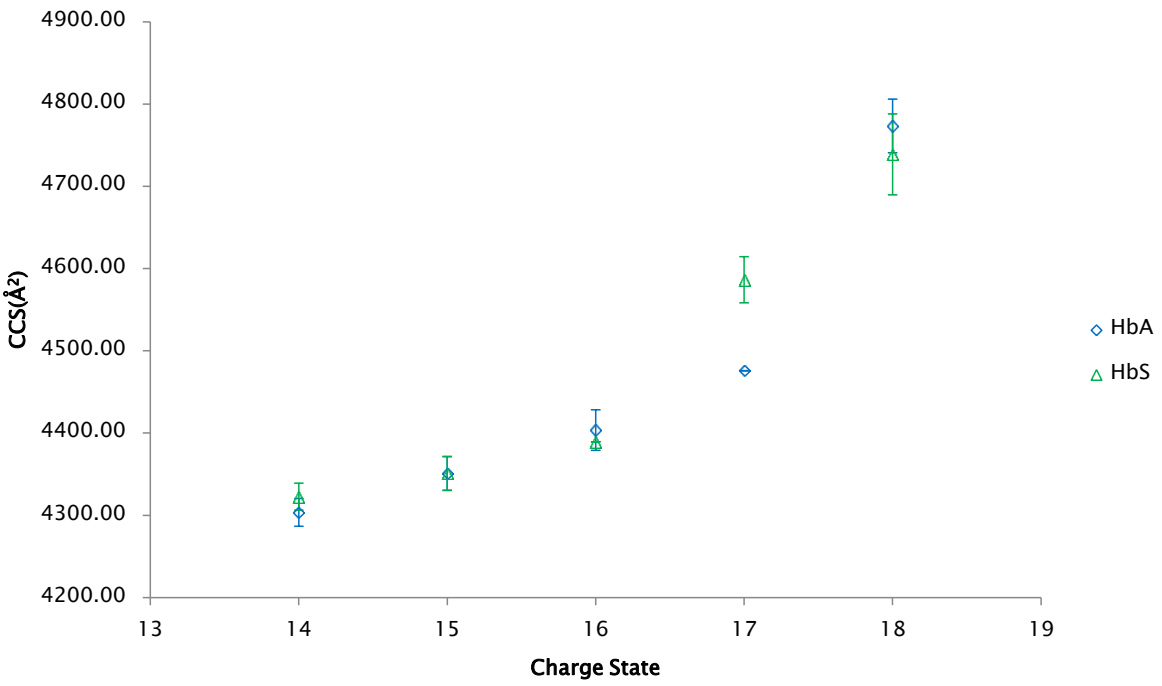


Figure 4-20 – A comparison of the experimental HbA and HbS CCS values. Both HbA and HbS show increased CCS with charge, consistent with previous observations. HbA and HbS show no differences in CCS at all charge states except the 17^+ charge state. This may suggest differences in the gas phase stability, as HbS more readily adopts a more extended conformation.

Figure 4-20 indicates that both HbA and HbS have similar cross sections in the 14^+ charge states. This is in contrast to the work previously published (Scarff, Patel *et al*. 2009) which indicated HbS had a larger cross section than HbA. The data presented here suggests that, whilst the two proteins have a similar CCS within experimental error, HbS estimates are still marginally larger than those of HbA. It has already been shown that the improved calibration framework and sample preparation methods provide CCS estimates closer to those predicted by computational methods. Based on this observation, it is likely that the small differences observed here were amplified as a result of the previous calibration.

With increasing charge we observe an increase in cross section, consistent with previous publications (Scarff, Thalassinos *et al*. 2008, Scarff, Patel *et al*. 2009) indicating that with increasing charge a protein begins to unfold. Whilst HbA and HbS exhibit similar CCS values for the 14^+, 15^+, 16^+ and 18^+ charge states, there is a significant difference in the 17^+ charge state. HbS has a significantly larger CCS value than HbA at the 17^+ charge state consistent with the increase in drift time shown in Figure 4-20. This suggests that, at the 17^+ charge state, HbS may adopt a

more extended gas phase structure than HbA, however at the 18^+ charge state we observe similar CCS values again. Whilst HbA and HbS do follow a similar gas phase unfolding pathway, suggested by broadly similar drift times/ arrival time distributions and CCS estimates, HbS is able to adopt this extended conformation more readily.

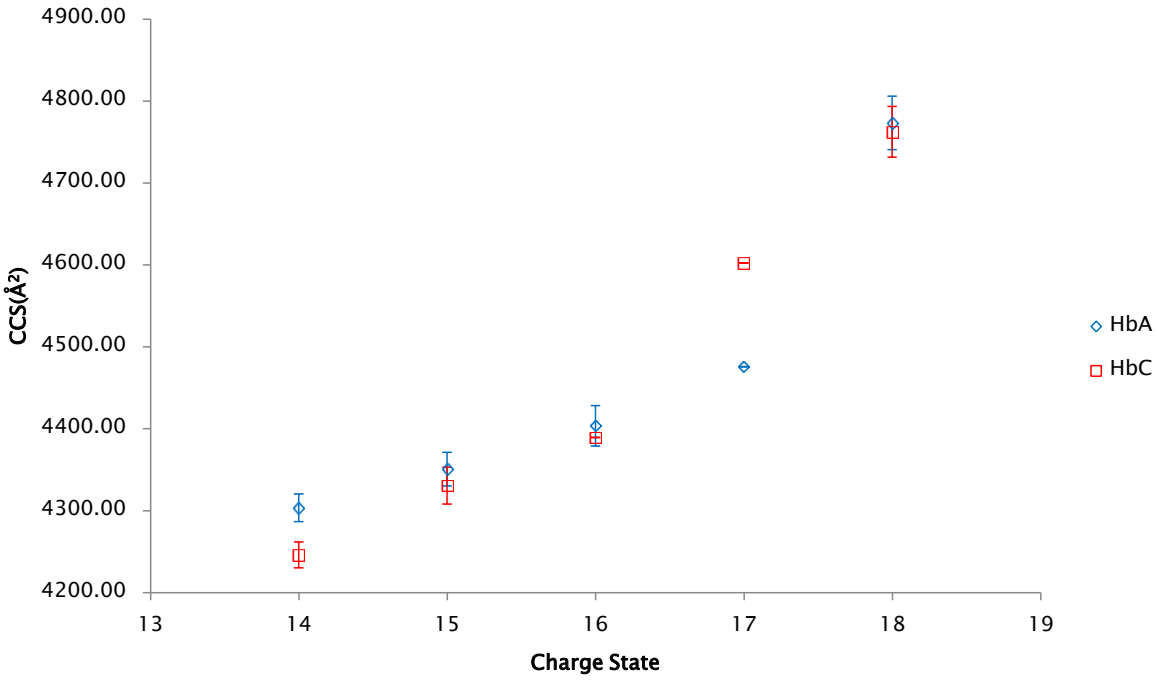


Figure 4-21 – A comparison of HbA and HbC CCS value.
Both HbA and HbC show increased CCS with charge, consistent with previous observations. HbC appears to have a smaller CCS than HbA at the 14^+ charge state, suggesting a more compact conformation. As with HBS, there is an observed difference between HbC and HbA at the 17^+ charge state. This may suggest differences in the gas phase stability, as HbC more readily adopts a more extended conformation.

A comparison of HbC with HbA suggests that the 14^+ charge state of HbC has a smaller CCS value, consistent with the lower drift time observed in Figure 4-21. With increasing charge the cross section of HbC increases, consistent with the HbA and HbC data. Interestingly, HbC also exhibits a larger cross section than HbA at the 17^+ charge state. From Figure 4-22 illustrates that HbC 17^+ charge state adopts a similar cross section to that of HbS at the 17^+ charge state, indicating a potentially shared conformational ensemble. This suggests that both HbS and HbC follow a similar gas phase unfolding pathway when unfolding due to increased charge. For the 18^+ charge state HbA and HbC have similar CCS values, suggesting they both occupy similar conformational ensembles at higher charges.

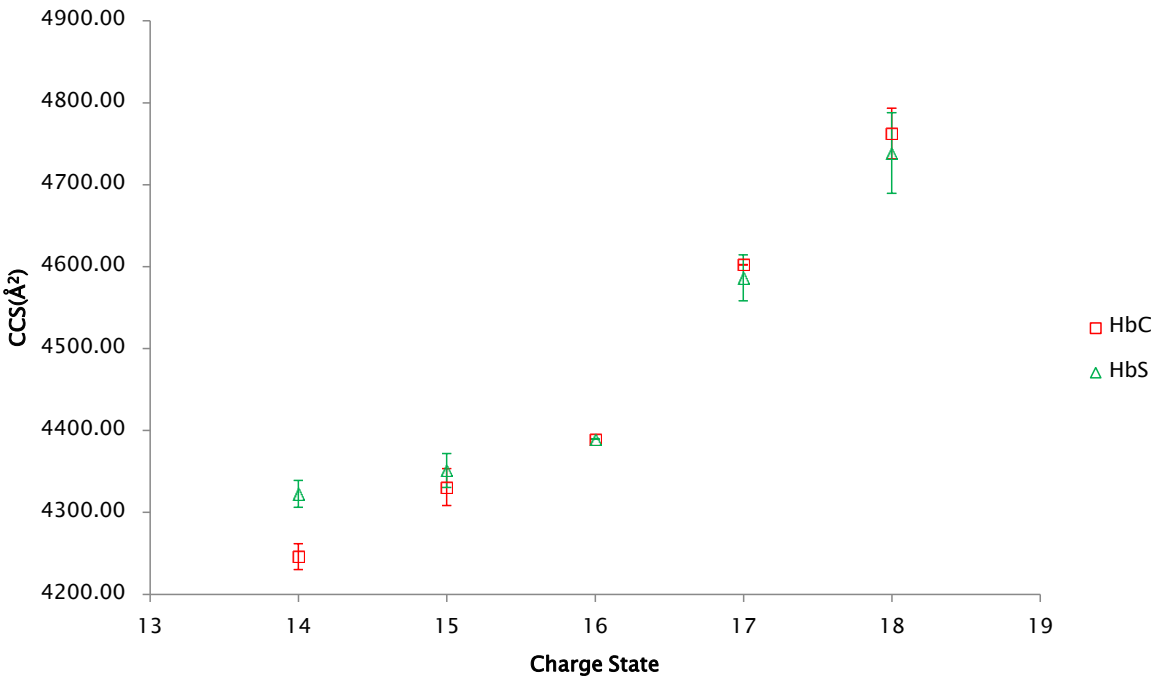


Figure 4-22 – A comparison of HbC and HbS CCS value.
HbC appears to have a smaller CCS than HbS at the 14^+ charge state, suggesting a more compact conformation. HbS and HbS appear to have the same CCS at the 17^+ charge state, suggesting a shared, extended gas phase conformation.

The effects of mutation were also investigated on the dimers and the results are shown in Figure 4-23 and Figure 4-24. Both Figure 4-23 and Figure 4-24 indicate that conformational differences observed in the tetramer are not observed in the dimer for either the native state, or as a result of unfolding due to increased charge.

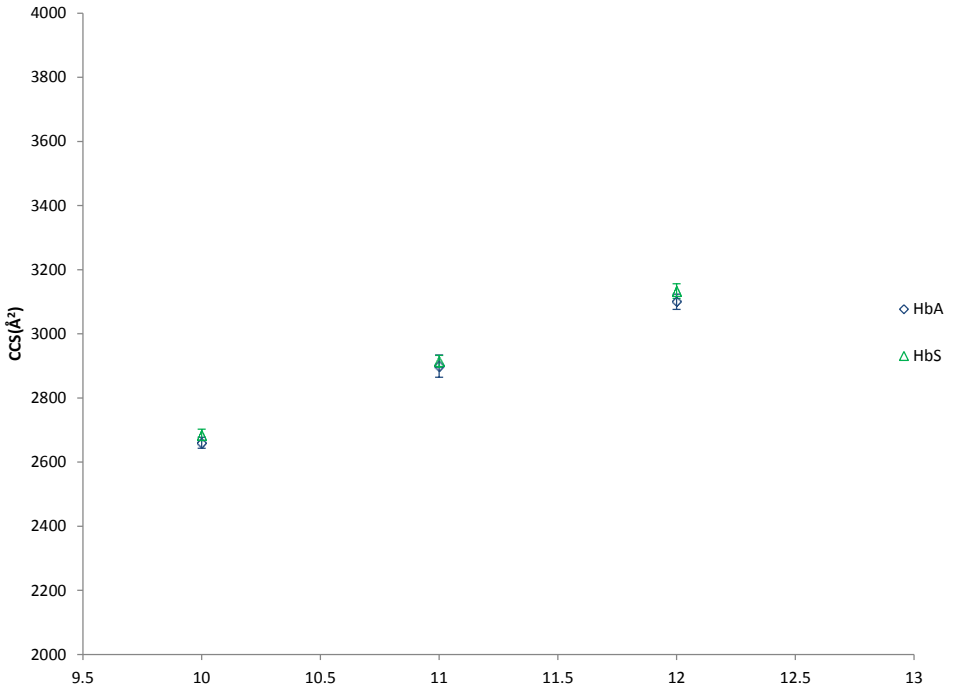


Figure 4-23 – A comparison of HbA and HbS CCS value for the dimer species.
Bothe HbA and HbS dimers have the same CCS at all charge states, indicating no conformational differences

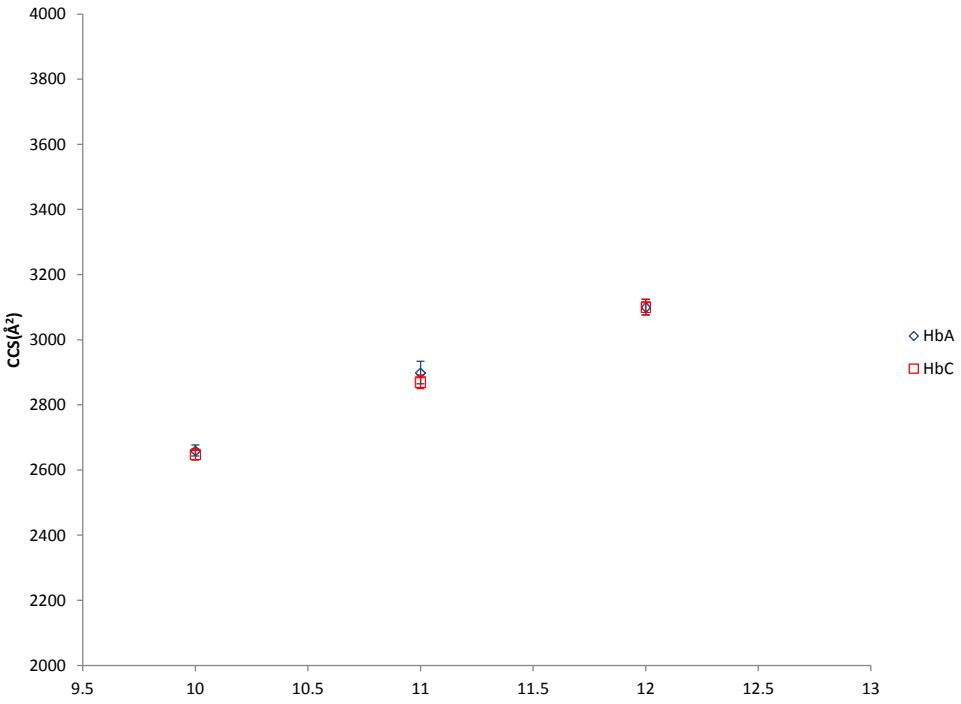


Figure 4-24 – A comparison of HbA and HbS CCS value for the dimer species.
Bothe HbA and HbS dimers have the same CCS at all charge states, indicating no conformational differences

4.2.6 Collisional activation of haemoglobin tetramer

In order to investigate the observed differences in the unfolding transitions of haemoglobin collisional activation was used to induce unfolding in the protein. Collision activation is described in more detail in Introduction section: Collisional Activation but in summary, increasing the acceleration voltage of the protein as it enters various regions of the cell results in a number of higher energy collisions which can induce dissociation. In this study the aim was not to induce dissociation, but to increase the internal kinetic energy of the protein via these collisions in order to induce unfolding much in the same way as temperature could be using to induce unfolding. The voltage to the skimmer cone was systematically increased between 60V and 200V; the results for HbA are shown in Figure 4-25.

It has already been observed that increasing the skimmer cone voltage leads to collisional activation and dissociation in a similar manner to that of CID (Benesch, Ruotolo *et al*. 2007), and that is evident in Figure 4-25 as increasing the cone voltage leads to a loss of intensity of the HbA tetramer, and an increase in the relative intensities of the dimer and monomer species.

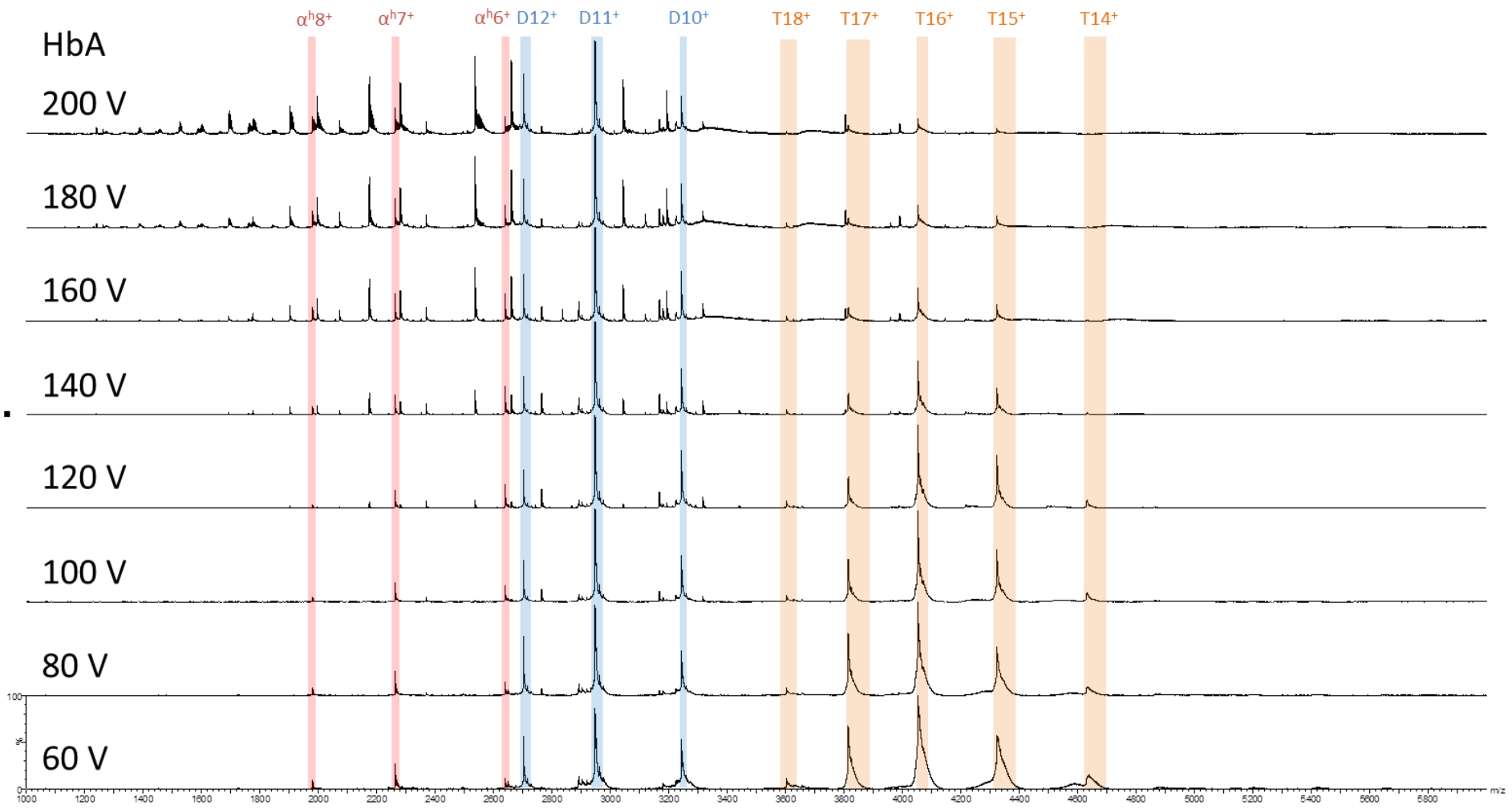


Figure 4-25 – The native MS spectra of HbA at varying cone voltages.
With increasing cone voltage the relative intensity of the tetramer (orange) decreases, and the intensity of the dimer (blue) and monomer (red) increases. At cone voltages higher than 120 V a number of additional peaks begin to appear, resulting from the fragmentation of the protein backbone by CID.

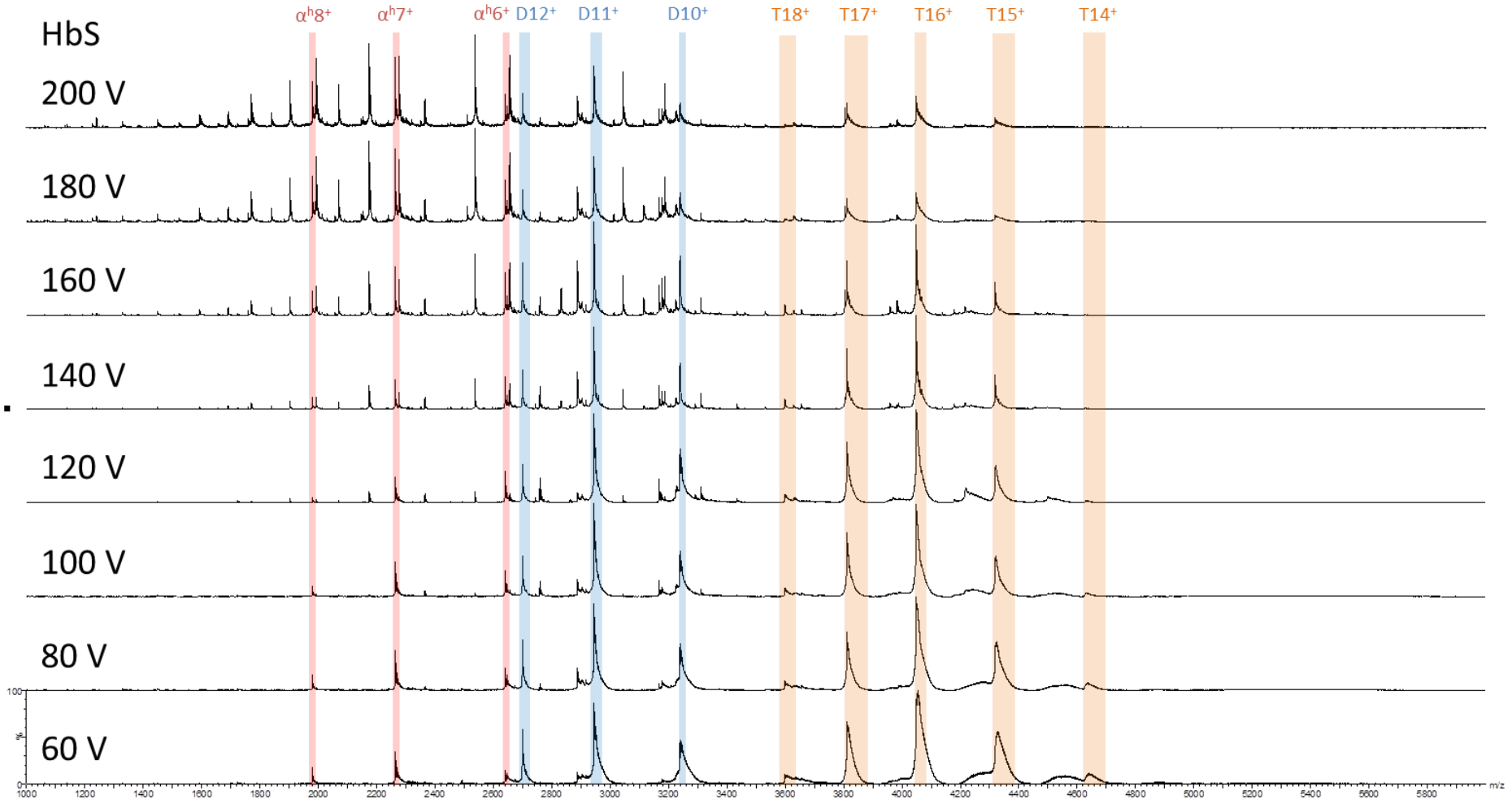


Figure 4-26 - The native MS spectra of HbS at varying cone voltages.
With increasing cone voltage the relative intensity of the tetramer (orange) decreases, and the intensity of the dimer (blue) and monomer (red) increases. At cone voltages higher than 120 V a number of additional peaks begin to appear, resulting from the fragmentation of the protein backbone by CID.

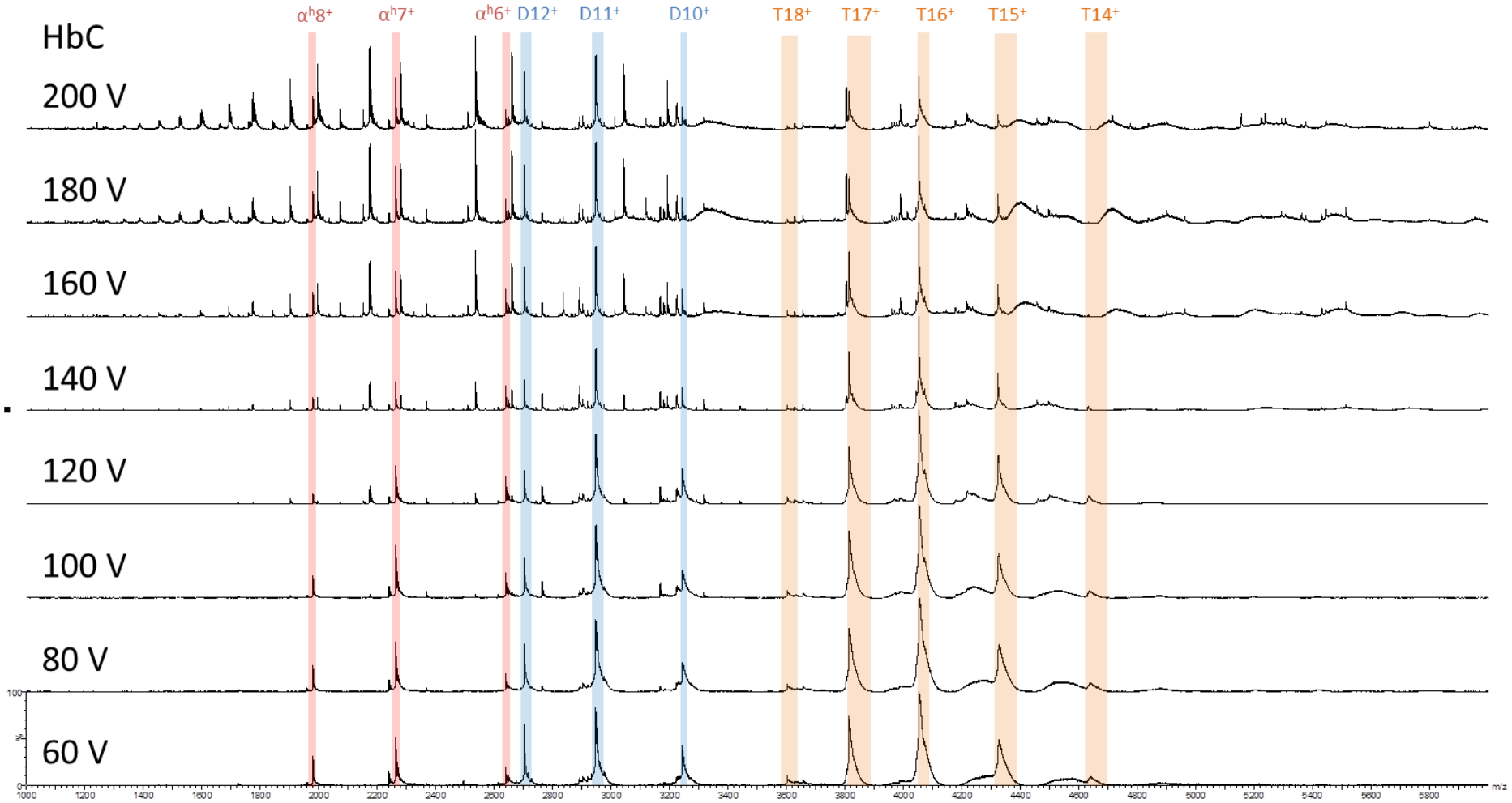


Figure 4-27 - The native MS spectra of HbC at varying cone voltages.
With increasing cone voltage the relative intensity of the tetramer (orange) decreases, and the intensity of the dimer (blue) and monomer (red) increases. At cone voltages higher than 120 V a number of additional peaks begin to appear, resulting from the fragmentation of the protein backbone by CID.

Figure 4-26 and Figure 4-27 show results for HbS and HbC, respectively that are similar; HbS and HbC spectra show a far higher degree of fragmentation, based on the intensities of the multiply charged ions observed below 2500 m/z. From this is can be concluded that these both HbC and HbS more readily fragment than HbA, however which oligomeric state is fragmenting cannot be accurately determined without further experiments. These results are in agreement with those previously observed for the reduction of stability in the diseased variants of haemoglobin relative to normal haemoglobin.

Figure 4-27 – 4-29 show that tetramer intensity drops off with increasing charge. It is also clear that above 160 V, the native-like 14^+ ion of the tetramer has a significant of signal intensity; reducing to below 10% in all cases. At these values the peak has a very poor signal: noise ratio making it difficult to analyse the ion mobility data for the ion. The arrival time distributions of the 14^+ ions for HbA, HbS and HbC were therefore compared for 60 V → 160 V and are shown in Figure 4-28.

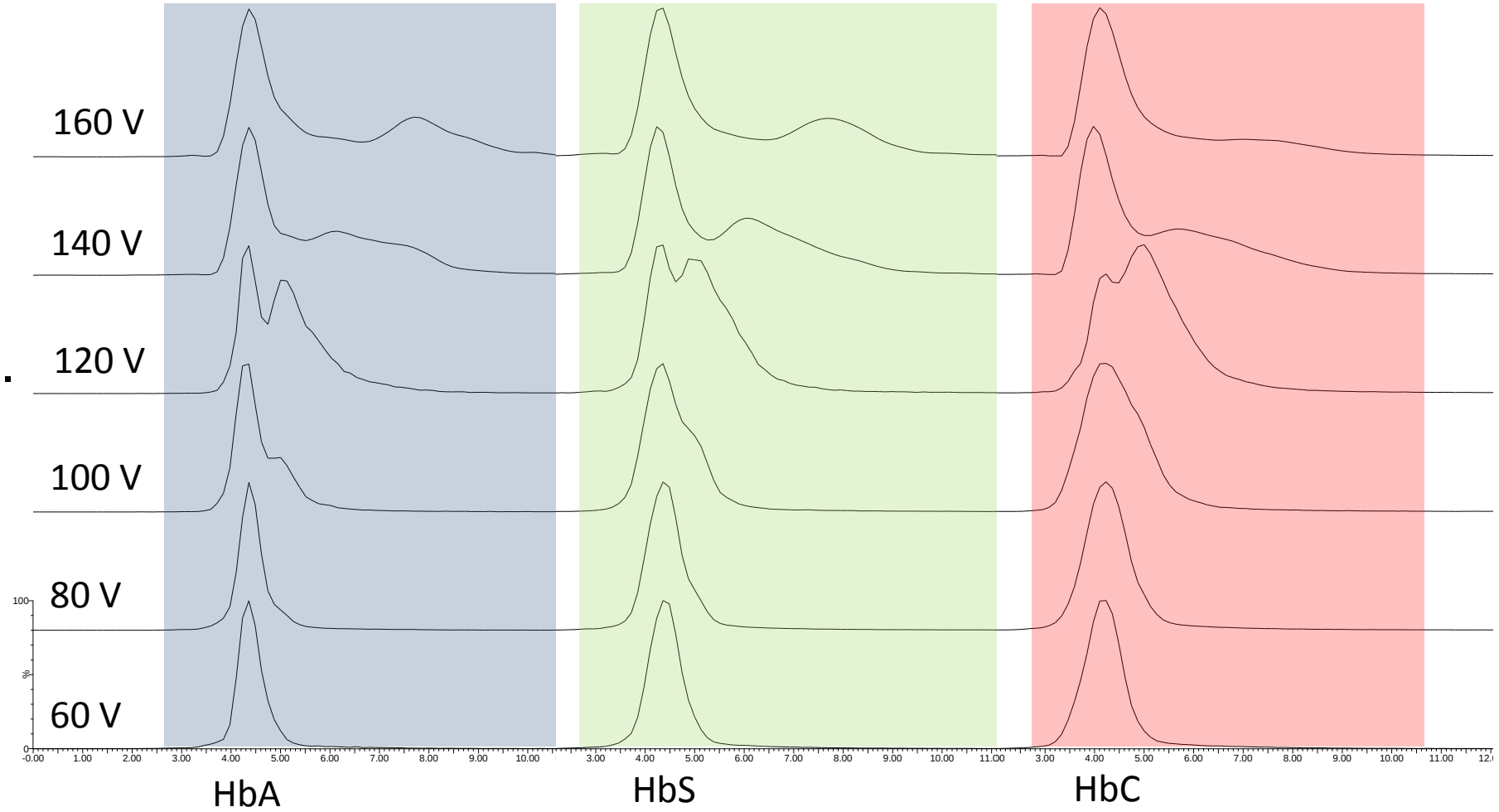


Figure 4-28 - The drift time chromatograms for HbA (blue), HbS (green) and HbC (red) showing changes in arrival time distributions with increasing cone voltage for the 14^{+} charge state.
Increasing cone voltage results in the emergence of a number of additional, extended conformations. The relative intensity of these extended conformations in more intense in the HbS and HbC ATDs relative to the HbA ATD

The HbA data shows a single conformation at 60 V for the 14^+ ion, consistent with previous observations. As the skimmer cone voltage increases a shoulder appears on the ATD, starting at 80 V but shown most prominently at 100 V, suggesting the presence of a more extended conformation that cannot be fully resolved from the native conformation. At 120 V this extended conformation is fully resolved and is itself showing evidence of shouldering, suggesting a number of further unfolded subpopulations. By 140 V and 160 V this large peak further extends, resulting in a number of smaller unresolved peaks. The data suggests that, with collisional activation, HbA can be made to undergo gas phase unfolding to form first one, and then a number of extended gas phase structures. Each of these extended populations have broad ATDs, indicating dynamic conformations; consistent with a less folded structural state. Interestingly, for HbA the native conformation, present at 60 V, remains the dominant peak at all cone voltages, indicating that there is incomplete unfolding of the HbA. Increased collisional activation may push the equilibrium of folded: unfolded protein towards the unfolded state, but at 160 V there is still not enough energy to completely force the equilibrium towards the unfolded. There are still some changes evident in the native conformation. At the higher energy levels (140 V and 160 V) the native conformation shows a small increase in drift time, suggesting the equilibrium may be shifting towards a more extended conformation, and the ATD is broader, suggesting the conformation has become more dynamic.

HbS shows a similar trend to that observed for HbA. At 60 V the ATD is broader, consistent with the previous data. The shouldering observed in the HbA sample at 80 V and 100 V is also present, as is the additional peak at 120 V. In HbS, however, each of these features appears at higher relative intensities than in the HbA sample, suggesting they are more easily induced in HbS. At 140 V and 160 V there are multiple, unresolved peaks which are observed at a higher relative intensity in HbS than in HbA. These observations support the previous suggestion that HbS is more readily able to unfold compared to HbA.

HbC shows a similar trend to that observed for HbA and HbS; the formation of a shoulder in the ATD at 80V $\rightarrow$100 V which becomes a resolved peak by 120 V. This extended population has a higher relative intensity in HbC compared to both HbA and HbS suggesting that HbC is the most easily unfolded of the three proteins. This is most evident at 120 V where the native conformation is no longer the dominant

peak in the ATD, having been replaced with the more extended conformation. At 140 V and 160 V it appears as if the intensity of these extended conformations is lower than in the previous examples. Whilst the intensities are lower, the overall arrival time window of all the conformations of HbC is wider, suggesting that HbC may adopt a more dynamic extended conformation at a lower intensity at these higher activation energies. This is consistent with the suggestion that HbC more readily unfolds than HbA or HbS.

Figure 4-29 – 32 show the same unfolding data for the 60 – 200 V unfolding for the 15^+ - 18^+ charge states.

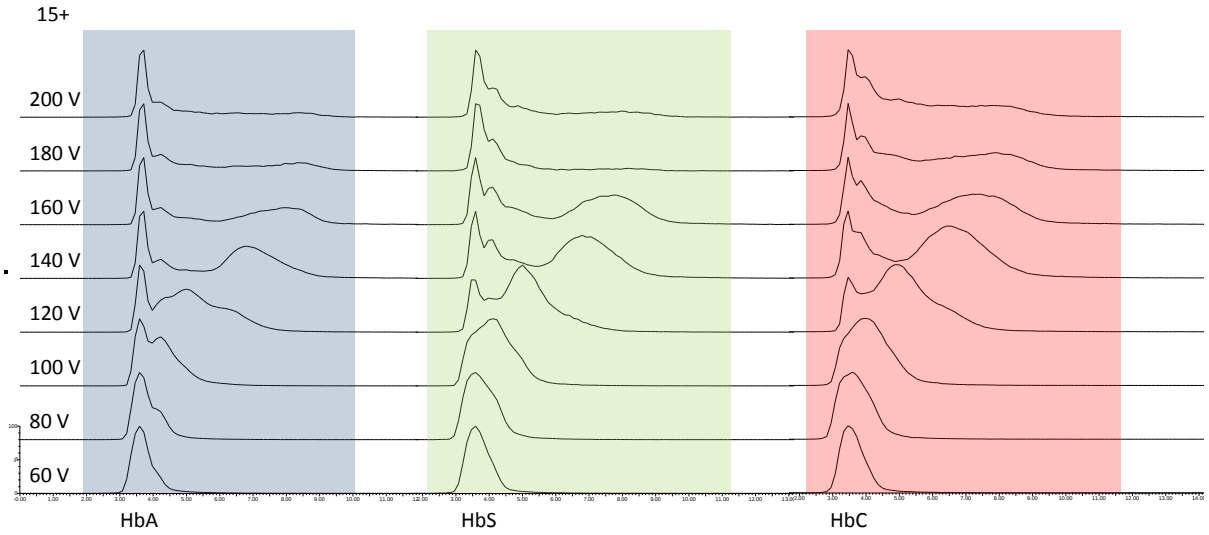


Figure 4-29 – The drift time chromatograms for HbA (blue), HbS (green) and HbC (red) showing changes in arrival time distributions with increasing cone voltage for the 15^+ charge state. Increasing cone voltage results in the emergence of a number of additional, extended conformations. The relative intensity of these extended conformations in more intense in the HbS and HbC ATDs relative to the HbA ATD

The 15^+ charge state (Figure 4-29) shows a similar trend to that observed for the 14^+ charge state. All three proteins show a similar ATD as a factor of unfolding, with the HbS and HbC samples showing higher intensity of the extended conformations relative to HbA. The same is true of the 16^+ (Figure 4-30) and 17^+ (Figure 4-31) charge states. At 17^+ charges the HbS and HbC samples show evidence that more extended conformations are more intense than the more compact. This supports the previously shown data which discussed the differences between the HbA CCS and both the HbS and HbC CCS values (Figures 4-20, 4-21 and 4-22).

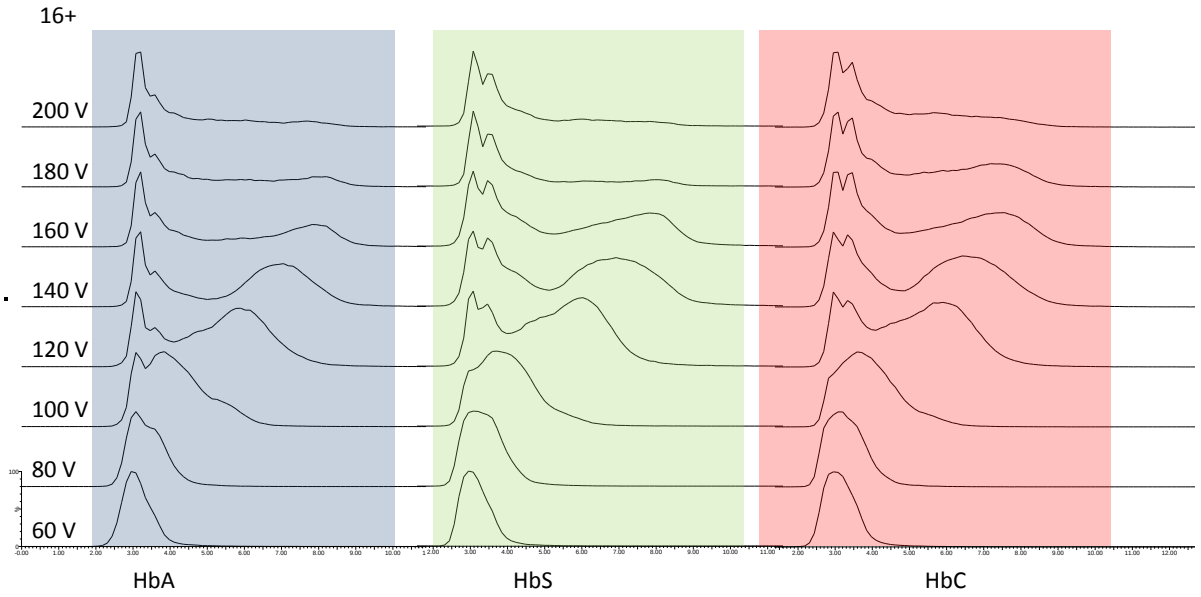


Figure 4-30 - The drift time chromatograms for HbA (blue), HbS (green) and HbC (red) showing changes in arrival time distributions with increasing cone voltage for the 16^{+} charge state Increasing cone voltage results in the emergence of a number of additional, extended conformations. The relative intensity of these extended conformations in more intense in the HbS and HbC ATDs relative to the HbA ATD

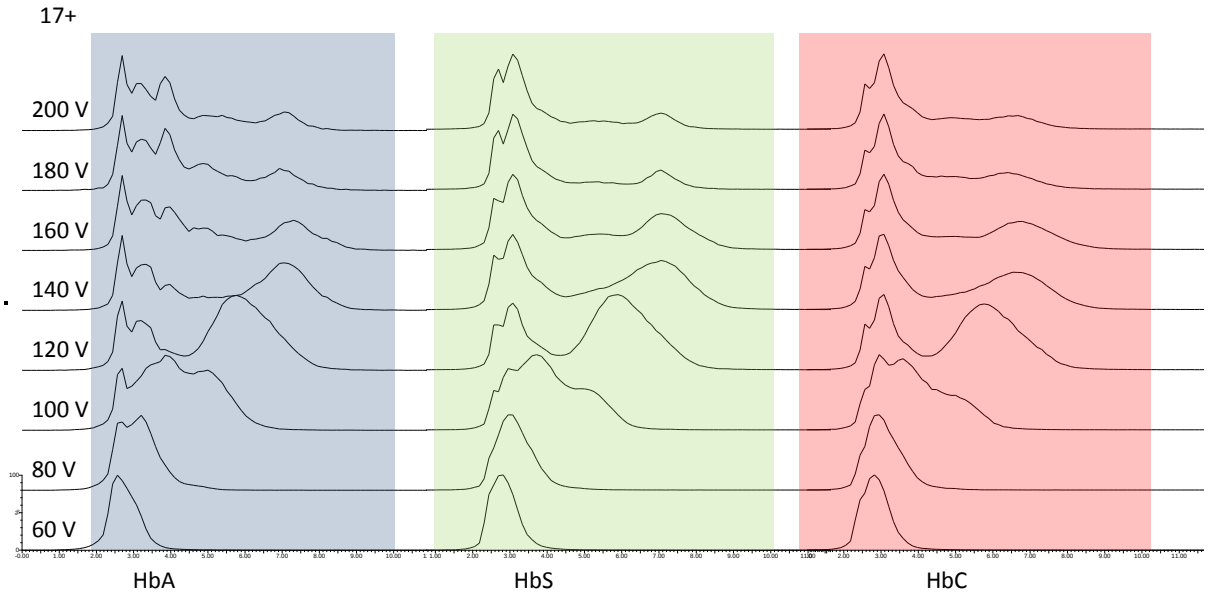


Figure 4-31 - The drift time chromatograms for HbA (blue), HbS (green) and HbC (red) showing changes in arrival time distributions with increasing cone voltage for the 17^{+} charge state. Increasing cone voltage results in the emergence of a number of additional, extended conformations. The relative intensity of these extended conformations in more intense in the HbS and HbC ATDs relative to the HbA ATD

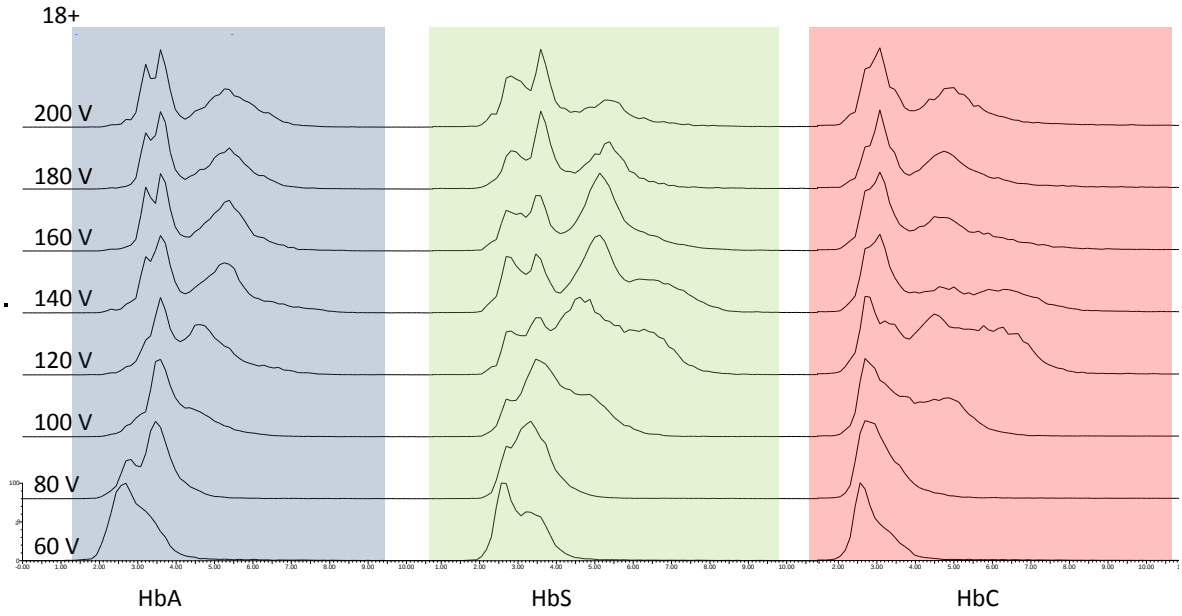


Figure 4-32 - The drift time chromatograms for HbA (blue), HbS (green) and HbC (red) showing changes in arrival time distributions with increasing cone voltage for the 18^+ charge state Increasing cone voltage results in the emergence of a number of additional, extended conformations. The relative intensity of these extended conformations in more intense in the HbS and HbC ATDs relative to the HbA ATD

The 18^+ charge state shows the largest differences between the HbA and both HbC and HbS samples. Up until this point the data has suggested that, whilst normal and diseased haemoglobin unfolded at different rates, they ultimately followed a similar unfolding pathway based on the observed similarity between the CCS values and the ATDs. The 18^+ data suggests that, when unfolded due to increased charge, the three proteins show a similar ATD profile, and therefore have a similar CCS value, as indicated by Figures 4-20, 4-21 and 4-22. When the 18^+ charge state is unfolded by collisional activation there are some key observed differences between the samples. In this instance, the HbA sample seems more readily able to unfold; the most compact conformation is lost by 120 V and is replaced by three dominant extended conformations. In contrast, HbS and HbC both show evidence of this more compact 18^+ structure throughout the activation process, with evidence of the compact shoulder persisting until 200 V. In contrast to previous observations, HbC retains this compact structure to the greatest degree. In addition, HbS appears to be able to adopt a number of different extended conformation which do not appear to a lesser extent in HbA, and at different intensities to those of HbC; these differences are is most evident at 120 V.

4.3 Conclusions

Ion mobility mass spectrometry has become a useful tool for studying protein structure and dynamics. Since commercialisation a number of groups have worked on methods for improving the experimental data, and it was the aim of this study to revisit a previous publication by Scarff *et al* from 2009, and apply recently described improvements in experimental protocols to that study. By using gentler sample preparation and lower source temperatures and pressure it has been possible to obtain spectra of haemoglobin with lowest observed charge states and a better resolution, obtaining accurate masses to within a 2 Da of the theoretical.

This improved sample preparation method has been used to analyse the two previously studied haemoglobin variants, HbA and HbS, alongside HbC which has not previously been analysed by native MS or IMS. The data presented here indicated that differences between HbA and HbS initially reported were limited by the extrapolation of calibration employed, and the cross sections these two protein adopt may be similar than initially reported. HbC does however have a significantly smaller estimated CCS value than HbA or HbS, suggesting a more compact conformation.

All three proteins show an increase in CCS as a factor of charge, however HbS and HbC show larger CCS values for the 17^+ charge states than HbA suggesting a possible stability difference between normal and diseased protein. The variants may more easily adopt unfolded structures. All three haemoglobin samples do share similar CCS values at the 18^+ charge state suggesting a possible shared mechanism for gas phase unfolding. This has been further investigated using collisional activation. The data obtained supports the inferences obtained from the CCS estimations; HbS and HbC do indeed appear to unfold more readily in the gas phase, adopting similar conformations to HbA but at a higher relative intensity. It can be suggested that HbA, HbS and HbC follow broadly similar unfolding pathways. There does appear to be some differences for are ions at higher internal suggesting this shared mechanism may diverge at a certain point.

The results suggest that the β6 mutations causing HbS and HbC induce conformational changes in the protein. For HbC these are more immediately obvious, but for HbS they only manifest as differences in stability. Nevertheless, ion

mobility mass spectrometry has proven useful as a tool for identifying and investigating these differences in molecules with strong structural homology; differences that would not be easily studied using other techniques.

Future perspectives

Collisional activation experiments focused primarily on the changes in the ATD of the haemoglobin tetramer and not the dissociation of the tetramer into its subunits. The molecular assembly of haemoglobin was one area of the 2009 Scarff *et al* paper that has not been reinvestigated here. Given observed stability differences between HbA and HbS it would be interesting to characterise the dissociation energy thresholds of the protein using collisional activation in order to see if there are any changes to the protein-protein interaction strengths induced by the mutations; the data presented here suggests that this may be possible but it is unclear as to whether this is due to monomer-monomer interactions or dimer-dimer interactions.

This leads to two other possible future considerations; self-assembly and collisional activation. All collisional activation in work here has been achieved using the skimmer cone, which is a relatively blunt method of activation. Future work could compare activation using the cone with activation in the instrument collision cell in order to draw parallels and investigate the selectivity of the two activation techniques. With regards to self-assembly, both HbS and HbC are diseases of self-assembly and it would be of interest to look at this mechanism. Collisional activation could also play a role; inducing dissociation of the aggregate may not only provide information regarding assembly mechanisms but also provide an insight into the conformation the protein adopts within the aggregate, allowing an observation of a possible link between stability differences and self-association.

Haemoglobin has only been investigated here in the deoxygenated state whereas in biological systems there is constant flux between oxygenated and deoxygenated states. It would be interesting to investigate conformational and stability differences between the two states. This is of particular interest for HbC since it is only capable of forming pathological crystals in it oxygenated state.

4.4 References

Bain, B. J. (2006). Haemoglobin and the genetics of haemoglobin systhesis. Haemoglobinopathy Diagnosis, Second Edition.

Benesch, J. L. P., *et al*. (2007). "Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics." Chemical Reviews **107**(8): 3544-3567.

Benesch, J. L. P., *et al*. (2003). "Thermal Dissociation of Multimeric Protein Complexes by Using Nanoelectrospray Mass Spectrometry." Analytical Chemistry **75**(10): 2208-2214.

Bleiholder, C., *et al*. (2011). "A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method." International Journal of Mass Spectrometry **308**(1): 1-10.

Boys, B. L. and L. Konermann (2007). "Folding and Assembly of Hemoglobin Monitored by Electrospray Mass Spectrometry Using an On-line Dialysis System." Journal of the American Society for Mass Spectrometry **18**(1): 8-16.

Bush, M. F., *et al*. (2010). "Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology." Analytical Chemistry **82**(22): 9557-9565.

Buxbaun (2007). Haemoglobin and myoglobin: co-operativity. Fundamentals of Protein Structure and Function

Clarke, G. M., and Higgins, T. N. (2000). "Laboratory Investigations of Hemoglobinopathies and Thalassemias@ Review and Update." Clinical Chemistry **46**(8): 1284-1290.

Daniel, Y. A., *et al*. (2005). "Rapid and specific detection of clinically significant haemoglobinopathies using electrospray mass spectrometry–mass spectrometry." British Journal of Haematology **130**(4): 635-643.

Fernandez de la Mora, J. (2000). "Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism." Analytica Chimica Acta **406**(1): 93-104.

Griffith, W. P. and I. A. Kaltashov (2003). "Highly Asymmetric Interactions between Globin Chains during Hemoglobin Assembly Revealed by Electrospray Ionization Mass Spectrometry†." Biochemistry **42**(33): 10024-10033.

Griffith, W. P. and I. A. Kaltashov (2006). "Mass spectrometry in the study of hemoglobin: from covalent structure to higher order assembly." Current Organic Chemistry **10**(5): 535-553.

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

Hunt, J. A. and V. M. Ingram (1958). "Allelomorphism and the Chemical Differences of the Human Haemoglobins A, S and C." Nature **181**(4615): 1062-1063.

Ingram, V. M. (1956). "A Specific Chemical Difference Between the Globins of Normal Human and Sickle-Cell Anaemia Haemoglobin." Nature **178**(4537): 2.

Jurneczko, E. and P. E. Barran (2011). "How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase." Analyst **136**(1): 20-28.

Kirshenbaum, N., *et al*. (2010). "Analyzing Large Protein Complexes by Structural Mass Spectrometry." J Vis Exp(40): e1954.

Konijnenberg, A., *et al*. (2013). "Native ion mobility-mass spectrometry and related methods in structural biology." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(6): 1239-1256.

Lawrence, C., *et al*. (1991). The unique red cell heterogeneity of SC disease: crystal formation, dense reticulocytes, and unusual morphology.

Lessin, L. S., *et al*. (1969). "Molecular mechanism of hemolytic anemia in homozygous hemoglobin C disease. Electron microscopic study by the freeze-etching technique." J Exp Med **130**(3): 443-466.

Mesleh, M. F., *et al*. (1996). "Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential." The Journal of Physical Chemistry **100**(40): 16082-16086.

Michaelevski, I., *et al*. (2010). "Gas-Phase Compaction and Unfolding of Protein Structures." Analytical Chemistry **82**(22): 9484-9491.

Murayama, M. (1967). "Structure of Sickle Cell Hemoglobin and Molecular Mechanism of the Sickling Phenomenon." Clinical Chemistry **13**(7): 578-588.

Ofori-Acquah, S. F., *et al*. (2001). "Mass Spectral Analysis of Asymmetric Hemoglobin Hybrids: Demonstration of Hb FS (α2γβS) in Sickle Cell Disease." Analytical Biochemistry **298**(1): 76-82.

Patriksson, A., *et al*. (2007). "Protein Structures under Electrospray Conditions." Biochemistry **46**(4): 933-945.

Pauling, L., *et al*. (1949). "Sickle Cell Anemia, a Molecular Disease." Science **110**(2865): 543-548.

Putnam, F. (2012). The Plasma Proteins V4: Structure, Function, and Genetic Control, Elsevier.

Ruotolo, B. T., *et al*. (2008). "Ion mobility-mass spectrometry analysis of large protein complexes." Nat. Protocols **3**(7): 1139-1152.

Salbo, R., *et al*. (2012). "Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers." Rapid Communications in Mass Spectrometry **26**(10): 1181-1193.

Scarff, C., *et al*. (2009). "Probing hemoglobin structure by means of traveling-wave ion mobility mass spectrometry." Journal of the American Society for Mass Spectrometry **20**(4): 625-631.

Scarff, C. A., *et al*. (2008). "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 Communications in Mass Spectrometry **22**(20): 3297-3304.

Shimizu, A., *et al*. (2006). "Detection and characterization of variant and modified structures of proteins in blood and tissues by mass spectrometry." Mass Spectrometry Reviews **25**(5): 686-712.

Shvartsburg, A. A. and M. F. Jarrold (1996). "An exact hard-spheres scattering model for the mobilities of polyatomic ions." Chemical Physics Letters **261**(1–2): 86-91.

Sobott, F. and C. V. Robinson (2004). "Characterising electrosprayed biomolecules using tandem-MS--the noncovalent GroEL chaperonin assembly." International Journal of Mass Spectrometry **236**(1-3): 25-32.

Steinberg, M. Z., *et al*. (2007). "The dynamics of water evaporation from partially solvated cytochrome c in the gas phase." Physical Chemistry Chemical Physics **9**(33): 4690-4697.

Tahallah, N., *et al*. (2001). "The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument." Rapid Communications in Mass Spectrometry **15**(8): 596-601.

Thein, S. L. (2011). Abnormalities of the structure and synthesis of hemoglobin. Blood and Bone Marrow Pathology (2nd Edition), Churchill Livingston.

Vekilov, P. G., *et al*. "Intermolecular Interactions, Nucleation, and Thermodynamics of Crystallization of Hemoglobin C." Biophysical Journal **83**(2): 1147-1156.

Wild, B. J., *et al*. (2001). "Rapid Identification of Hemoglobin Variants by Electrospray Ionization Mass Spectrometry." Blood Cells, Molecules, and Diseases **27**(3): 691-704.

Witkowska, H. E., *et al*. (1991). "Sickle Cell Disease in a Patient with Sickle Cell Trait and Compound Heterozygosity for Hemoglobin S and Hemoglobin Quebec–Chori." New England Journal of Medicine **325**(16): 1150-1154.

Chapter 5 : Investigating the structural differences of engineered immunoglobulin Fc domains

5.1 Introduction

5.1.1 Immunity, Antibodies, Structure and History

The Immune Response

The world is full of microorganisms capable of causing disease. These microorganisms are collectively known as pathogens. In order to survive, mammals have evolved a number of strategies capable of preventing a pathogen from causing disease within the body and to purge the pathogen from the body. This response is known as the immune system, and may be broadly separated into two distinct mechanisms based on the types of cell that act to clear the pathogen; the innate immune response and the adaptive immune response. The innate immune response is immediately available to combat pathogens, and is responsible for control of a number of common, less virulent, infections. More aggressive or less common infections, however, are more challenging to overcome. The adaptive immune response has evolved to allow the rapid selection, evolution and proliferation of specialist cells which can specifically target and destroy a single pathogen. The adaptive immune response also confers long-lasting immunity against infection. The two systems do not operate in isolation, but in tandem in order to best combat infections; the innate immune system is responsible for the triggering of the adaptive response which, in turn, targets the pathogen for destruction by the cells of the innate immune response (Janeway 2005).

There are a number of cells and biomolecules that are vital to the immune response. White blood cells, or leukocytes, are the main cell type involved in the immune response, and can be subdivided based on functional differences. The innate immune response is principally comprised of granulocytes and macrophages, which directly destroy pathogens via cellular ingestion, or by the release of toxic agents capable of destroying parasites. The adaptive immune system is principally comprised of lymphocytes, which can be further subdivided into either T-lymphocytes (thymus-derived) or B-lymphocytes (bone-marrow-derived). T-lymphocytes, or T-cells, have two main functional roles; the triggering of apoptosis in cells expressing viral proteins (cytotoxic T-cells) or the activation of additional immune cells (helper T-cells). B-lymphocytes, or B-cells, which, when activated, rapidly differentiate into plasma cells capable of secreting antibodies. Antibody mediated immune response is

also known as the humoral immune response (Janeway 2005). Antibodies (Ab) are vital biomolecules in the immune response, and form the main focus of this research due to their biological, medical and economic impact.

The antibody molecule

In 1890 Emil von Behring discovered that it was an antibody that carried immunity to a given pathogen within the blood of an individual (Von Behring and Kitasato 1890). In 1973 M. Edelmann and R. R. Porter were awarded a Nobel Prize for their contributions in elucidating the structure of antibodies (Nobelprize.org 1972).Their early work focussed on the analysis of proteins and protein fragments obtained from myelomas which turned out to be antibodies. In 1959 R. R. Porter published the first X-ray crystal structure of an antibody fragment; an 80,000 Da fragment obtained from a papain digest which had been shown to crystallise readily (Porter 1959). This was followed by the first description of antibody structure (Fleischman, Porter *et al*. 1963), shown in Figure 5-1. Further studies on gene control (Oudin 1956) and primary structure (R. L. Hill, R. Delaney *et al*. 1966, O'Donnell, Frangione *et al*. 1970, Johnstone and Mole 1977) of this type of antibody provided evidence for the structure of antibodies.

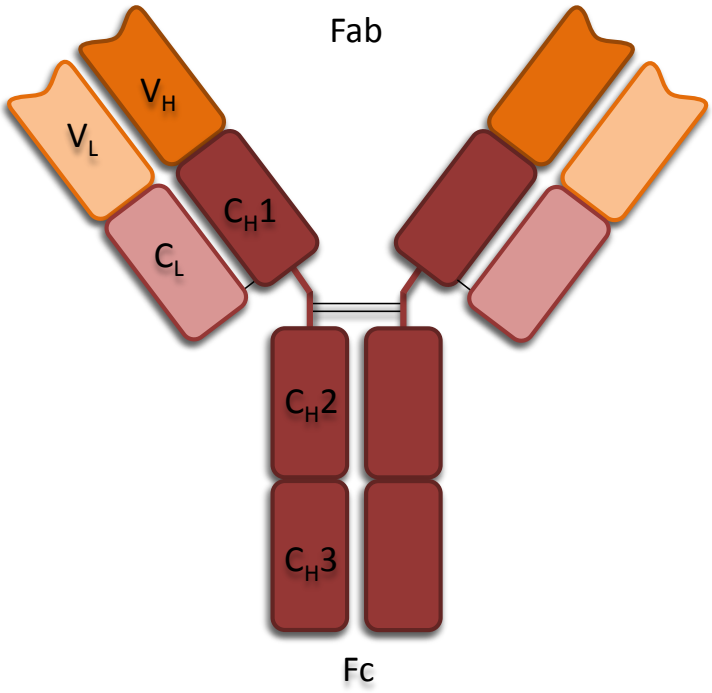


Figure 5-1 – The canonical structure of an IgG1 antibody.

The antibody structure shown in Figure 5-1 is that of an immunoglobulin γ, or IgG. There are five subclasses of antibody, determined by the type of C domains they contain. The five C domain heavy chain isotypes are α, δ, ε, μ and γ which give rise to the IgA, IgD, IgE, IgM and IgG isotypes. IgA and IgG both have further subclasses based on structural differences. IgA has three subclasses; IgA1, IgA2 and an IgA multimer consisting of two IgA molecules linked by a connective J-region (Janeway 2005). IgG contains four distinct subclasses, IgG 1-4, that differ in disulphide bond arrangement.

In addition to these heavy chain isotypes, there are two distinct light chain isotypes; κ and λ. Whilst there is no known functional difference between the two isotypes, it has been shown that the relative abundance of each chain changes between species; the relative abundance of κ:λ is 20:1 in mice compared to 2:1 in humans, suggesting a difference in expression of the two genes between species (Janeway 2005).

Each heavy chain antibody isotype has its own distinct structural properties, function, abundance and location within the body, and contributes in different ways to the immune response (Seiler, Gronski *et al*. 1985); IgA is associated with mucosal membranes and plays an important role in immunity within these membranes and

their secretions; IgD is a predominantly membrane bound receptor on lymphoid cells responsible for cellular activation; IgE is the main antibody associated with allergic responses; IgM is the first antibody circulated in an immune response, and is commonly circulated in the body; the IgG subtype is the most abundant of all Ig molecules, representing 75-80% of the antibodies in any given individual, and contains the conserved structural features of antibodies. IgG isotypes are the only Ig capable of traversing the placental barrier, allowing passive immune transfer from mother to foetus during pregnancy.

IgG, specifically the IgG1 subtype, will form the main focus of the work described here.

All antibodies are symmetrical, and consist of paired heavy and light chains, denoted H and L in Figure 5-1 (Seiler, Gronski *et al*. 1985). The heavy chain has a molecular weight of approximately 50 kDa, whilst the light chain has a molecular weight of approximately 25 kDa (Janeway 2005). The light chains flank the amino end of the heavy chain, and are connected to the heavy chain via disulphide bonds, the number of which varies in different Ig classes, and by a number of non-covalent interactions (Prabakaran, Vu *et al*. 2008). The heavy chain is divided into two regions by a linker region consisting of approximately 17 amino acids (Saphire, Stanfield *et al*. 2002), separating the heavy chain into two distinct regions. Pairs of heavy/ light chain dimers come together to form an intact heterotetramer of approximately 150 kDa; the heterotetramer is connected together by a number of disulphide bonds in the hinge region, the number of which differs based on IgG subclass, as well as a number of non-covalent interactions in the C_H3 domain (Prabakaran, Vu *et al*. 2008).

Both heavy and light chains contain a number of domains with a shared structural motif (Seiler, Gronski *et al*. 1985), known as the immunoglobulin fold. The motif is a β-sandwich structure comprised of approximately 110 amino acids, with a molecular weight of approximately 12 kDa. The light chain contains two of these domains, whilst the heavy chain contains four, or five in the case of certain Ig subclasses. Each domain contains a strong degree of sequence homology, with the exception of the most N-terminal domain in both the heavy and the light chains. The most N-terminal 110 amino acids form the variable (V) domain, responsible for antigen binding. It has been suggested that IgG molecules contain approximately

95% sequence homology, with the 5% variance located in the V-domains (Harris, Shire *et al*. 2004). IgGs also contain a intrachain disulphide bond in each domain, helping to stabilise the domain structure (Wang, Singh *et al*. 2007).

The V domain of both heavy and light chain contain three exposed loops, residues 30-36, 49-65 and 95-103 in the heavy chain (Janeway 2005) and residues 28-35, 49-59 and 92-103 in the light chain (Janeway 2005), known as the hyper variable region (HV); each region is denoted as HV1, HV2 or HV3 (based on ascending amino acid resides). These loops are also known as the complementarity-determining regions (CDRs), since they determine the antigen specificity of the Ig molecule (again, denoted CDR1, CDR2 and CDR3 based on ascending amino acid number). The CDR3 loop is always the most variable. It is the combination of CDRs 1-3 from both the heavy and light chains that provide the overall antigen binding site, rather than any single chain (Janeway 2005), for an antibody and is known as the antibody paratope. The CDR's are flanked by conserved regions known as framework regions (denoted as FR 1-4 based on position) (Janeway 2005).

The constant (C) domain of the antibody shows sequence variation, based on the different alleles from which the protein is transcribed, but no higher order structural variation.

The constant domain is, in most antibodies, post translationally glycosylated. The IgG subclass contains a single conserved glycosylation site at Asn 297 of the C_H2 domain that has been shown to be vital for the receptor binding of the antibody (Wright and Morrison 1997, Wang, Singh *et al*. 2007, Prabakaran, Vu *et al*. 2008), discussed in more detail later in this chapter. The common glycan structure for human IgG is shown in Figure 5-2.

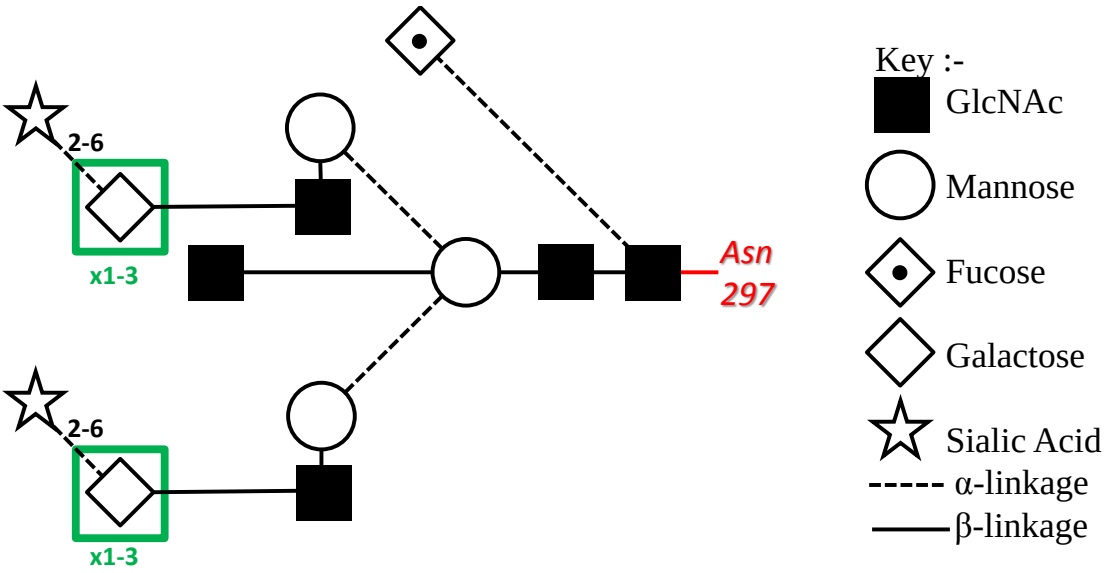


Figure 5-2 – The glycan structure of a human IgG.

All IgG's contain an N-linked core biantannary glycan consisting of three N-acetylglucosamine (GlcNAc) and three mannose moeties. To this core, additional galactose moieties can be added to either the Man (α1-6) or Man (α1-3) arms, with a possible branched fucose present at the α1-6 linkage of the primary GlcNAc moiety (Jefferis, Lund *et al*. 1990). Under some circumstances the Man (α1-3) and (α1-6) can have a terminal sialic acid in some rare cases (Jefferis, Lund *et al*. 1990).

Each C_H2 domain of the antibody contains a single glycan, and has been shown to influence the C_H2 domain spacing (Prabakaran, Vu *et al*. 2008). Removal of the glycan leads to significant changes in the Fc function, which will be discussed later.

The antibody molecule is divided into two sections; the upper arms which consist of two copies of the antigen binding domains and a constant tail domain, linked via the hinge region. The upper arms are known as the Fragment Antigen Binding, or Fab region, based on their function. The constant tail region is known as the Fragment Crystallisable, or Fc, region based on the ease by which is was crystallised; it was this region that was first crystallised by R. R. Porter in 1959 (Porter 1959).

In vivo role of IgG

Antibodies molecules have two main functions within the body; to specifically bind the pathogenic molecules responsible for the immune response, and to recruit other immunogenic molecules to assist in the removal of pathogens.

The IgG1 molecule, as shown in Figure 5-1, has evolved to fulfil both of these functions. As described in section 1.2, the antibody contains two antigen binding sites each capable of specifically binding a single, target antigen, making them bivalent and capable of cross-linking between two target antigens. The coming together of CDR 1-3 from the heavy and light chain creates a single antigen binding site which is unique in both sequence, and structure, leading to different Ig idiotypes (antibodies with a shared heavy/ light chain isotype and subclass with differences only in the variable regions). The structure recognised by the antigen binding site is known as the epitope. Epitopes fall into one of two categories, conformational/ discontinuous epitopes and linear/ continuous epitopes. Conformational/ discontinuous epitopes are those which have a specific three dimensional shape recognised by the antibody, resulting in an epitope that is composed of amino acids that are close in space and not sequence. Similarly, linear/ continuous epitopes are those that are recognised based on sequence and not shape. The majority of antigen binding sites recognise conformational epitopes; however some may be raised against linear epitopes if the target is a small, linear peptide (Janeway 2005).

Once an IgG has bound target antigen *via* the Fab region, the antibody may trigger further immune responses via interactions between the Fc region and host cell receptors. A large number of immune cells exhibit surface proteins capable of recognising antigens. Others contain an Fc receptor, an extracellular protein capable of specifically binding to a subclass of antibody, allowing the antibody molecule to provide antigen recognition system for these cells (Daëron 1997). An Fc Receptor (FcR) family exists for every subclass of antibody and are named for the antibody they bind; IgG is, therefore, recognised by FcγR proteins. Binding of IgG to FcγR triggers one of a number of immune responses, included phagocytosis, endocytosis, antibody-dependent cell mediated cytotoxicity (ADCC), the release of cell mediators of inflammation, and B-cell activation and antibody production (Raghavan and Bjorkman 1996). There are three types of FcγR, denoted FcγRI, FcγRII and FcγRIII

(Gessner, Heiken *et al*. 1998). All three FcγRs contain a transmembrane domain, an intracellular domain and a number of extracellular domains (Raghavan and Bjorkman 1996). Whilst all three FcγR's are capable of binding IgG Fc, they do so with different affinities: FcγRI has the highest affinity for IgG and is able to bind IgG at low concentrations, suggesting a role in early immune response (Shen, Guyre *et al*. 1987). FcγRII and III show a lower affinity for IgG and preferentially bind IgG when found in large, cross-linked antibody-antigen complexes (Raghavan and Bjorkman 1996).

Antibodies are incapable of eliciting a full immune response alone, and must come together in large numbers in order to trigger Fc effector function. Antibodies bind and cross-link a number of antigen molecules creating a larger antibody-antigen complex which, when bound by FcγR, induces receptor cross-linking at the cell surface and triggers cellular signalling (Raghavan and Bjorkman 1996, Daëron 1997). Despite affinity differences between FcγRI and FcγRII/FcγRIII, all FcγR's require crosslinking in order to trigger intercellular signalling (Davies and Metzger 1983).

IgG Fc is also capable of binding other receptors in addition to FcγRI, FcγRII and FcγRIII; the Fc neonatal receptor (FcRn) (Raghavan and Bjorkman 1996, Ghetie and Ward 1997, Ghetie and Ward 2000) and the C1q receptor (Cooper 1985, Ward ES and V. 1995) may also be bound.

The C1q receptor, like the FcγR, acts as a bridging receptor linking the humoral immune response with complement-dependent cytotoxicity (CDC).

Interaction between C1 proteins, the first protein complex in the complement pathway, with antibodies was first characterised in the 1960s (Müller-Eberhard and Calcott 1966), and was investigated thoroughly in the 1970s (Augener, Grey *et al*. 1971, Schumaker, Calcott *et al*. 1976). The C1q receptor is one peptide that makes up C1, a 750 kDa complex consisting of C1q, and two copies each of C1r and C1s ($C1qr_2s_2$). The binding of Fc triggers a rearrangement in the C1 protein which induces limited C1 auto-proteolysis, turning the C1 complex into an active serine protease (Ziccardi 1983). The active C1 complex can further activate additional complement proteins, C2 and C4, and begin a proteolytic cascade which eventually results in the destruction of the Ig-bound pathogen (Ziccardi 1983). This series of

activation steps is known as the classical complement pathway. CDC can be activated in other, antibody-independent ways, which are not covered here.

FcRn is another receptor capable of binding to IgG Fc. Unlike C1q and other FcγRs, FcRn is not responsible for eliciting effector function, but for antibody transport and recycling. FcRn shows specificity for the IgG class of antibody only, but unlike the FcγR it is capable of binding IgG alone, rather than IgG-antigen complexes (Raghavan and Bjorkman 1996). FcRn has been implicated in the transfer of IgG across the placental barrier from mother to foetus (Simister, Story *et al*. 1996), as well as in binding extracellular IgG in the gut, transporting it across the gut epithelium and into the blood stream (facilitating the transfer of passive immunity from mother to child via maternal milk (Mostov and Simister 1985)). FcRn has also been heavily linked to control of IgG serum lifetime (Raghavan and Bjorkman 1996). FcRn can act as a salvage receptor, binding IgG in intracellular digestive vesicles and exporting the IgG back to the cell surface for release (Ghetie and Ward 1997). The affinity of FcRn for IgG has been found to be pH dependant; at slightly acidic pH (6.0-6.5) the FcRn has higher affinity than at more neutral/ basic pH. This pH-dependent affinity is crucial to antibody reclycing, as it allows the FcRn to bind to IgG in the acidic intracellular vesicles, after which the IgG is transported to the cell surface. Upon exposure to the neutral pH of blood the FcRn loses affinity and releases the IgG back into circulation (Raghavan and Bjorkman 1996).

Exploitation of Antibodies

In 1984 Niels K. Jerne, Georges J.F. Köhler and César Milstein were awarded the Nobel Prize in Medicine and Physiology for “theories concerning the specificity in development and control of the immune system and the discovery of the principle for production of monoclonal antibodies" (Nobelprize.org 1984). For Kohler and Milstein this was awarded for their work in creating an immortalised cell line capable of producing a single antibody subtype with identical isotype and idiotype; a monoclonal antibody (mAb) (Kohler and Milstein 1975). The development of mAbs heralded a new age for antibodies in research and diagnostics since the selectivity, heterogeneity, speed of action and reproducibility enjoyed by mAbs, over polyclonal antibodies, made them far more suitable for routine experiments (Seiler, Gronski *et al*. 1985). For a long while the integration of mAbs in medicine was considered the

Holy Grail of applications, and a large amount of research was directed at addressing the inherent problems associated with using mAbs in medicine, discussed in section 4.1.1.5

Monoclonal Antibodies as therapeutics; history, current status and future perspectives

The exploitation of biological agents for medicinal purposes dates back as far as 1796, when Edward Jenner first used cowpox to immunize individuals against smallpox (Riedel 2005). Since these humble beginnings the use of biological agents in therapeutics has risen dramatically. Described as "magic bullets" in 1975, monoclonal antibodies were due to herald a new age in the use of biologics as therapeutic agents (Kohler and Milstein 1975). With their high selectivity, and ability to exploit a patient's immune response for therapeutic benefit, mAbs seemed to be the perfect drug for a number of indications. The first mAbs produced in the 1980s were murine based, and presented a number of challenges. Primarily, murine mAbs could be recognised as non-self by the patient, leading to the production of anti-drug antibodies (ADAs)(Buss 2012). Murine mAbs also exhibited short serum half-lives, as a result of weak binding to FcRn (Ober, Radu *et al*. 2001), and were poor recruiters of effector function, including complement activations and ADCC (Stern and Herrmann 2005). It has subsequently been shown that mAbs from different species can have distinct glycosylation patterns (Sheeley, Merrill *et al*. 1997), which can be a key factor affecting in Fc receptor binding.

In an attempt to address these issues a number of chimeric mAbs were created (Vaughan, Osbourn *et al*. 1998), containing ~65% human sequence in the constant domains, whilst the variable domains remained murine (Morrison, Johnson *et al*. 1984). These chimeras showed increased recruitment of effector function, complement activation and ADCC activity, alongside improved serum half-lives, but still led to the production of ADA's in >50% of patients (Vaughan, Osbourn *et al*. 1998). A limited number of chimeric mAbs however were produced that successfully progressed beyond clinical trials (Vaughan, Osbourn *et al*. 1998)

Chimeric mAbs are still capable of producing ADA's due to the conserved nature of the V-domain. With the exception of CDR regions on VH and VL both chains are

homogeneous within an organism, and it was with this realisation that work began on creating humanised mAbs (Vaughan, Osbourn *et al*. 1998); mAbs containing the human sequence in all but the CDR regions, which would be raised in murine systems and then grafted onto the human mAb gene (Jones, Dear *et al*. 1986, Riechmann, Clark *et al*. 1988, Verhoeyen, Milstein *et al*. 1988). Humanising mAbs overcame many of the issues associated with ADA's and the poor half-life/ induction of effector functions but it was a laborious process.

Current generation mAbs are no longer humanised, but fully human, making use of transgenic mice capable of expressing human variable domains (Green, Hardy *et al*. 1994, Lonberg, Taylor *et al*. 1994) and phage display technology to rapidly screen, select and optimise for V domains with enhanced antigen affinity (Winter, Griffiths *et al*. 1994). As with humanised mAbs, these show expected levels of serum half-life and effector function activity, with a diminished level of ADA production (Buss 2012).

In addition to producing human/ humanised mAbs, focus in recent years has shifted towards optimising these mAbs to make them more suitable for purpose. The Fab regions of the antibody contain the V domains responsible for antigen binding, are normally left unmodified and are normally engineered for enhanced antigen affinity and binding kinetics. The Fc domain is responsible for elucidating effector function and controlling antibody half-life, and a large degree of research has been carried out in order to investigate and modulate interactions with Fc receptors.

A number of groups have studied the Fc domain in conjugation with its receptors (FcRn, FcγRI-III and C1q) in an attempt to map the residues important in receptor binding on the Fc (Idusogie, Presta *et al*. 2000, Shields, Namenuk *et al*. 2001), most of which were localised to the upper C_H2 domain. Research has also been undertaken into how glycosylation contributes to Fc function (Krapp, Mimura *et al*. 2003, Prabakaran, Vu *et al*. 2008, Abès and Teillaud 2010). It has been observed that FcγR binding is highly sensitive to glysocylation at residue Asn 297, and that the glycan structure can lead to changes in the binding affinity (Sibéril, de Romeuf *et al*. 2006, Takahashi, Kuroki *et al*. 2009, Abès and Teillaud 2010). The use enzymes which enrich for certain types of glycans (such as N-acetylglucosamyltransferase III) (Davies, Jiang *et al*. 2001), as well as cell lines engineered for certain glycan

structures (Yamane-Ohnuki, Kinoshita *et al*. 2004) have been investigated in an attempt to optimise glycosylation to favour specific functions.

Fc receptor engineering may follow one of two strategies – increasing receptor binding strength to enhance binding and thus increase activity, or destabilisation/ removal of binding affinity and thus eliminating activity. In all cases, strategic mutation of various residues in and around the receptor binding sites leads to changes in function.

There are a number of documented examples of IgGs with enhanced receptor binding, and thus enhanced effector function, for ADCC, antibody-dependent phagocytosis (ADP) (Shields, Namenuk *et al*. 2001, Lazar, Dang *et al*. 2006, Stavenhagen, Gorlatov *et al*. 2007, Horton, Bernett *et al*. 2008, Richards, Karki *et al*. 2008, Zalevsky, Leung *et al*. 2009) and antibody-dependent CDC (Idusogie, Presta *et al*. 2000).

The principle behind enhancing mAb-dependant effector function seems obvious, but there are many instances where silencing these signals is preferable. For example, targeting cell surface proteins on immune cells, which would result in ADCC, ADCP and/ or CDC, may pose a serious safety risk to the patient and it is therefore preferable to silence these functions (Strohl 2009). Solaris® (eculizumab) is one example of an effector silenced antibody designed to block the complement cascade pathway without inducing further immune response (Rother, Rollins *et al*. 2007). A number of Fc engineered variants exist which silence Fc mediated activation of effector function (Bolt, Routledge *et al*. 1993, Alegre, Peterson *et al*. 1994, Hutchins, Kull *et al*. 1995, Cole, Stellrecht *et al*. 1999, Reddy, Kinney *et al*. 2000, Xu, Alegre *et al*. 2000, Oganesyan, Gao *et al*. 2008, Sazinsky, Ott *et al*. 2008, Strohl 2010, Bell and Rother 2012).

In addition Fc variants engineered to modulate effector function, there are also a number of variants designed to have increased binding affinity to FcRn, and thus increased serum half-life (Acqua, Woods *et al*. 2002, Hinton, Xiong *et al*. 2006, Oganesyan, Damschroder *et al*. 2009, Yeung, Leabman *et al*. 2009, Ober 2012) . Increasing serum half-life can increase drug exposure and reduce patient dosing frequency thereby enhancing a therapeutic mAbs overall effectiveness.

It has been nearly 40 years since Kohler and Milstein first developed the technology to create mAbs. In the last 20 years particularly, antibody therapeutics have become the dominant force in global therapeutics. In 2012 five of the top 10 prescription drugs were biologics, four of which were monoclonal antibodies. The combined sales of these four mAbs accounted for $29.3 billion USD (Lindsley 2013). Antibodies are currently used to treat a range of disorders including, autoimmunity and inflammation, infectious diseases, metabolic disorder and cancer (Larrick 2012). With hundreds of potential mAb based therapeutics in development, as well as a number of mAbs reaching the end of their patent exclusivity, leading to the development of mAb biosimilars, it is more important than ever to have a number of robust analytical tools available for their characterisation.

5.1.2 Characterising Biotherapeutics – Mass Spectrometry and Antibodies

Characterising any recombinant biotherapeutic is non-trivial. They are often chemically modified with small modifications, such as deamidation (Liu, van Enk *et al*. 2009), or post transnationally modified with larger modifications, such as glycosylation. These modifications, particularly in the case of glycosylation, can contain an inherent heterogeneity, making structure/ function relationships more difficult to characterise. In addition to variable modifications, proteins have the potential to misfold/ rearrange during expression or in storage which can lead to conformational differences. There are therefore a number of different aspects to consider when attempting to fully characterise these complex molecules (Chen 2011).

Mass spectrometry has been used to characterise a number of different structural features of mAbs, including post-translational modifications (glycosylation) (Lewis, Guzzetta *et al*. 1994, Ashton, Beddell *et al*. 1995, Wang, Amphlett *et al*. 2005, Reid, Tait *et al*. 2010), chemical modifications (deamidation and N-terminal glutamine cyclisation) (Yu, Remmele *et al*. 2006, Zhang, Pan *et al*. 2009, Reid, Tait *et al*. 2010), primary sequence (Downard 2000, Wang, Amphlett *et al*. 2005, Zhang and Shah 2007, Zhang, Pan *et al*. 2009), disulphide bond arrangement and positions (Zhang, Pan *et al*. 2009), stability (Liu, Chumsae *et al*. 2010), aggregation and self-assembly (Kukrer, Filipe *et al*. 2010), antibody/ antigen complex formation and stoichiometry (Tito, Miller *et al*. 2001).

There is a growing need for methodologies to characterise the higher order structure of mAb therapeutics to better understand key structure/ function relationships (Zhang, Cui *et al*. 2014). Monoclonal antibodies are difficult to study by NMR difficult due to their size, and the flexible linker region makes obtaining a crystal for X-ray diffraction difficult (Saphire, Stanfield *et al*. 2002). Recent advances in MS technology have enabled the technique to contribute to the study of higher order structure.

Native MS has been used on a modified, high resolution Orbitrap^{TM} mass spectrometer to analyse a mAb intact (Rosati, Rose *et al*. 2012, Thompson, Rosati *et al*. 2014). The resolution is sufficient to identify the heterogeneity, if the sample is prepared correctly, from different Fc glycoforms. The mass accuracy is also sufficient to observe C-terminal lysine cleavage, or primary sequence mutations. Analysis in this way also provides direct information regarding a proteins conformation since charge state distributions correlate with changes in protein structure (Katta and Chait 1991).

Ion mobility mass spectrometry (Section 1.1.6 Ion Mobility Mass Spectrometry) can be used to probe higher order protein structure, and has been successfully used to differentiate between structural isoforms in IgG2 (Bagal, Valliere-Douglass *et al*. 2010) and to investigate the dynamics of intact IgG and IgG fragments (Pacholarz, Porrini *et al*. 2014).

Hydrogen deuterium exchange mass spectrometry has also been used to characterise protein structure, proving an effective tool in highlighting disruptions in local protein structure between different C_H3 mutants (Rose, van Berkel *et al*. 2013)

5.1.3 A study of Fc engineered antibodies

Fc engineering is becoming increasingly important in the field of biotherapeutics. Mass spectrometry is uniquely able to characterise the structural characteristics of these large, heterogeneous and dynamics biomolecules,

Molecules to study

Three IgG Fc's were used in this study; two triple mutants known as a the triple mutant (TM) construct (Oganesyan, Gao *et al*. 2008) and the YTE construct (Acqua, Woods *et al*. 2002, Oganesyan, Damschroder *et al*. 2009). The third Fc used was the

TM-YTE construct (Figure 5-3) containing all six mutations present in the TM and YTE constructs.

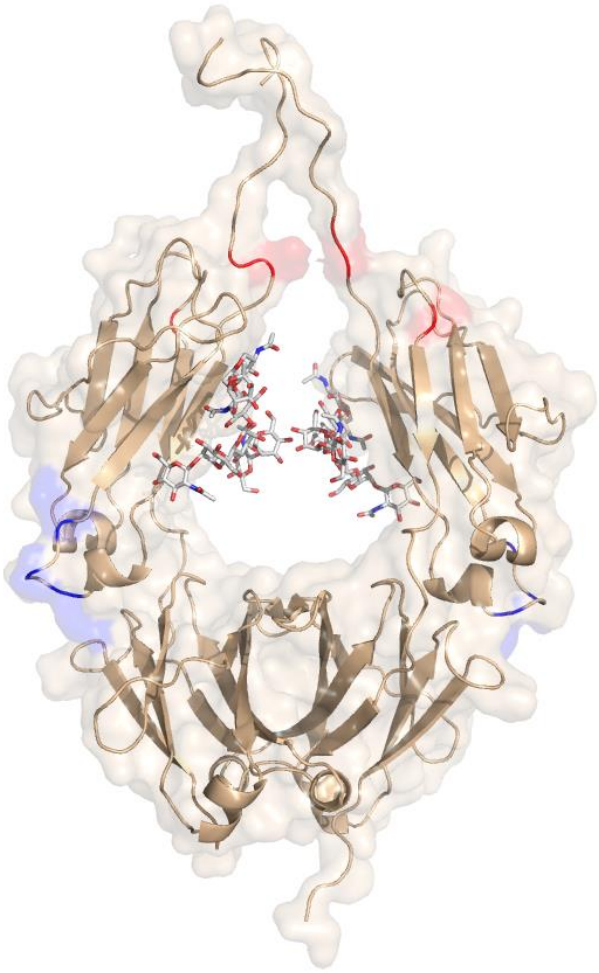


Figure 5-3 – A homology model of the IgG1 Fc domain complete with hinge.
Residues highlighted in RED are those corresponding to the TM construct. Residues highlighted in BLUE are those corresponding to the YTE construct.

The TM construct contains three mutations located in the upper C_H2 domain (L234F, L235E, P331S) (shown in Figure 5-3) which lead to a decrease in ADCC, ADCP and antibody dependent-CDC. Residues 234 and 235 had been previously identified as having an important role in Fcγ receptor binding (Canfield and Morrison 1991, Chappel, Isenman *et al*. 1991, Armour, van de Winkel *et al*. 2003) whilst residue 331 has been shown to play an important role in C1q binding (Tao, Smith *et al*. 1993, Idusogie, Presta *et al*. 2000). Mutagenesis of these three residues resulted in a reduced binding affinity to a number of FcγR's and C1q. It should be noted,

however, that there was no electron density for residues 234 and 235 since these both lie in the lower hinge.

The YTE construct contains three mutations located in the lower C_H2 domain (M252Y, S254T, T256E) (Figure 5-3) which result in a 10-fold increase in the binding of Fc to FcRn at pH 6.0, resulting in a 4-fold increase in the serum half-life on mAbs engineered with these mutations (Acqua, Woods *et al*. 2002, Dall'Acqua, Cook *et al*. 2006).

Analysis of the Fc-TM and Fc-YTE crystal structures suggests some structural differences between the two, shown in Figure 5-4.

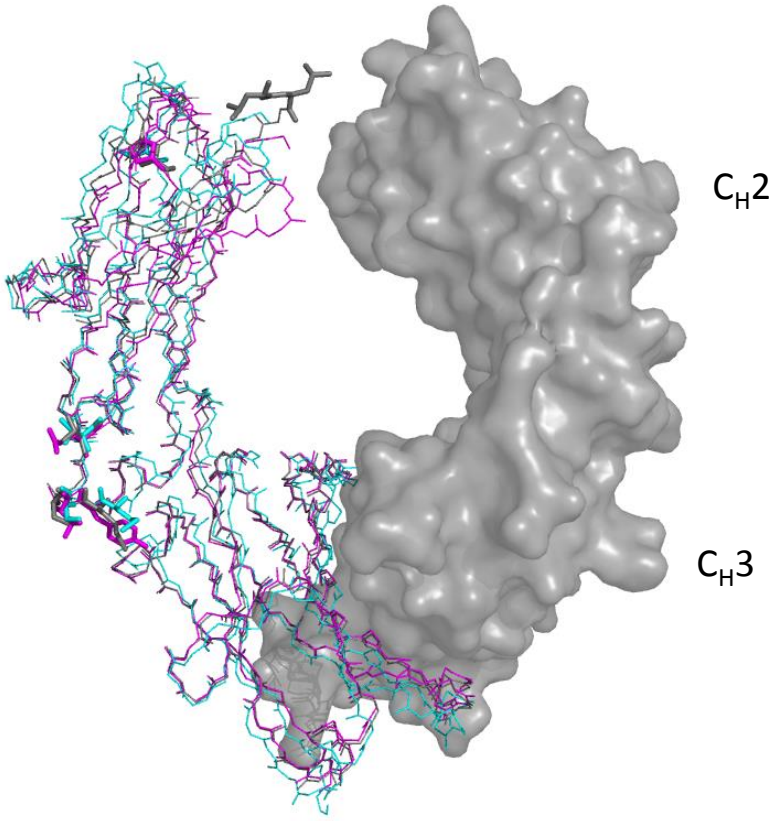


Figure 5-4 – A comparison of the X-ray crystal structures of the Fc-TM (PDB: 3C2S), Fc-YTE (PDB: 3FJT) and Fc-WT (PDB: 3AVE).
Fc-TM is shown in cyan, Fc-YTE is shown in magenta and Fc-WT is shown in Grey. Glycans have been omitted from this structure, but were present on the full crystal. Figure generated using structural alignment in Pymol

A comparison of the three crystal structures suggests that Fc-TM has a larger C_H2 interdomain space than that of Fc-YTE or Fc-WT. This may be related to function, or may be an artefact of the crystal structure.

Despite this subtle difference, a closer inspection of the sites of amino acid substitution in the three constructs suggests that these residues adopt a similar conformation between the different constructs (Figure 5-5 and Figure 5-6) indicating no significant changes to protein conformation in either construct at this site.

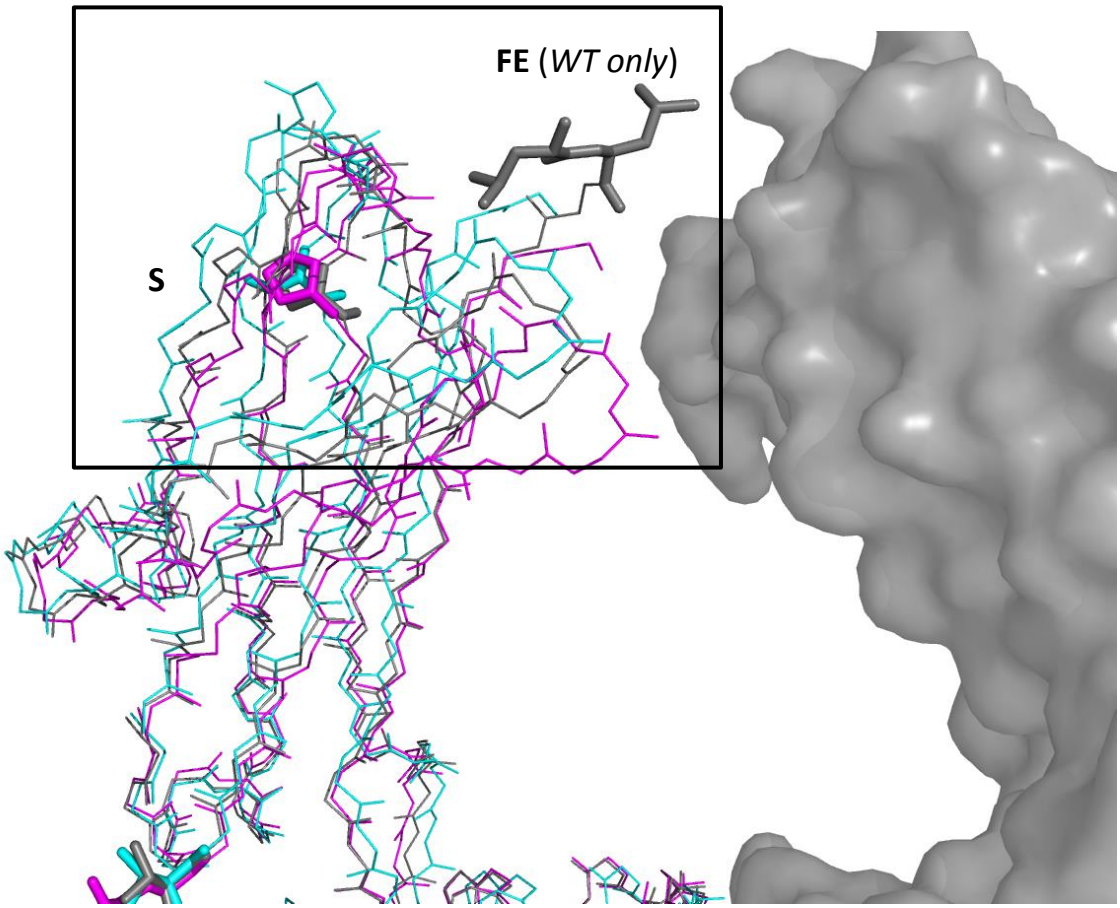


Figure 5-5 – An enhanced image of the TM mutation sites in the upper C_H2 domain shown in Figure 5-4.
Residues 234 (L → F) and 235 (L → E) only showed electron density in the WT structure, and are located at the top of the C_H2 domain. Residue 331 (P → S) is highlighted in the figure. All three amino acids are predicted to overlay, suggesting no conformational change. Fc-TM is shown in cyan, Fc-YTE is shown in magenta and Fc-WT is shown in Grey. Figure generated using structural alignment in Pymol

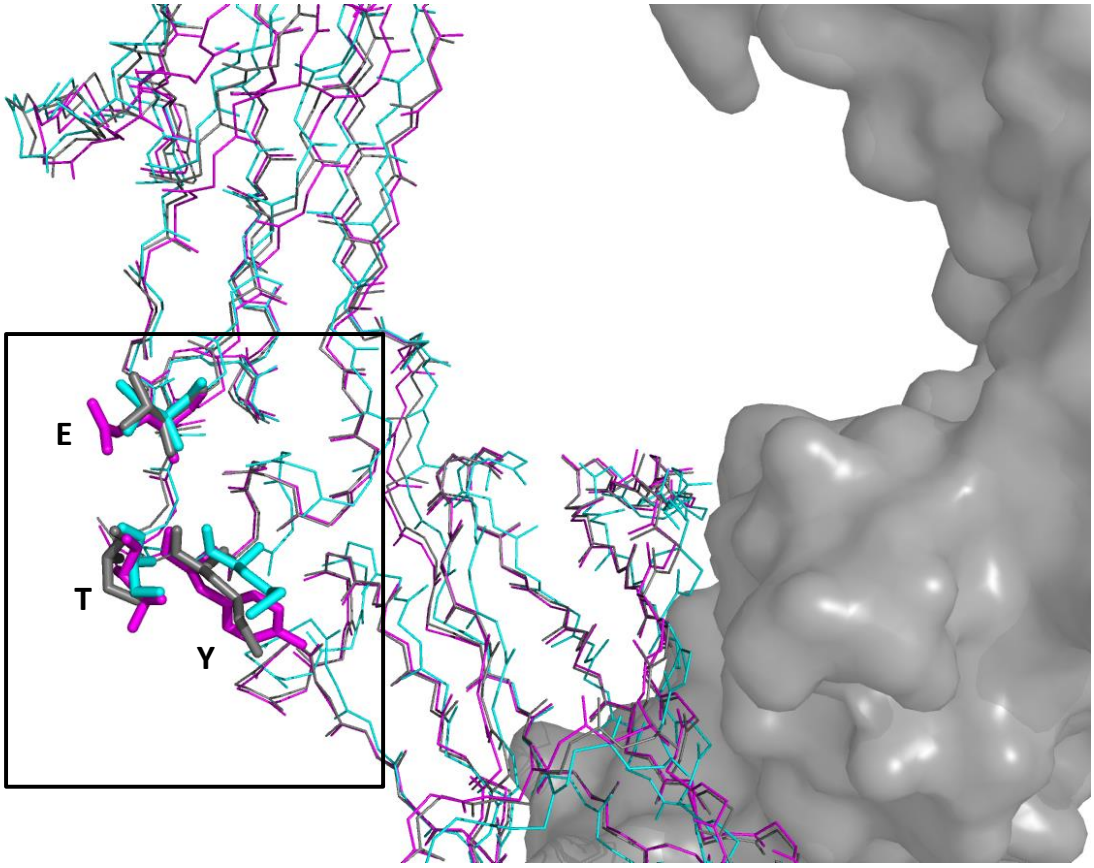


Figure 5-6 - An enhanced image of the YTE mutation sites in the lower C_H2 domain shown in Figure 5-4.
Residues 252 (M→Y), 254 (S→T), and 256 (T→E are located at the bottom of the C_H2 domain. All three amino acids overlay, suggesting no conformational change. Fc-TM is shown in cyan, Fc-YTE is shown in magenta and Fc-WT is shown in Grey. Figure generated using structural alignment in Pymol

When an Fc variant was generated containing both TM and YTE modifications (a TM-YTE construct containing all six substitutions) there was an observed shift in the thermal stability of the C_H2 domain observed by DSC. A summary of the data is shown in Figure 5-7.

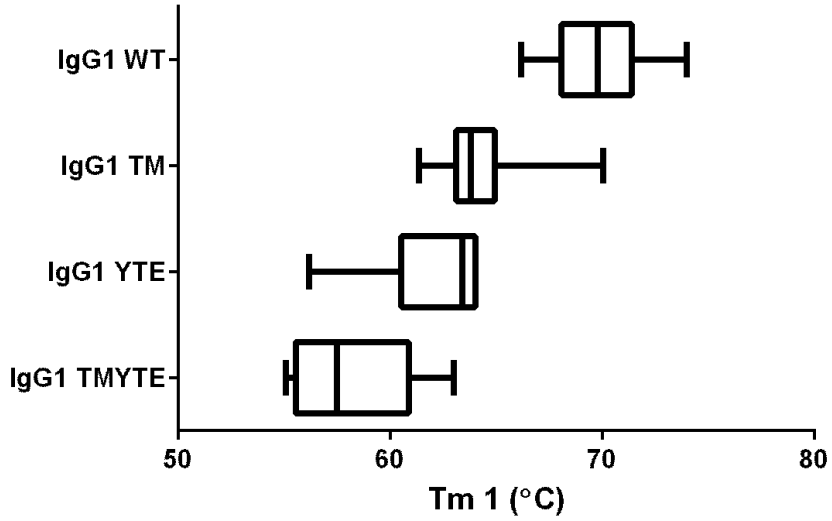


Figure 5-7 – A box and whisker plot illustrating the differences in the thermal stability of the C_H2 domain for Wild Type (WT) TM, YTE and TMYTE constructs.
Data is based on the average Tm1 shifts (known to refer to the melting of the C_H2 domain) of 44 fully human IgG1 mAbs – 12 IgG-WT, 21 IgG-TM, 6 IgG-YTE and 5 IgG-TMYTE. Box shows the first to third quartile, bisected by the mean average value. Whiskers show the upper and lower limits.

Both IgG-TM and IgG-YTE show a similar reduction in thermal stability of approximately 6-7^{o}C in the C_H2 domain upon mutation; Tm1 changed from 70.1^{o}C in WT to 64.1^{o}C in IgG-TM and 62.2^{o}C IgG-YTE constructs, based on the mean Tm from sampled data.. This destabilisation is compounded in IgG-TMYTE, which shows a C_H2 Tm of approximately 58^{o}C (12^{o}C change) based on the mean Tm from sampled data. The effects of the two sets of mutations are not purely additive, suggesting a cooperativity between the sets of two mutations. As previously described, there is no direct structural information underpinning these changes in thermal stability between the individual TM and YTE constructs compared to the wild type.

Experimental Aims

This study aims to use a combination of recently introduced mass spectrometry based techniques to characterise the structural differences between different IgG constructs engineered for different effector functions, and to attempt to rationalise the observed thermal destabilisation in IgG-TMYTE. The hypothesis is that there will be observable structural differences between the wild type and the mutants relating to the thermal stability differences observed in Figure 5-7.

It was the aim of this study to use ion mobility mass spectrometry to identify the conformation of the Fc-WT, and to compare this with the conformations of the three engineered Fc variants to provide insight into conformational changes across the domain. It was also the aim of the study to use hydrogen/ deuterium exchange MS to investigate the conformational changes between the Fc WT and the different engineered variants, and to use this as a measure of local conformational change within the Fc domain.

5.2 Results and Discussions

5.2.1 Preparation of the Fc

Enzyme digestion strategy

Antibodies may be digested into separate Fab/ Fc Fragments using a number of different strategies, depending on the desired outcome. It was decided that analysis of the isolated Fc domain would be the most appropriate for two reasons. Firstly this would ensure that there was no variation in HDX kinetics as a result of differences in the V-regions, since one of the four IgG1s sourced contained a different V-region, based on sample availability. The second reason was specific to the ion mobility experiments. Given the inherent flexibility of IgG1, by virtue of the hinge region, it was believed that this flexibility would mask any Fc structural differences caused by the mutations which could be observed by IMS. To have the best chance of observing conformational differences, only the Fc's were used. A number of digestion strategies were considered for obtaining isolated Fc regions.

Pepsin has long been considered as a key proteolytic enzyme for obtaining antibody fragments. Pepsin cleaves on the carboxy-terminal side of the hinge disulphide bridge, generating a number of hinge and upper C_H2 fragments along with the disulphide linked Fabs, known as the $F(ab')_2$ fragment. The Fc is released, but also undergoes a number of additional cleavage steps in the C_H2 region, generating a partial Fc (pFc') fragment (Figure 5-8) (Janeway 2005)

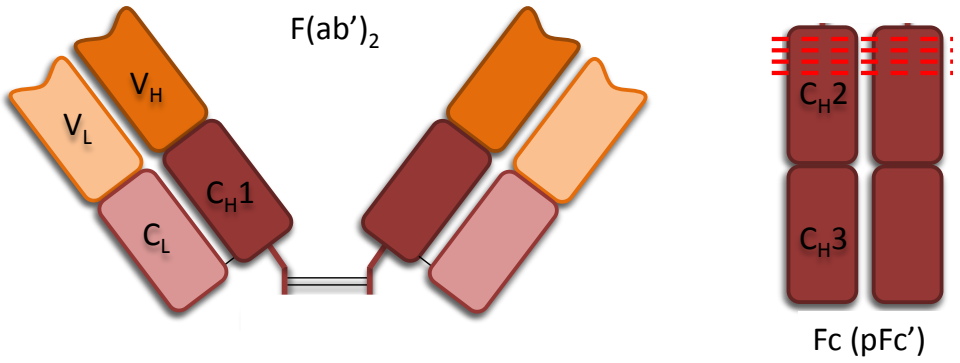


Figure 5-8 – The fragments generated by Pepsin partial digestion; $F(ab')_2$ and pFc'. Dashed red lines on the Fc construct show additional sites of digestion.

Pepsin was therefore ruled out since it cleaved at the site of TM mutation (upper C_H2) and would therefore provide an incomplete Fc structure.

A novel protease, Immunglobulin-degrading enzyme S (IdeS), was also ruled out for the same reason. IdeS (von Pawel-Rammingen, Johansson *et al*. 2002) is a protease which cleaves a specific site within the hinge on IgG's on the carboxy-terminal side of the disulphide bridge. Unlike pepsin, IdeS shows specificity for the target sequence shown in Figure 5-9.

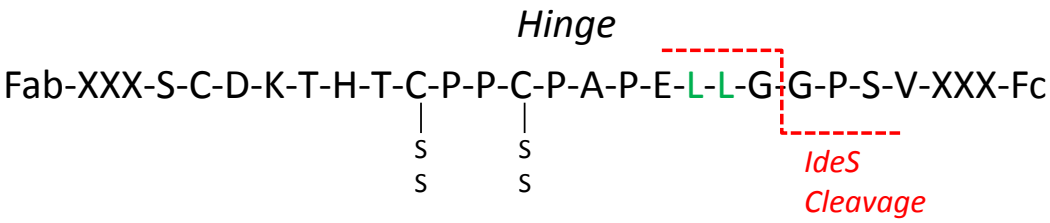


Figure 5-9 - A focused image of the IgG1 hinge sequence, illustrating the IdeS cleavage site.
IdeS cleave on the C-terminal side of L234 and L235, two residues mutated in the Fc-TM construct.

Whilst IdeS does not lead to ragged digestion, like pepsin, it still cleaves below the disulphide, yielding an incomplete Fc. IdeS digestion would also remove two of the three residues substituted in the TM construct.

Papain has been considered as being an enzyme of choice for proteolytic cleavage of IgG into Fab and Fc fragments. Papain cleaves at the amino-terminal side of the disulphide bridge in the hinge, generating a complete Fc fragment, complete with disulphide bonds, and Fab fragments. Papain does have however a number of undesirable characteristics associated with digestion. Due to its nonspecific nature,

papain digestion can result in microheterogeneity within the Fc fragments as a result of differences in digestion site in the hinge, making fragment identification difficult. In addition, papain is a thiol protease and requires the presence of reducing agents in order to function. It has been observed that, if the conditions are not completely optimised, the presence of cysteine can lead to cleavage of the interchain disulphide bonds (Adamczyk, Gebler *et al*. 2000). Termination of papain digestion occurs via addition of iodoacetamide, an alkylation agent. It has been observed in other experiments performed that this can lead to derivitisation of the partially cleaved disulphide bridges, which would cause problems to assessing structure/ function performance of the antibody Fcs.

The enzyme chosen for these experiments was endoproteinase Lys C. Like IdeS, endoproteinase Lys C has a specific digestion site; it cleaves on the C-terminal side of lysine (Jekel, Weijer *et al*. 1983). Under carefully controlled conditions this can be used for the partial digestion of antibodies, this leads to the digestion of a single position within the mAb hinge (Figure 5-10).

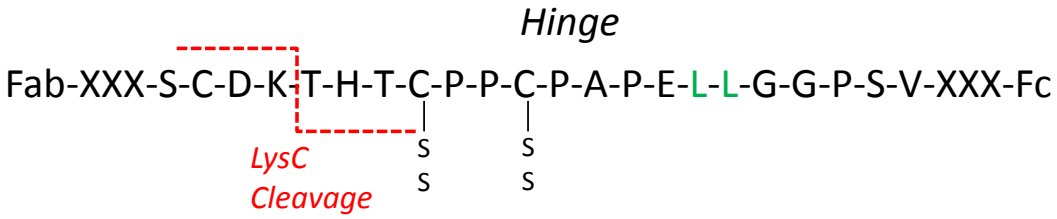


Figure 5-10 –Tthe digestion site of endoproetinase lys C.
Like IdeS, endoproteinase Lys C has a specific cleavage site. Unlike IdeS, endoproteinase lys C cleave on the N-terminal side of the disulphide bridge, retaining the intact Fc and all the TM mutations.

As shown in Figure 5-10, endoproteinase lys C appears to be a perfect enzyme for the digestion of an IgG to obtain a homogeneous Fc fragment that retains the full structure of the Fc. The endoproteinase lys C reaction is terminated by the addition of diisopropyl fluorophosphate (DIPF); a specific inhibitor which binds covalently to the catalytic serine and prevents further digestion. As a protease inhibitor, DIPF should ensure there is no derivitisation of the antibody during the digestion.

Protein Purification

The in-house method for purification of IgG's and Fc fragments involved large scale use of protein A (Duhamel, Schur *et al*. 1979). Protein A has, for many years, been

used as the affinity matrix of choice for purification of antibodies (Hober, Nord *et al*. 2007). The facilities at MedImmune are equipped for a larger scale (10's of millilitres to 10's of lires scale) the antibody Fcs, and so steps were taken to develop a smaller, analytical scale purification method more suitable to a few millilitres of sample.

For the affinity matrix, CaptureSelect^{TM} affinity columns were purchased from Genovis and used to purify the Fc fragments (see methods). At the time of analysis, the YTE had yet to be sourced, and only experimental data for the WT, TM and TMYTE samples is presented here. In order to test the applicability of the CaptureSelect^{TM} affinity matrix the three samples were bound and eluted from the column following the method outlined above. The acid-released eluent was subject to LC-MS analysis, the chromatograms for which are shown in Figure 5-11.

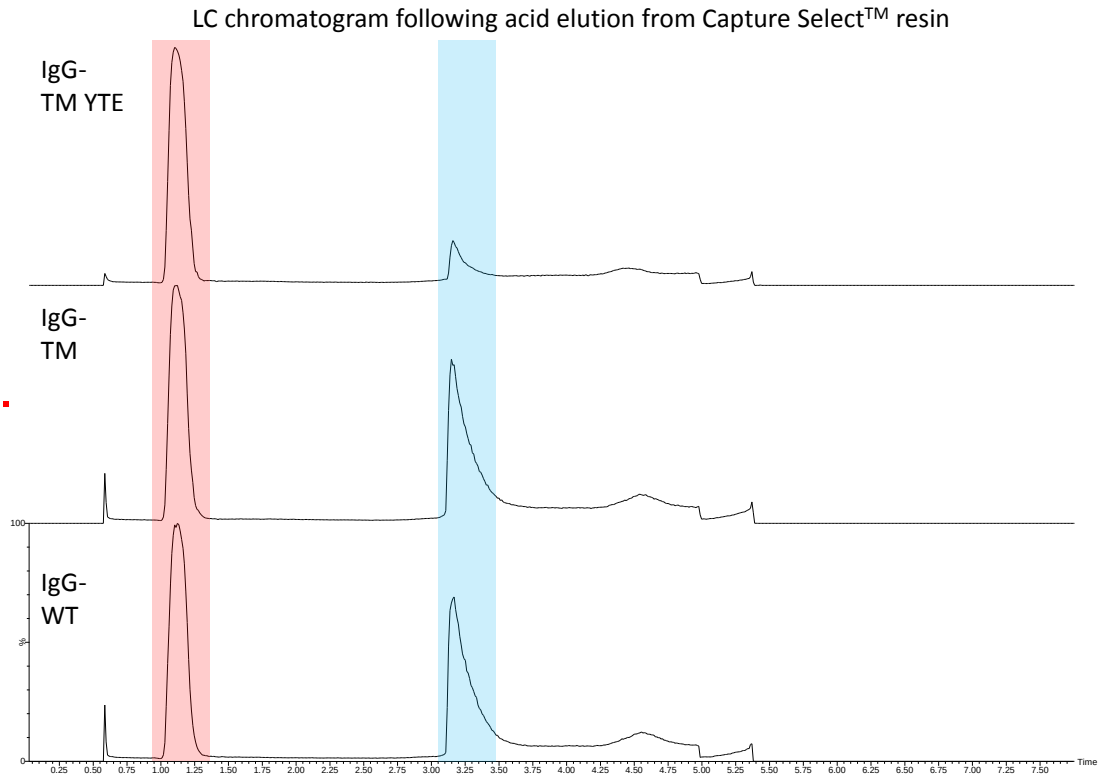


Figure 5-11 – Chromatogram showing the elution's from the CaptureSelect column following acid release.
Two major peaks can be observed; the first (red) corresponds with low signal noise as a result of solvent/ salt elutions. The second (blue) corresponds to the full length IgG

All three samples show a single large peak eluting between 1 and 1.25 mins, which corresponds to salts and other low m/z species. Both the WT and the TM mutant show a prominent peak eluting between 3 and 3.5 minutes. The 3-3.5min peak is

present in the TMYTE sample but at much lower intensity. The deconvoluted MS spectra associated with the 3-3.5 min peak are shown in Figure 5-12.

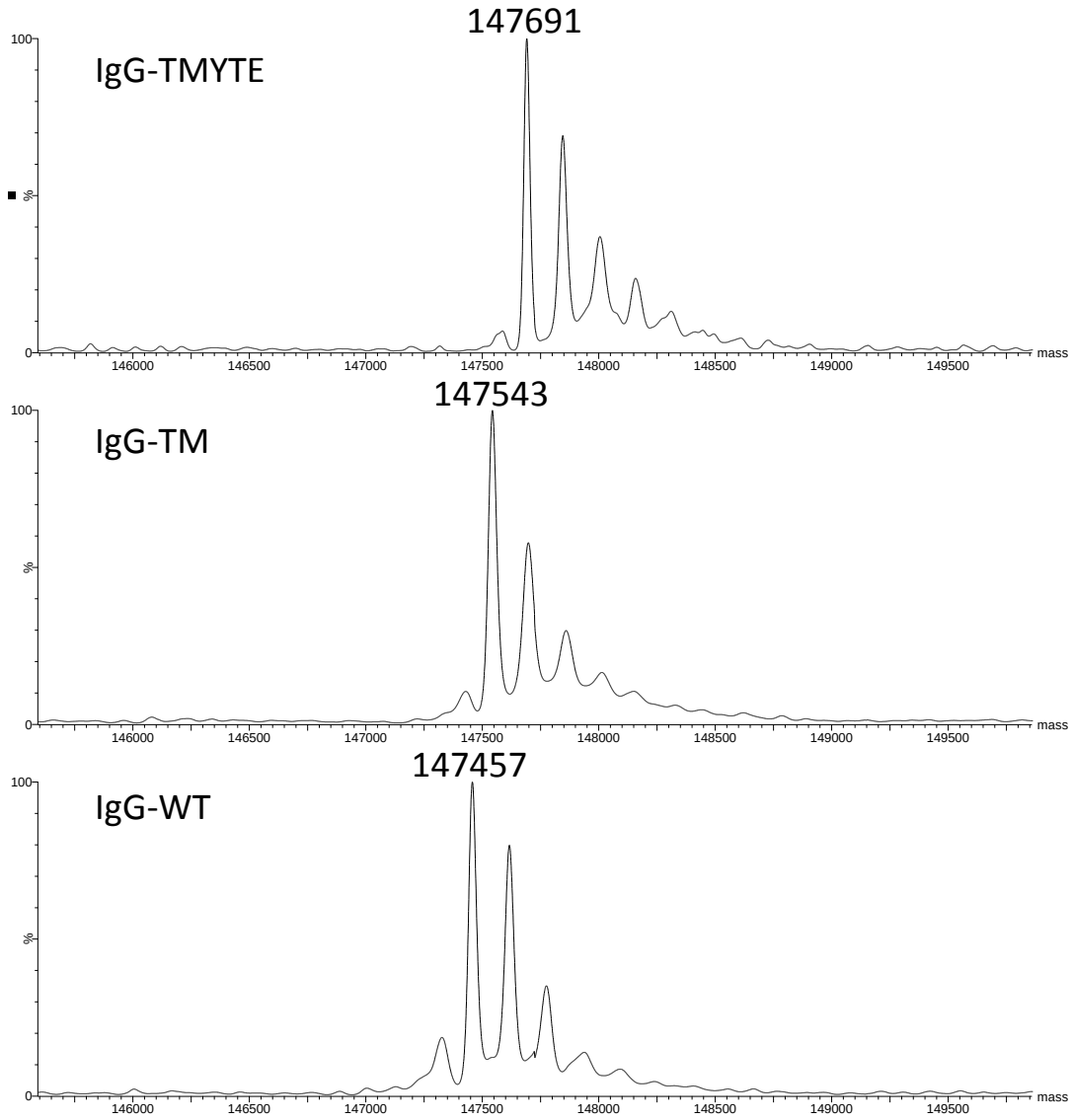


Figure 5-12 - The deconvoluted MS spectra of the IgG-WT, IgG-TM and IgG-TMYTE eluted from the blue peak shown in Figure 4-11.

The three peaks deconvolute to give masses of approximately 147 kDa, plus additional glycan species (see Figure 5-15 and Figure 5-17), corresponding to the intact antibody. Based on observations shown in Figure 5-11, this would suggest that the IgG-TMYTE shows a different binding affinity to the CaptureSelect columns than either the IgG-WT or the IgG-TM. Either the pH 2.8 glycine wash is too weak to successfully remove the IgG from the CaptureSelect^{TM} or the CaptureSelect^{TM} did not bind the IgG. In order to test this hypothesis, the three IgG's were bound to the

resin and washed with PBS. The eluent was then analysed using LC-MS; results are shown in Figure 5-13.

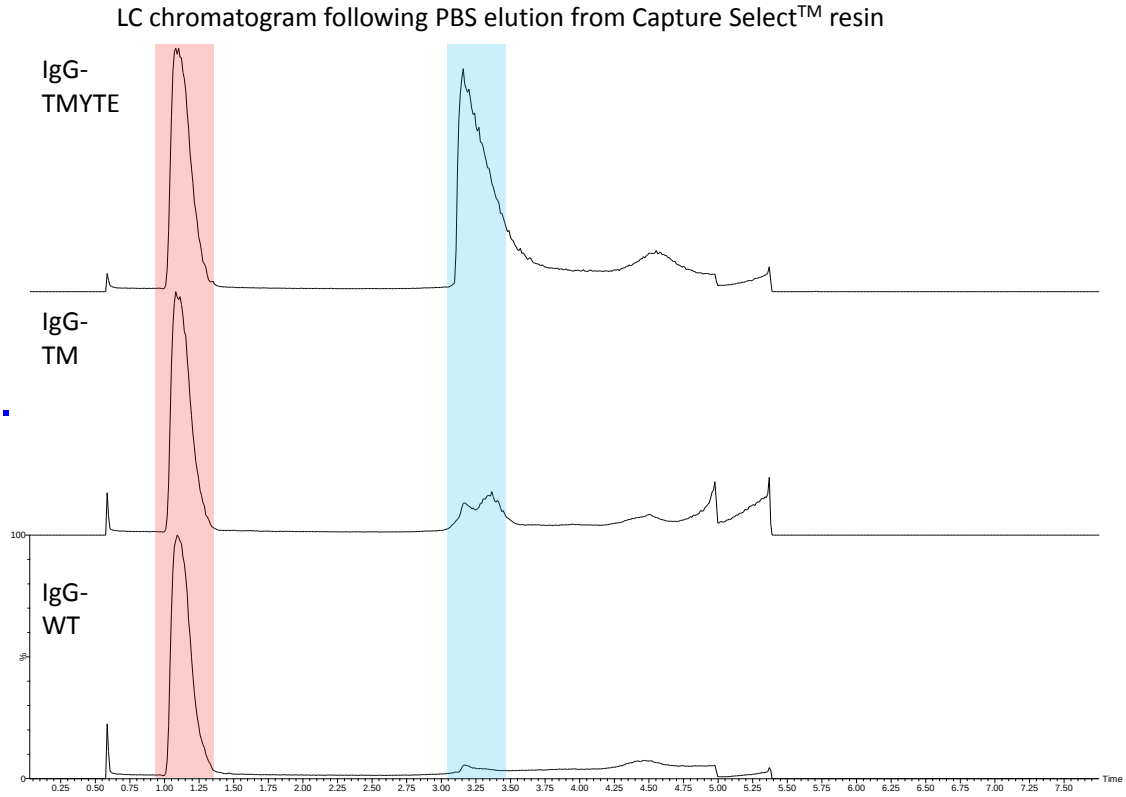


Figure 5-13 – A chromatogram showing the elution profiles of IgG-WT, IgG-TM and IgG-TMYTE after binding to CaptureSelectTM affinity matrix and washed with PBS (10mM sodium phosphate, 150mM NaCl, pH 7.4)

Figure 5-13 shows that, with a PBS wash, IgG-WT and IgG-TM have minor elution peaks at 3-3.5 mins, corresponding to the IgG. The IgG-TMYTE, by comparison, has a much larger elution peak, supporting the suggestion that it is unable to bind to the CaptureSelect resin™. The small peaks present in the PBS wash for the IgG-WT and IgG-TM are believed to be as a result of overloading the column, rather than the induced dissociate of the protein from the resin.

This data suggests that CaptureSelect™ affinity matrix is incompatible with the TM-YTE mutation. The matrix itself consists of an immobilised V domain from an antibody raised against human IgG Fc. Whilst the CaptureSelect™ matrix does have broad cross-species applicability, including human IgG1, the data here suggests that as a result of the TMYTE mutations the matrix no longer has affinity with the IgG-Fc. There are two possible explanations for this. Engineering of the double mutant may reduce the affinity of the CaptureSelect™ for the antibody, making the

observation in Figure 5-13 specific to the Fc-TMYTE and thus suggesting a conformational change in the Fc-TMYTE, or the YTE mutation itself changes the epitope of the CaptureSelect^{TM} matrix.

Following the unsuccessful attempt to use the low-volume CaptureSelect^{TM} affinity matrix, IgG's were digested and subjected to purification using the MabSelect SuRe protein A affinity matrix. Full IgG's were not screened using this resin prior to digestion as they were with the CaptureSelect^{TM} matrix since all antibodies provided had been previously purified using the MabSelect SuRe protein A affinity matrix during the production process.

Following elution, the Fc-WT, Fc-TM, Fc-YTE and Fc-TMYTE fractions were analysed using LC-MS to confirm identity; the chromatograms are shown in Figure 5-14.

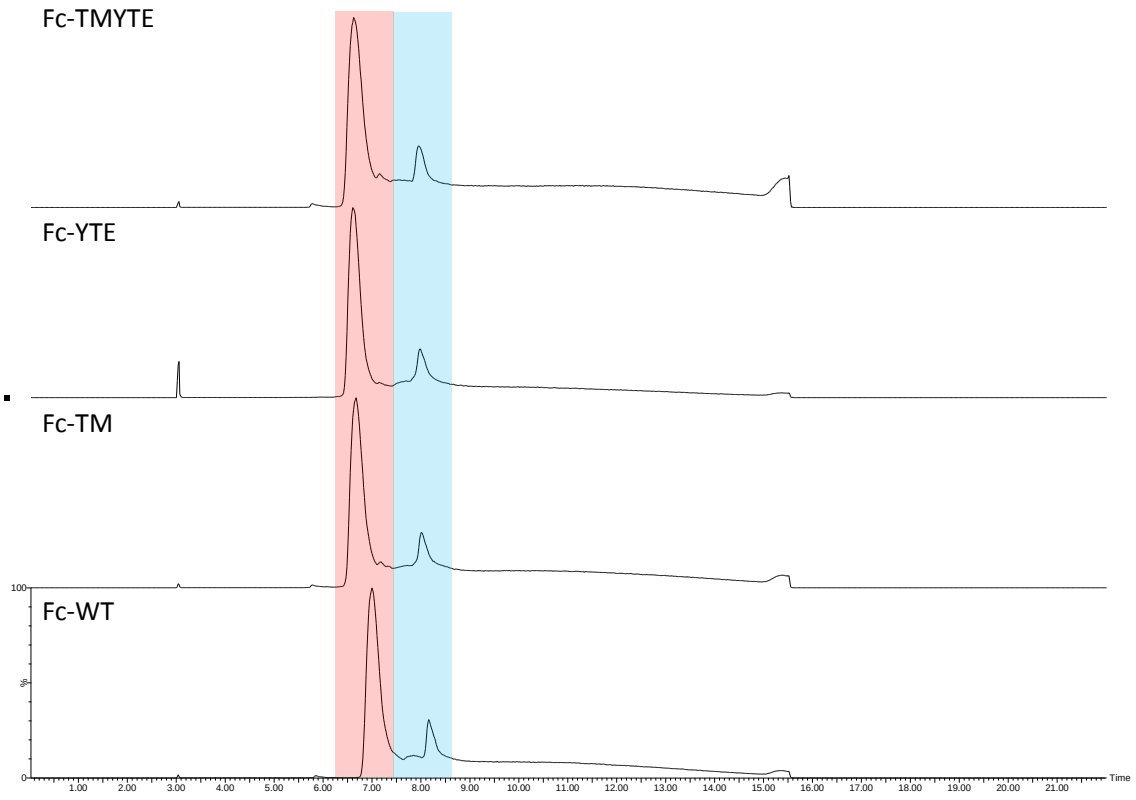


Figure 5-14 – LC-MS chromatogram for the four IgG-Fc samples. All four indicate two eluting species.

The chromatograms for all four samples show two major elution peaks; the first is highlighted in red and the second in blue in Figure 5-14. The corresponding

deconvoluted MS spectra for the red species are shown in Figure 5-15, whilst the deconvoluted spectra for the red peak is shown in Figure 5-18)

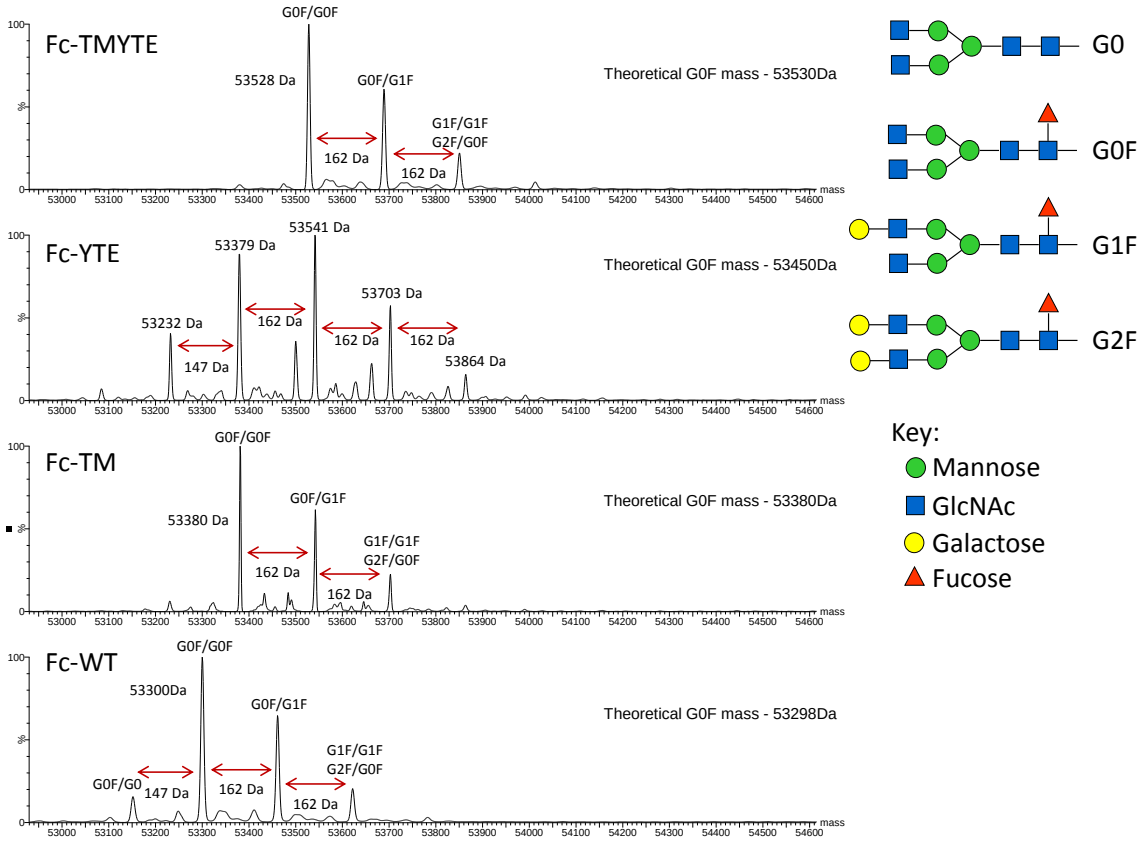


Figure 5-15 – The deconvoluted MS spectra of the FcWT, Fc-TM, Fc-YTE and Fc-TM-YTE. F-WT, Fc-TM and Fc-TMYTE show good agreement between experimental and expected

Figure 5-15 corresponds to the red LC peak highlighted in Figure 5-14. Considering the Fc-WT, Fc-TM and Fc-TMYTE samples, all four show clear evidence of being Fc. There is a strong agreement between the observed molecular weights expected based on the primary sequence for these three samples, indicating that they are the Fc's. There are also a number of peaks separated by 162 Da in the Fc-WT, Fc-TM and Fc-TMYTE spectra, indicating glycan patterning consistent with that observed in IgG Fc's (an increase in the galactose moieties on either the Man (α1-6) or Man (α1-3) arms). The data presented in Figure 5-15 corresponds to the non-reduced Fc, and therefore contains two copies of each glycan. Each increase of 162 Da corresponds to the addition of a single galactose to one of the two glycans, but it is impossible to identify which glycan has had the addition. For the addition for 324 Da, corresponding to two galactose moieties, there is some ambiguity as to whether

this corresponds to equal glycosylation across the two glycans (leading to G1F/ G1F) or both glacatose attaching to the same glycan (leading to G2F/G0F).

In all cases the dominant glycoform is the G0F/G0F, which is the expected major glycoform. There is a high degree of certainty that the digestion and purification method has yielded correct Fc fragments for these three samples. The Fc-YTE data, however, does not agree with the other samples, nor with the expected data. The glycan profile shows the addition of 147 Da, consistent with a fucosylation, followed by a number of 162 Da additions, which is consistent with the addition of a number of galactose moieties. This glycan pattern, alongside the observed molecular weight, supports the theory that it is an IgG Fc, however the glycan profile (the relative intensities of the glycans) is inconsistent with that of the other three IgG1 samples. Assuming glycans are similar between the Fc-WT shown in Figure 5-15 and "Fc-YTE" sample, this would mean that the G0F glycoform mass for the "Fc-YTE" is 53379 Da, a 71 Da mass error compared to the theoretical. This suggests that the sample provided was not that of an IgG-YTE or, if it was, not of the IgG1 subtype as previously thought.

Following this observation the IgG YTE sample was replaced with a new IgG-YTE from a different production batch. This YTE was subject to the same digestion and purification steps as before, and characterised using same LC-MS methods. Figure 5-17 shows the same data shown in Figure 5-15, with the incorrect "Fc-YTE" data being replaced by that obtained from new Fc-YTE.

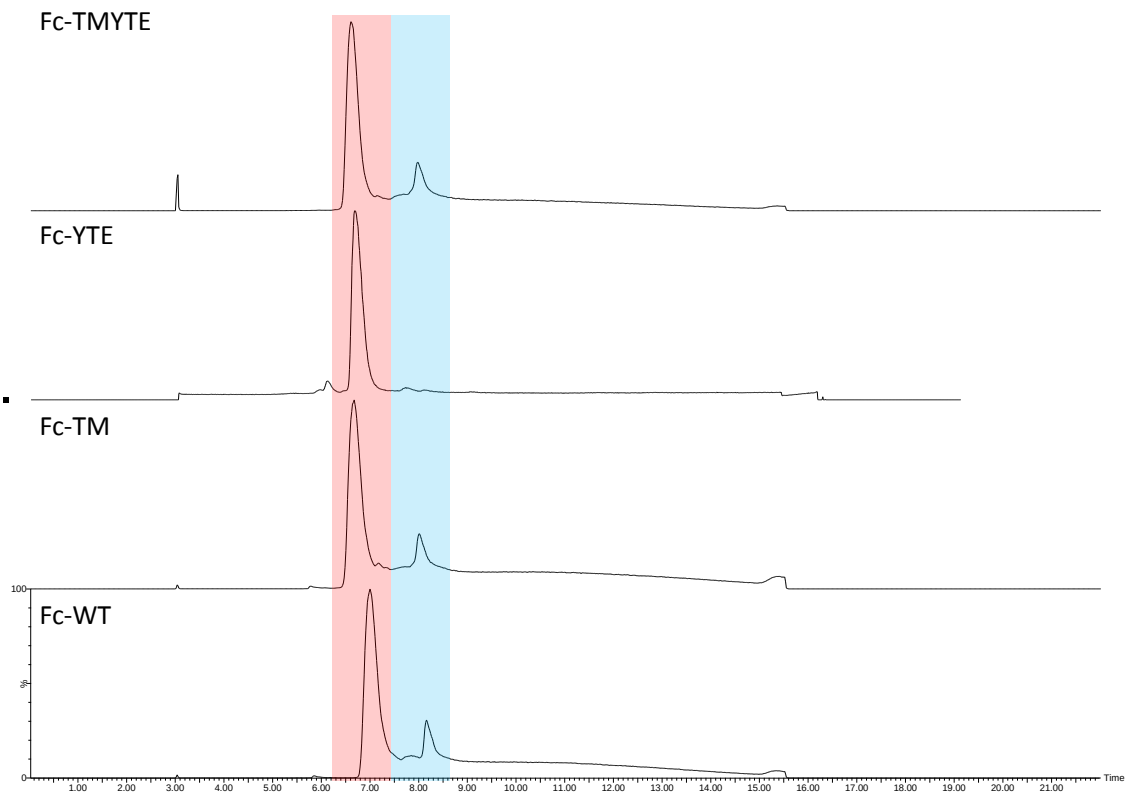


Figure 5-16 – LC-MS chromatogram for the four IgG-Fc samples. Following the second digestion and purification step, the Fc-YTE indicate only a single eluting species compared to the Fc-WT, Fc-TM and Fc-TMYTE.

Following a second preparation and purification of Fc-YTE (Fc-YTE preparation 2) the LC chromatogram looks similar to that presented in Figure 5-14, with the obvious exception of the missing peak highlighted in blue in the other species.

The deconvoluted MS spectra for the Fc-YTE preparation 2 elution peak (red) is shown in Figure 5-17.

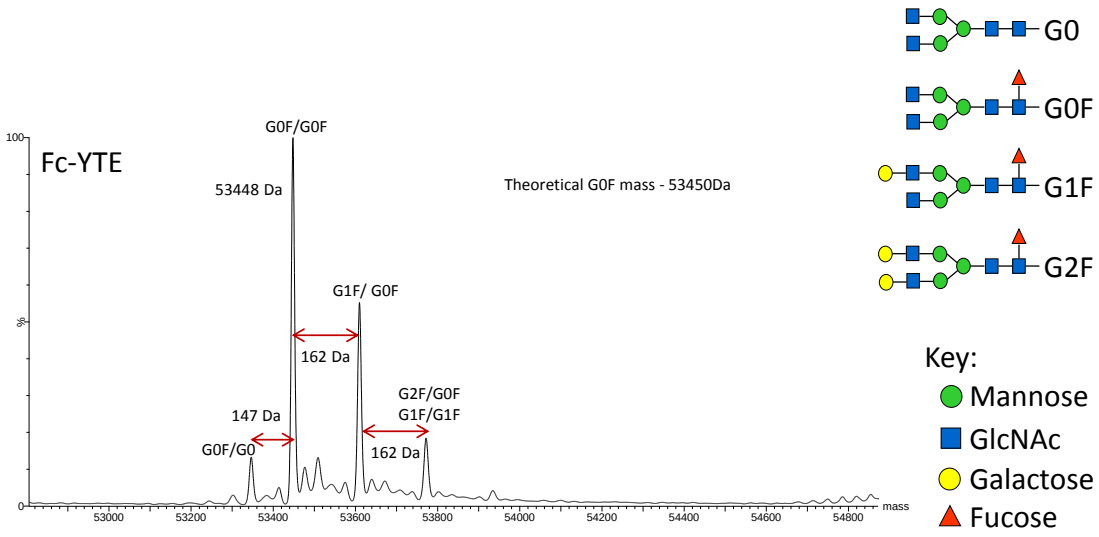


Figure 5-17 – The deconvoluted MS spectra for the Fc-YTE from preparation 2.

The deconvoluted MS spectra for Fc-YTE preparation 2 in Figure 5-17 shows significantly improved agreement with that of the Fc-WT, Fc-TM and Fc-TMYTE shown in Figure 5-15 illustrating once again previous "Fc-YTE" was not the correct sample. Figure 5-17 shows a glycan pattern consistent with previous data, and the expected data, along with an accurate mass that agrees with the theoretical mass to within 2 Da.

Deconvoluted MS spectra relating to the species highlighted in blue for Fc-WT, Fc-TM and Fc-TMYTE are shown Figure 5-18. All three spectra show the presence of two distinct molecules with molecular weight of 47115±1 Da and 47228±1 Da in all species. The strong agreement between the molecular weights, and retention times, between the three samples supports the suggestion that these are the same molecules. The IgG-WT, IgG-TM and IgG-TMYTE provided have identical V domain sequences with variations only in their Fc domains. The blue species present in the chromatogram in Figure 5-16 corresponds to the Fab region of the Fc which has co-eluted with the Fc. The theoretical mass of the Fab fragment of these three samples is 47116.6 Da, which agrees well with the masses shown here.

It is likely that some Fab protein was non-specifically bound to the protein A column used during the purification of these three samples. Given that these three samples were run at the same time, it would suggest that an extra wash step would be required to remove any non-specifically bound proteins during our purification.

The new Fc-YTE sample was run on a different occasion and contains little/ no Fab carryover. The reasons for this are not obvious, but may be as a result of minor changes to the elution protocol/ different buffers used between days.

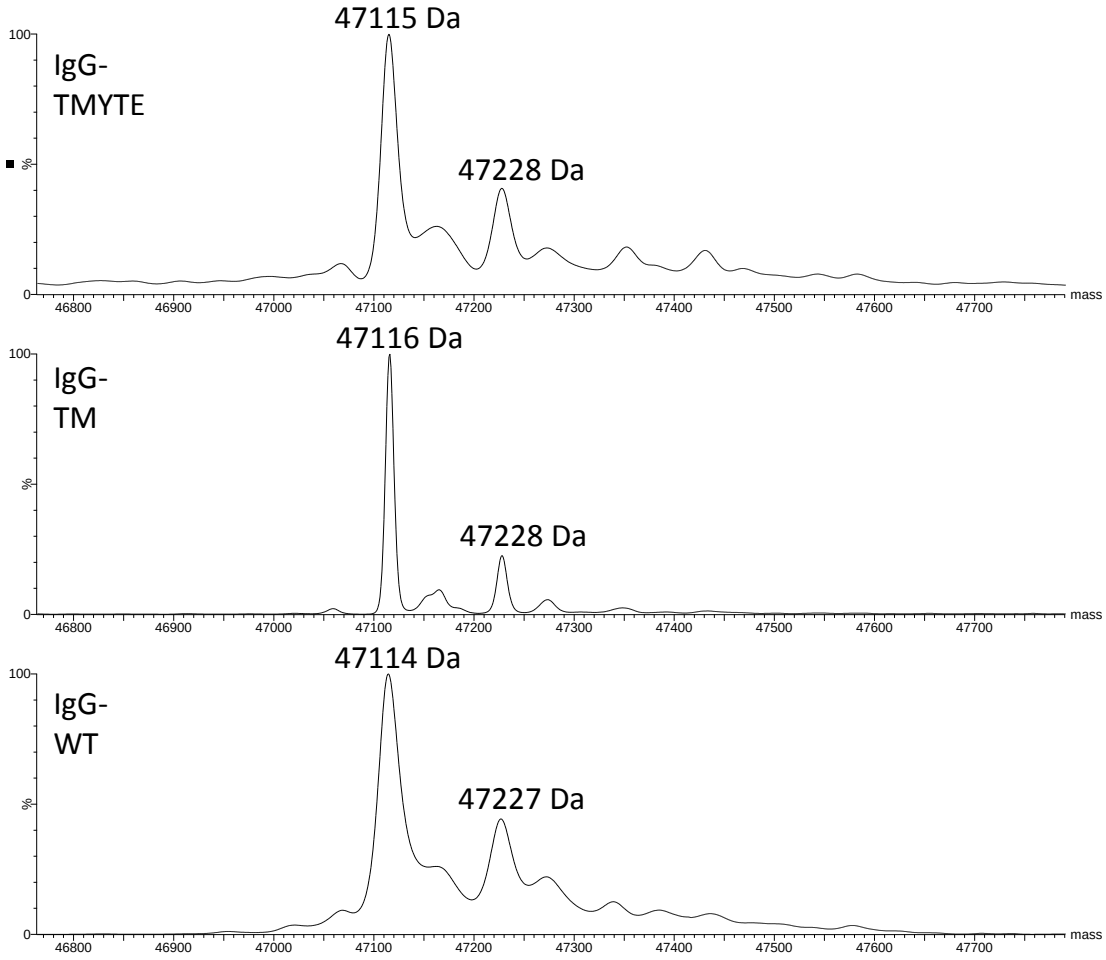


Figure 5-18 – The deconvoluted MS spectra of the elution species highlighted in blue in Figures 5-14 and 5-16

Figure 5-18, as mentioned, shows a second species with a molecular weight 113 Da larger than the theoretical mass of the Fab. Given that the focus of this study is on structural studies associated with the Fc region, no further investigation was performed on the Fab.

Having successfully identified and purified an authentic Fc-YTE sample an additional experiment was conducted to investigate Fc-YTE affinity with CaptureSelect™ matrix. The results are shown in Figure 5-19.

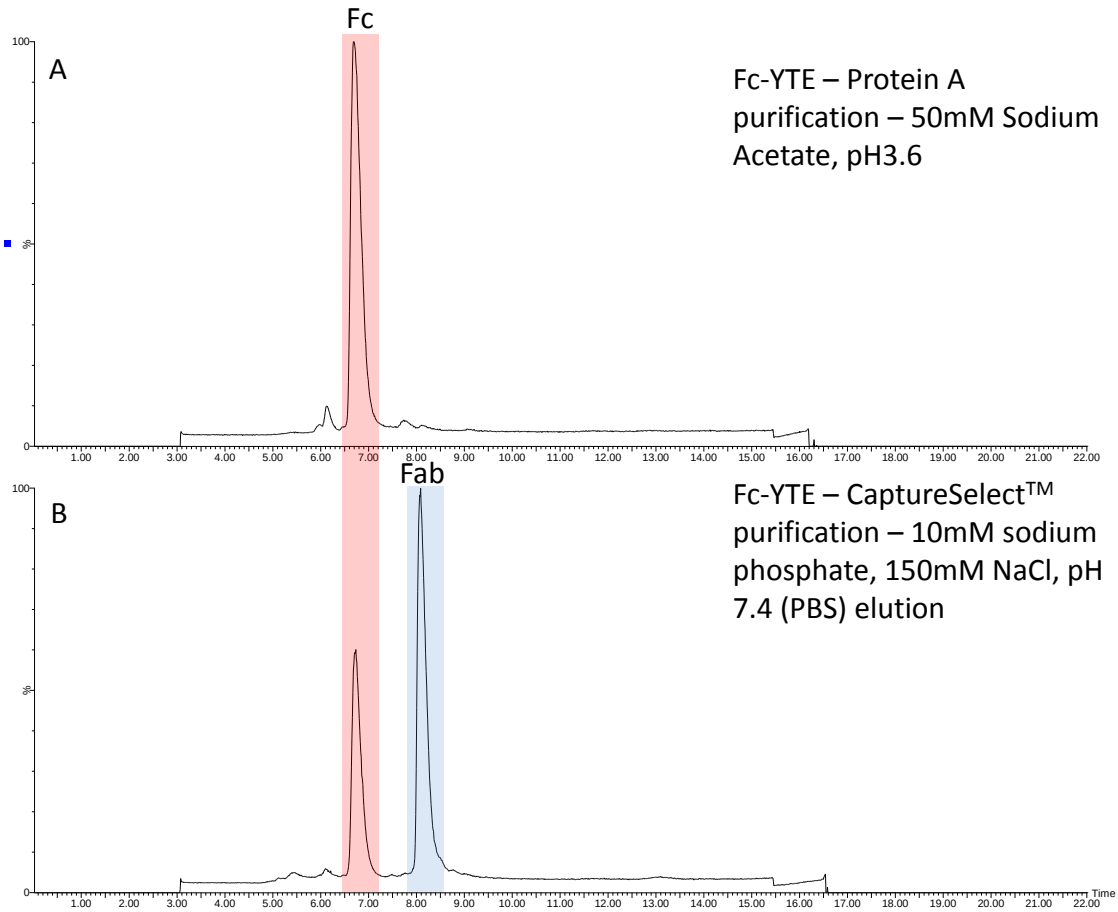


Figure 5-19 – A) The Fc peak for purified Fc-YTE obtained following the acid elution from a protein A affinity column. B) The result following a weak, neutral pH elution from and Fc-YTE/ Fab mixture incubated on CaptureSelect™ affinity matrix

Figure 5-19 shows that, following incubation with CaptureSelect™ affinity matrix, the Fc-YTE elutes under the same conditions as the Fab fragment, indicating that, as previously observed for the Fc-TMYTE, Fc-YTE does not bind to the CaptureSelect™. This observation suggests that the lack of affinity Fc-YTE and Fc-TMYTE is likely due to disruptions in the CaptureSelect™ epitope, rather than a conformational rearrangement induced by a combination of the TM and YTE mutations.

5.2.2 Native MS

Following confirmation of the sample identities, and successful purification of Fc from the Fc/ Fab mixture, the different antibody Fc's were analysed using native MS. It has been observed that changes in charge state distribution can be indicative of differences in conformation (Katta and Chait 1991, Benesch, Sobott *et al*. 2003, Heck and van den Heuvel 2004, Wyttenbach and Bowers 2011, Konijnenberg, Butterer *et al*. 2013), and it was possible that global changes in Fc conformation, as a result of TM/ YTE mutations, would result in changes in the charge distribution. None of the amino acids introduced as part of the TM or YTE mutations are positively charged, and therefore we would only expect to see a shift in the positive MS spectra as a result of conformational change. The results obtained are shown in Figure 5-20.

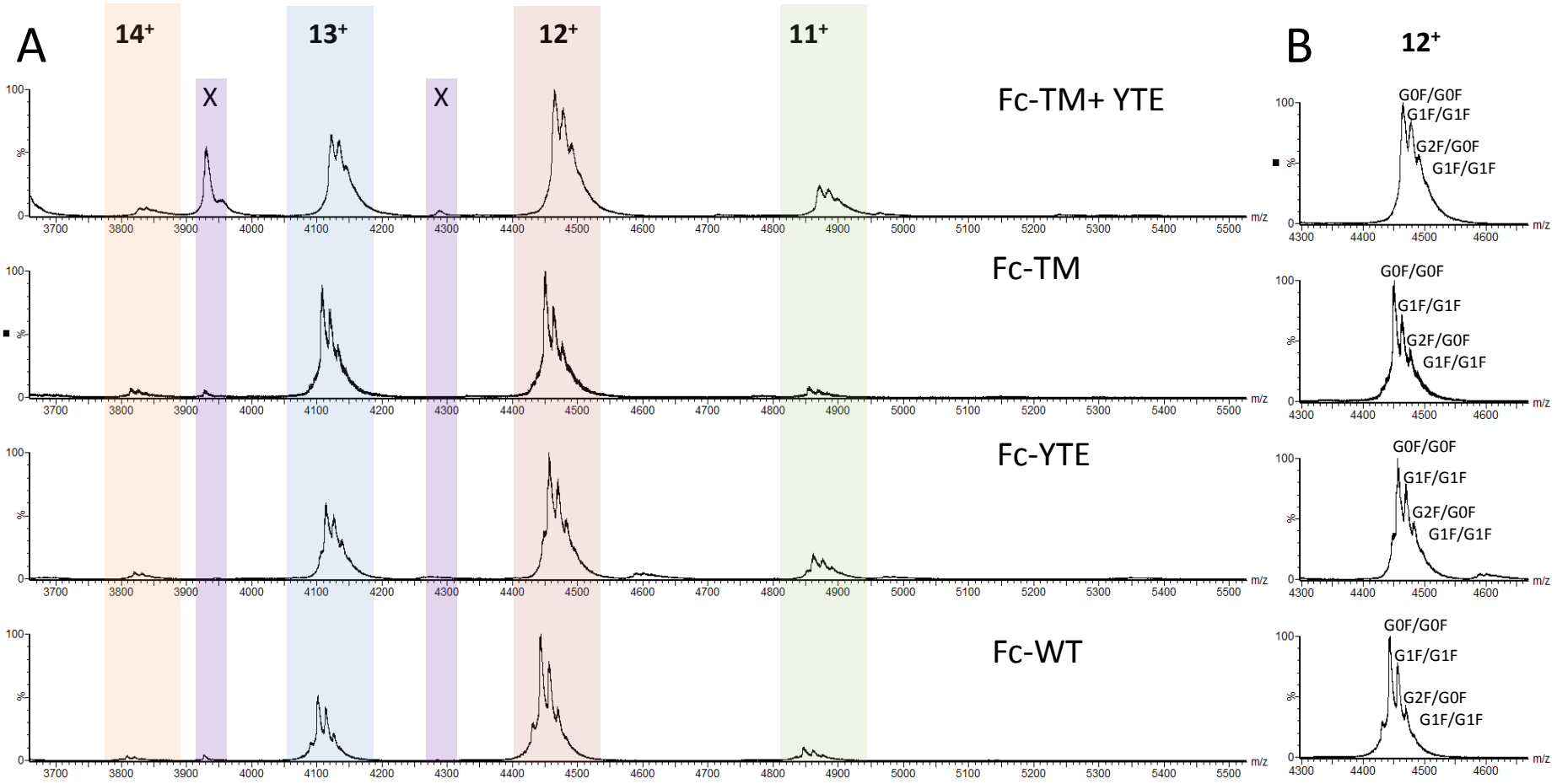


Figure 5-20 – A - the native MS spectra of the four Fc constructs, Fc-WT, Fc-YTE, Fc-TM and Fc-TM-YTE, labelled with their $11^+ \rightarrow 14^+$ charge states. Species labelled X, most prevalent in the Fc-YTE spectra, corresponds to native Fab fragments. B – A focused image of the 12^+ charge state, illustrating the major glycoforms present. Fc-YTE and Fc WT also show evidence of a minor G0 glycoform, consistent with the observations from the LC-MS experiment

The native MS of all four Fc samples indicate few differences between the samples. All four Fc's show similar charge states, $11^+ \rightarrow 14^{+,}$ with a broadly similar charge state distribution, centred on the 12^+ charge state, consistent with previous studies on IgG Fc domains (Pacholarz, Porrini *et al*. 2014). This suggests that the four molecules adopt similar conformations in solution.

Resolution obtained on the MS spectra was sufficient to observe the different glycoforms present on the Fc. The peaks are separated by a mass difference of 162 Da, which is consistent with the addition of a single glycan moiety. This observation is consistent with that presented in Figure 5-15 and Figure 5-17.

The Fc-TMYTE sample shows two peaks, labelled X in Figure 20, which are also present in low levels in the Fc-TM and Fc-WT samples. These peaks were identified as belonging to the Fab fragment, and were therefore ignored for the purposes of analysis and processing.

5.2.3 Ion mobility MS

Ion mobility mass spectrometry was then used probe to the conformation of the proteins, as described in Section 1.1.6 Ion Mobility Mass Spectrometry

Comparison of the conformation of different glycoforms in the Fc-WT

X-ray diffraction studies have suggested that different the inter-domain spacing is affected by different Fc glycoforms (Krapp, Mimura *et al*. 2003). This was investigated here using IMS, since any changes as a result of different glycosylation profiles would need to be considered when obtaining IMS measurements. The results for the Fc-WT construct are shown in Figure 5-21.

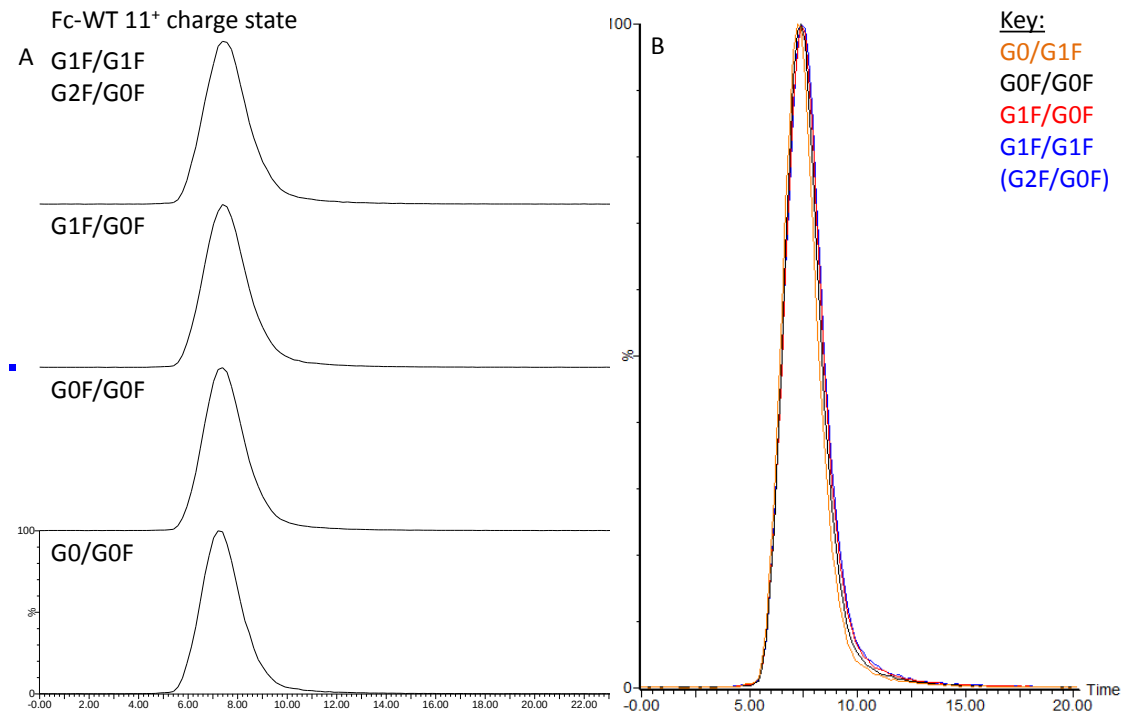


Figure 5-21 – A comparison of the arrival time distributions of the four observed glycoforms for the Fc-WT construct.
No differences are observed in either drift time, or arrival time distribution, suggesting that the glycan has no observable effects on the conformation of the Fc observed by IMS

Figure 5-21 shows no significant differences in the arrival time distribution (ATD) for the Fc-WT sample as a result of differences in glycosylation. Ion mobility provides a measure of the collision cross section (CCS) of a protein, determined by the broadly spherical shape it adopts when tumbling through the mobility cell. The longest dimension of a proteins structure is most representative of the diameter of this spherical shape, and thus is the most important dimension used to determine CCS. Small changes in protein conformation which do not affect this value are unlikely to be observed since they result in little change to the overall shape of the sphere. The structure of the Fc, complete with atomic measurements, is shown in Figure 5-22. As can be seen in Figure 5-22, the longest dimension is that that which spans the full height of the Fc with hinge, referred to as the Hinge→C_H3 dimension, with a length of 104.5 Å (shown in magenta). This structure represents just one possible hinge conformation, since there are no interactions that anchor the hinge to the core C_H2 domain, affording complete flexibility. A C_H2→C_H3 dimension, measuring 78.2 Å, was also determined spanning the full height of the Fc core without the hinge to illustrate that, even without hinge or with hinge lying flat against the core, the dimension spanning the two C_H2 domains remains the smallest

dimension. It is likely that any changes in this space, particularly changes in domain spacing of approximately 5 Å which have been previously reported (Krapp, Mimura *et al*. 2003), are unlikely to be observable using the IMS approach.

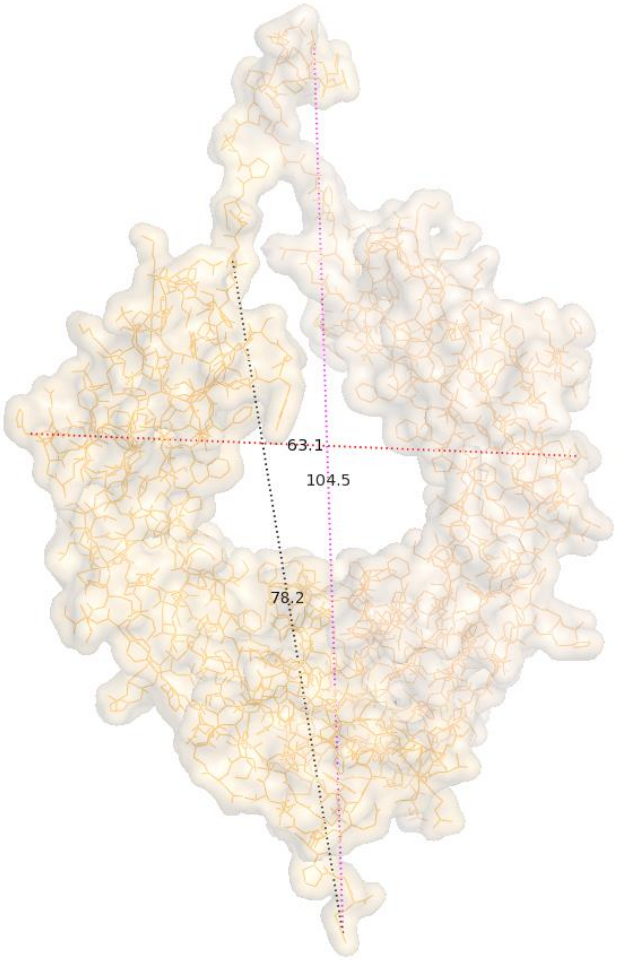


Figure 5-22 – A homology model of the Fc-WT construct complete with hinge in a hinge-up view. Three atomic measurements are shown on the figure. Black shows the $C_H2 \rightarrow C_H3$ atomic distance, magenta shows the Hinge$\rightarrow C_H3$ atomic distance, red show the C_H2 spanning atomic distance. All measurements are shown in Å.

A comparison of the Fc constructs by IMS

Having shown that the glycosylation profile did not significantly affect the ATD of the protein, the 11^+ charge state for all four Fc variants was analysed for differences. All glycoforms in the 11^+ charge state were combined for this analysis. Results are shown in Figure 5-23.

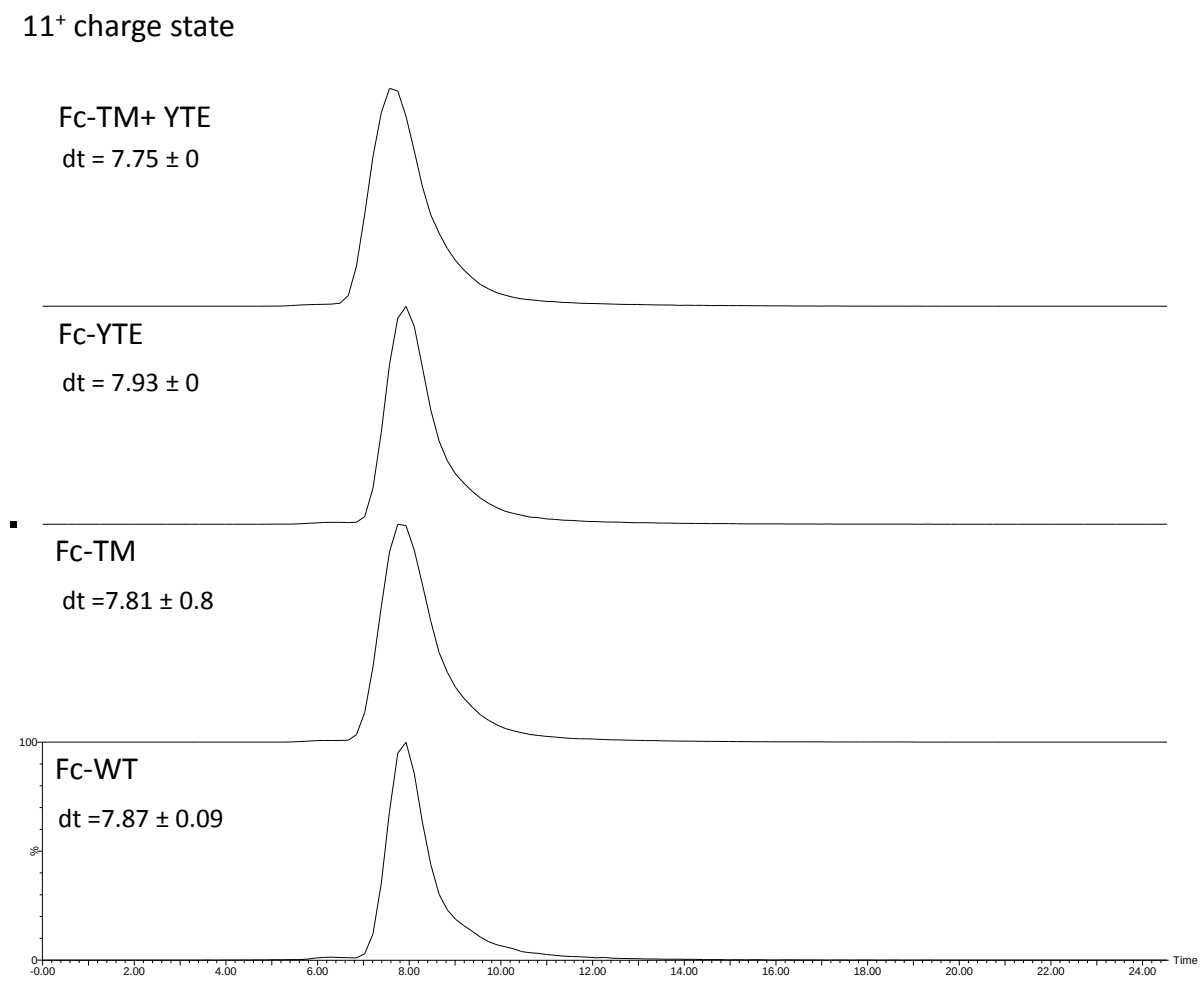


Figure 5-23 - The arrival time distributions for the 11^{+} charge state of the Fc constructs.
Drift times show small changes between samples that are within the experimental error of the instrument

Small differences were observed in drift times between the four Fc constructs, but repeated experiments indicate that these differences were within experimental error and are not significant. This suggests there were no measureable global changes in conformation associated with any of the constructs. This is not a surprising observation and is supported by previous X-ray diffraction studies by Oganeysan *et al* which suggested there were no large conformational changes associated with these mutations (Oganesyan, Gao *et al*. 2008, Oganesyan, Damschroder *et al*. 2009).

Table 5-1 shows a comparison of the arrival time distributions, measured as peak width at half maximum (PWHM).

	Drift time chromatogram width at 50%
Fc-WT	1.07 ± 0.12
Fc-TM	1.11 ± 0.09
Fc-YTE	0.94 ± 0.11
Fc-TMYTE	1.21 ± 0.14

Table 5-1 - A comparison of the FWHM from the drift time chromatograms.

The width of the ATD can be considered to be representative of the dynamics of the protein, and related to the flexibility of the conformation adopted by a protein. The data suggest that Fc-TMYTE may have a slightly more flexible conformation that either Fc-TM or Fc-YTE. The data also suggest that Fc-YTE has a more rigid structure than that of the Fc-WT. These observations may echo the minor C_H2 domain spacing differences highlighted in Figure 5-5, suggesting minor changes in domain flexibility. Concentration-dependent peak broadening was observed when analysing these proteins. Whilst peak broadening occurred when concentration changed as a factor of 10 (Figure 5-24), smaller fluctuations in spray conditions which influence the protein concentration in the droplet may play a role in the observations presented in Table 5-1 and may also play a part in the larger errors shown in Table 5-1.

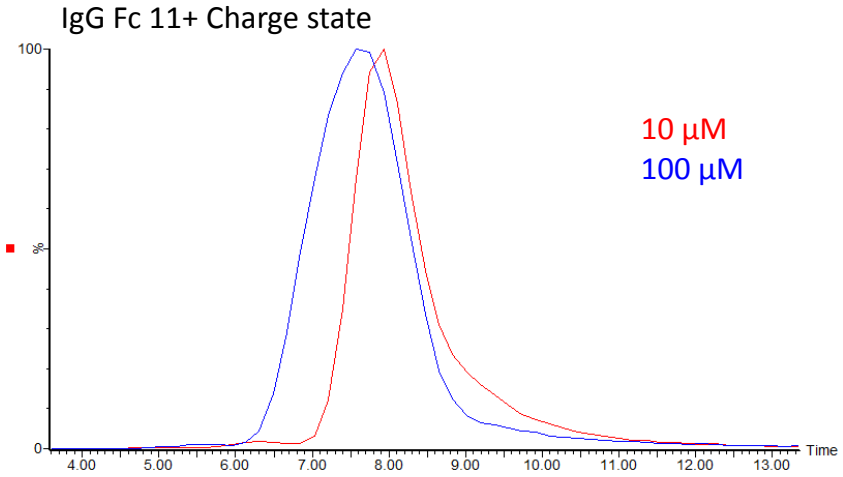


Figure 5-24 – An example of the peak broadening observes as a factor of concentration.

Following the analysis of the 11^+ charge state, all four charge states of the Fc-WT, Fc-TM, Fc-YTE and Fc-TMYTE were compared. In order to provide better clarity regarding the different conformations, CCS estimates for each of the drift times were calculated. The calibration curve used is shown in Figure 5-25, and calibration was performed using established methods (Scarff, Thalassinos *et al*. 2008, Williams and Scrivens 2008, Thalassinos, Grabenauer *et al*. 2009).

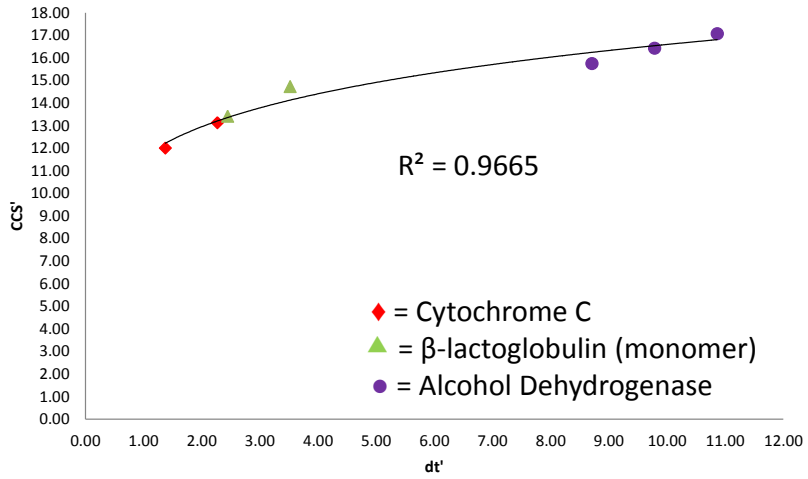


Figure 5-25 – The calibration curve used for the calculation of CCS values for the four Fc constructs

Figure 5-26 shows the arrival time distributions obtained for the four Fc constructs, complete with CCS estimates calculated for the different conformers.

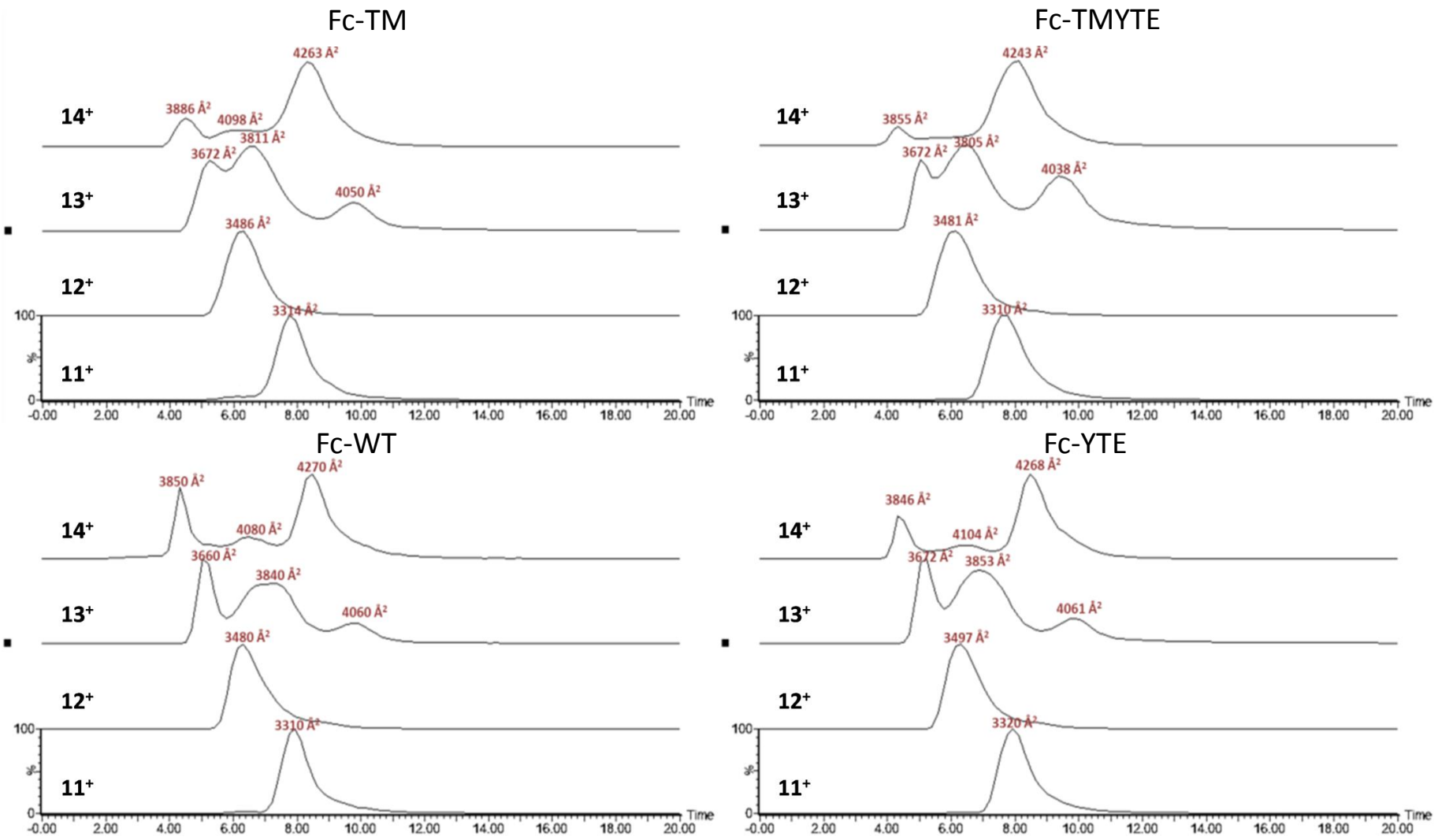


Figure 5-26 – The drift times, and CCS estimates, for the four Fc constructs.

Figure 5-26 shows that, with increasing charge the calculated protein CCS increases as the Fc begins to unfold. At higher charge states this results in the unfolding of the protein into multiple conformations, which is consistent with previous reports on the behaviour of proteins in the gas phase (Scarff, Thalassinos *et al*. 2008).

Similar to the results obtained from the 11^+ charge state, the 12^+ charge state shows only a single observable conformation. The ATD, however, shows tailing for the 12^+ charge state when compared with the 11^+, suggesting that the Fc is beginning to adopt more extended conformations.

The 13^+ charge state for all species shows major peaks corresponding to a ~3660 $Å^2$ species, a ~3840 $Å^2$ species and a ~4060 $Å^2$ species (as observed in the wild type). The peak width of the 3840 $Å^2$ species suggests at least two conformations which cannot be resolved; this observation is supported by the observation of a broad ~3811 $Å^2$ peak present in the Fc-TM and Fc-TMYTE 13^+ ATD not present in the Fc-WT or Fc-YTE. This suggests that, as a result of TM mutations the Fc-TM and Fc-TMYTE follow a slightly different unfolding mechanism, as they can more readily adopt an extended conformation not seen to the same degree in either Fc-construct without the TM mutations.

The 14^+ charge state also contains three dominant conformations; a 3850 $Å^2$ species, a 4080 $Å^2$ species observed only in the Fc-WT, Fc-TM and Fc-YTE, and a 4270 $Å^2$ species. The lack of an observed 4080 $Å^2$ species the 14^+ Fc-TMYTE data suggests the protein is able to completely unfold into the more extended conformation, rather than adopting a partially unfolded intermediary conformation.

Data for the four Fc constructs shows that whilst the proteins exhibit species with similar CCS values, and therefore potentially similar conformations, there are differences observed in intensities of the more extended conformations between the mutants, suggesting possible differences in stability between the constructs. Fc-WT consistently shows the highest relative intensity for the most compact conformation at all charge states, suggesting it is the most stable. The Fc-TM, Fc-YTE and Fc-TMYTE show higher observed intensity for intermediate conformations (3800 $Å^2$ species in the 13^+ and 4080 $Å^2$ in the 14^+ ATDs) compared to wild type, with the Fc-TMYTE showing the highest relative intensity of all three for this intermediate state, as well as the highest intensity for the most extended conformation observed (4040

$Å^2$ in the 13^+ and 4270 $Å^2$ in the 14^+). This would suggest that Fc-WT is the most stable construct, with Fc-TMYTE being the least stable; this observation agrees with data obtained from DSC shown in Figure 5-7. Fc-YTE appears more resistant to gas phase unfolding that Fc-TM based on Figure 25, which is not in agreement with the DSC data presented in Figure 5-7 and may suggest different mechanisms of unfolding for Fc's between solution and gas phase.

Ion mobility has shown a number of minor structural differences between the four Fc constructs, but no major structural changes at the lowest energy state. With increasing charge there are observable stability differences between the four Fcs. In order to investigate the conformational and dynamics differences on a more local level the four Fc's were analysed using HDX-MS.

5.2.4 Hydrogen Deuterium Exchange MS

In order to obtain data on the whole Fc it was essential to obtain as close to 100% sequence coverage as possible. It was also important to obtain sequence coverage across the regions of mutation; the upper C_H2 and hinge for the TM mutation sites, and the lower C_H2 for the YTE mutation sites. The sequence coverage maps for the four Fc regions are shown in Figure 5-27 for Fc-WT, Figure 5-28 the Fc-TM, Figure 5-29 for Fc YTE and Figure 5-30 for Fc-TMYTE.

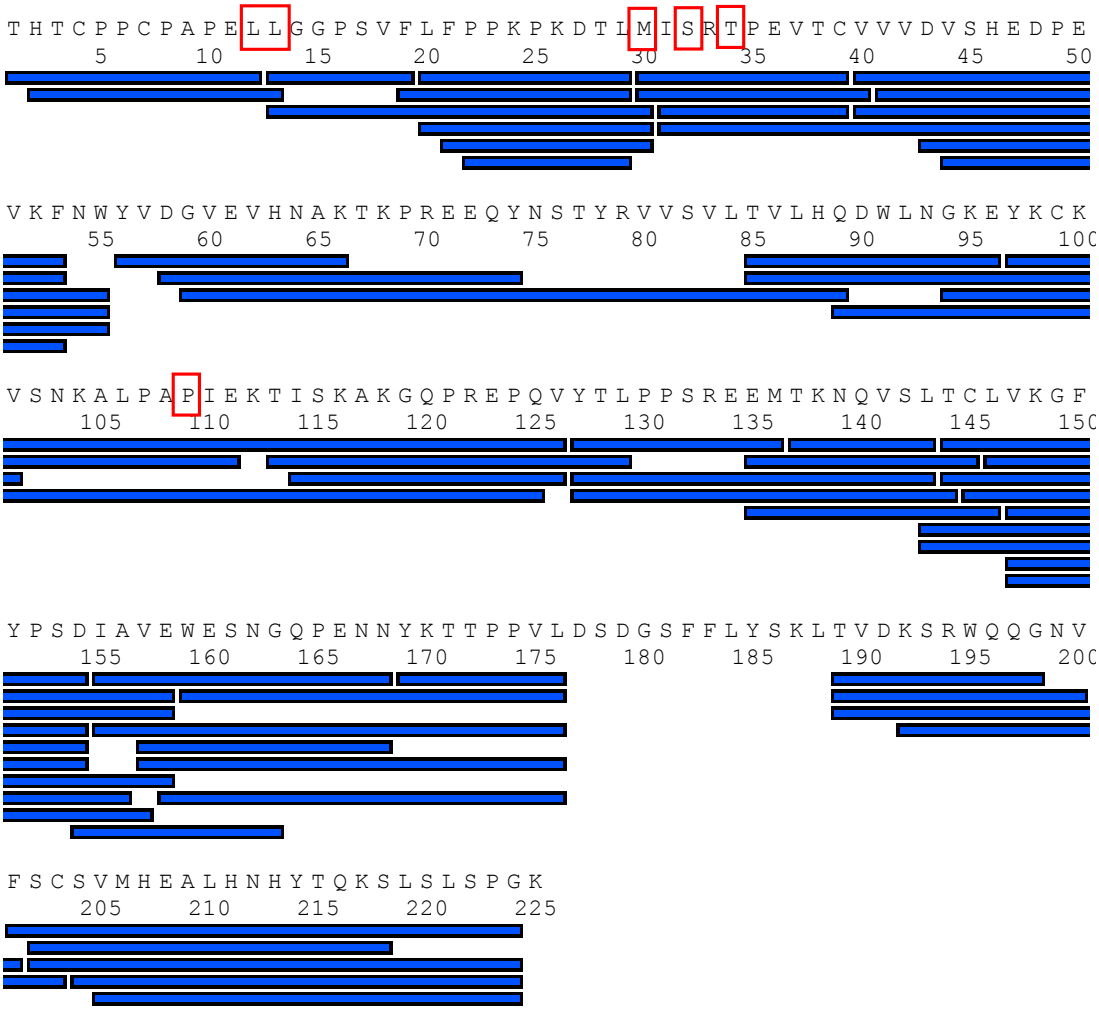


Figure 5-27 – The peptide coverage map for the Fc-WT construct. Mutation sites are highlighted with red boxes. Blue boxes located under sequence correspond to peptides identified using ProteinLynx Global Server. Sequence coverage corresponds to the number of residues identified, based on the total residues in the protein. Total sequence coverage was 94.2%

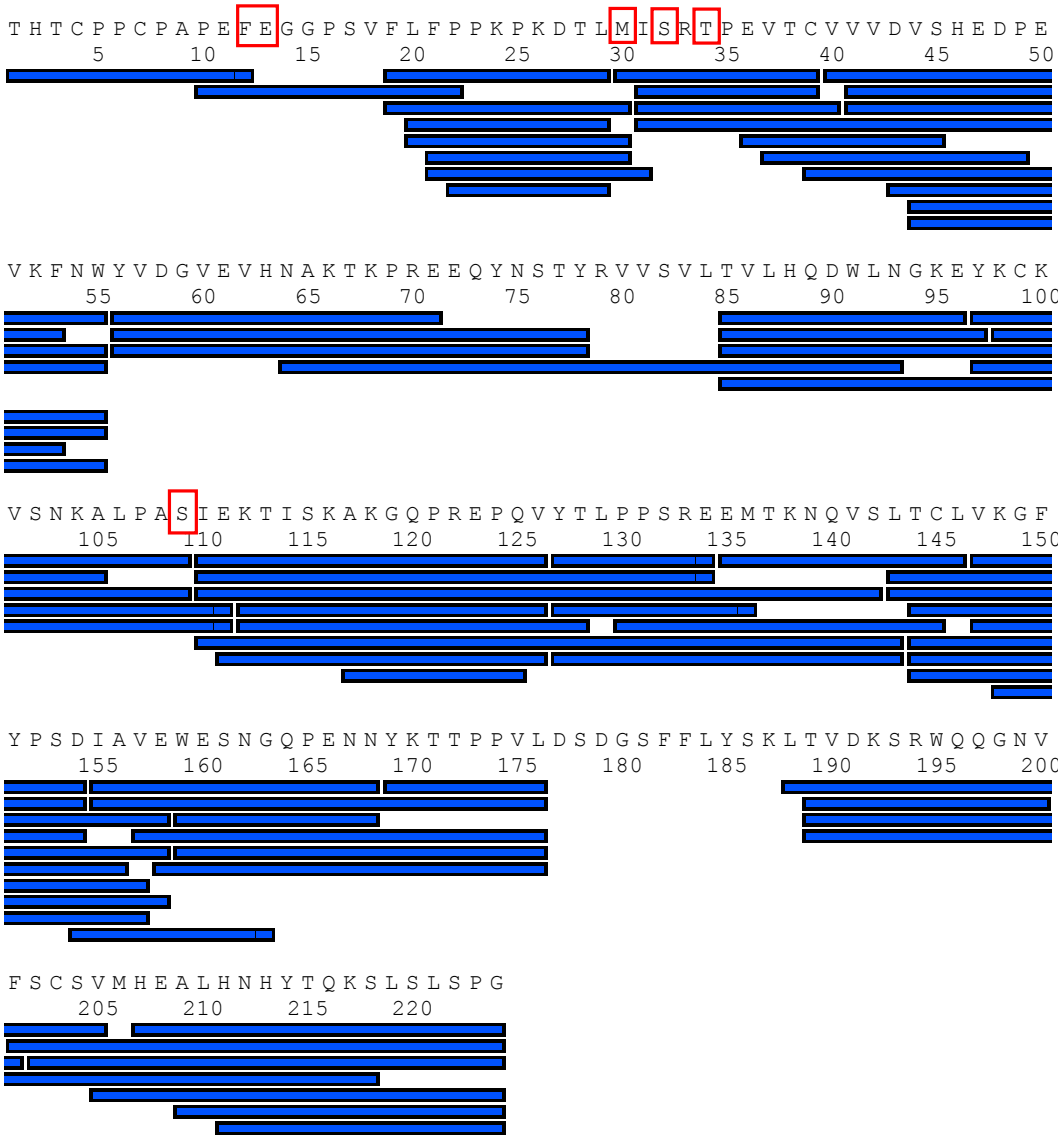


Figure 5-28 - The peptide coverage map for the Fc-TM construct. Mutation sites are highlighted with red boxes. Blue boxes located under sequence correspond to peptides identified using ProteinLynx Global Server. Sequence coverage corresponds to the number of residues identified, based on the total residues in the protein. Total sequence coverage was 95.1%

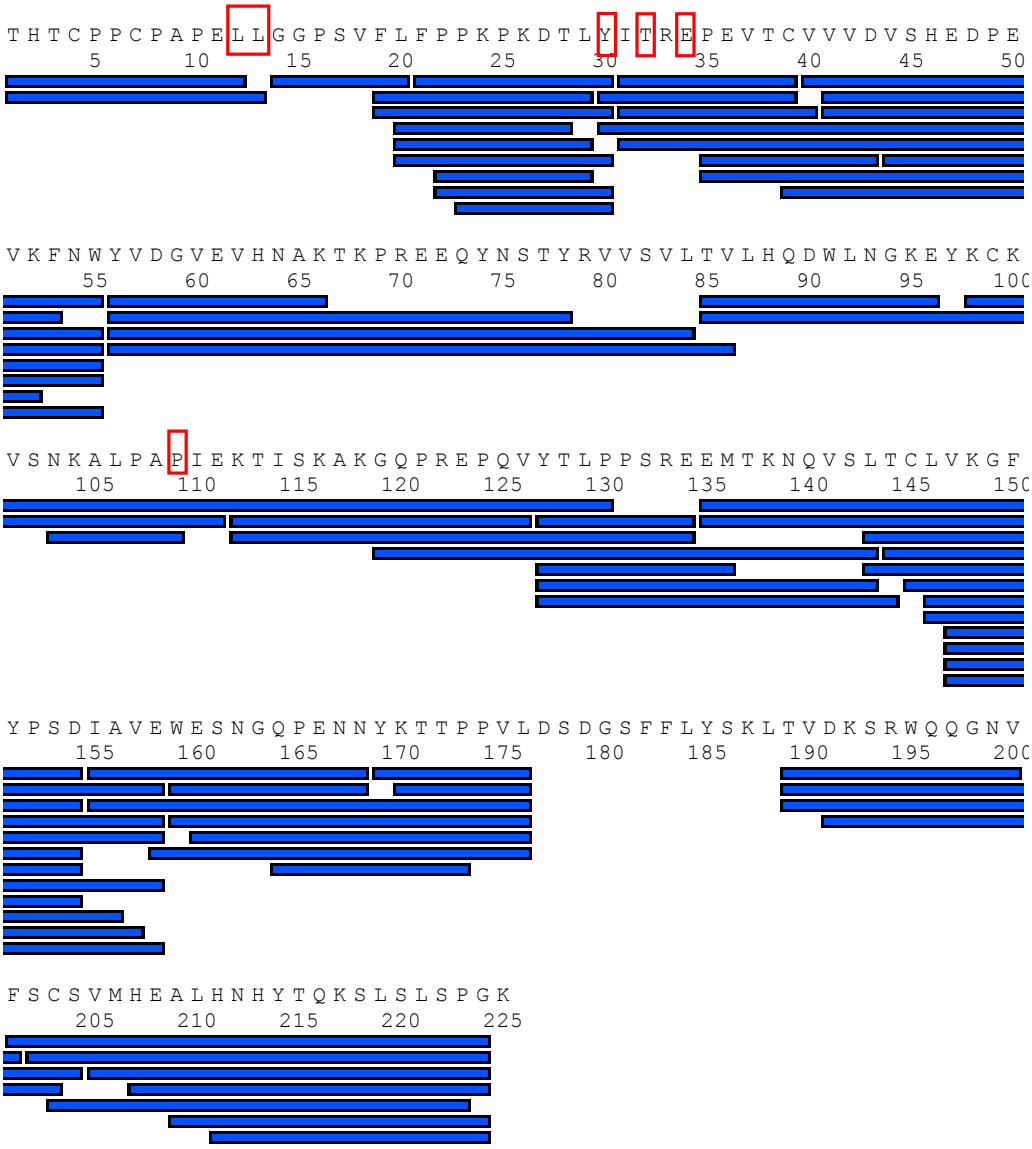


Figure 5-29 - The peptide coverage map for the Fc-YTE construct. Mutation sites are highlighted with red boxes. Blue boxes located under sequence correspond to peptides identified using ProteinLynx Global Server. Sequence coverage corresponds to the number of residues identified, based on the total residues in the protein. Total sequence coverage was 94.2%

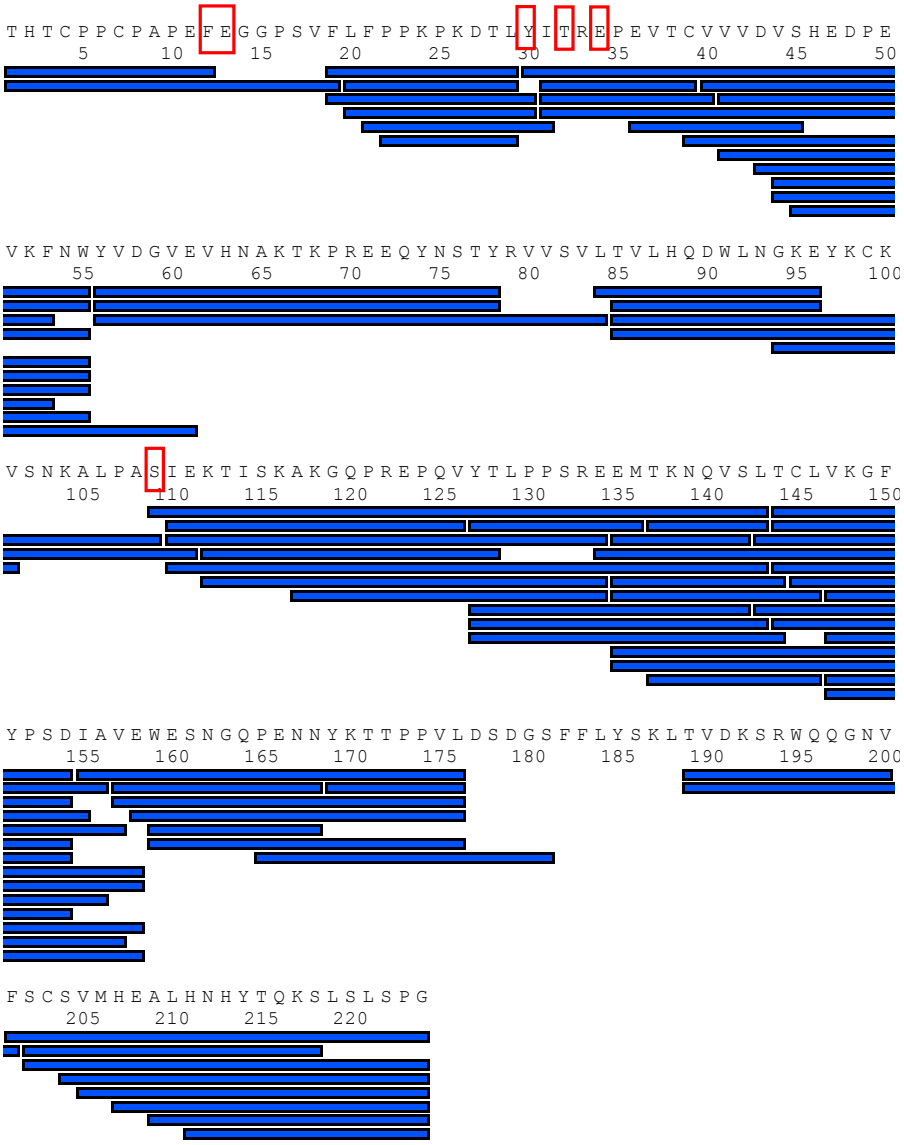


Figure 5-30 - The peptide coverage map for the Fc-TMYTE construct. Mutation sites are highlighted with red boxes. Blue boxes located under sequence correspond to peptides identified using ProteinLynx Global Server. Sequence coverage corresponds to the number of residues identified, based on the total residues in the protein.Total sequence coverage was 96.9%

Based on the criteria outlined earlier, all four peptide maps were suitable. All four peptide maps showed approximately 95% sequence coverage. Sequence coverage shows a high degree of redundancy within the data; there are a number of overlapping peptides across the protein which give higher confidence in the assigned sequence. There was also a degree of redundancy at the sites of mutation.

The region of the largest concern was that of the hinge. Peptides from the hinge were observed at a low abundance. It is possible a number of close proximity cleavages within the hinge resulted in over digestion of the hinge to low amino acid fragments. To ensure that key peptides of the hinge had been correctly assigned, these peptides

were selected for follow-up, targeted MS/MS fragmentation and manual fragment assignment. Results are shown in Figure 5-31.

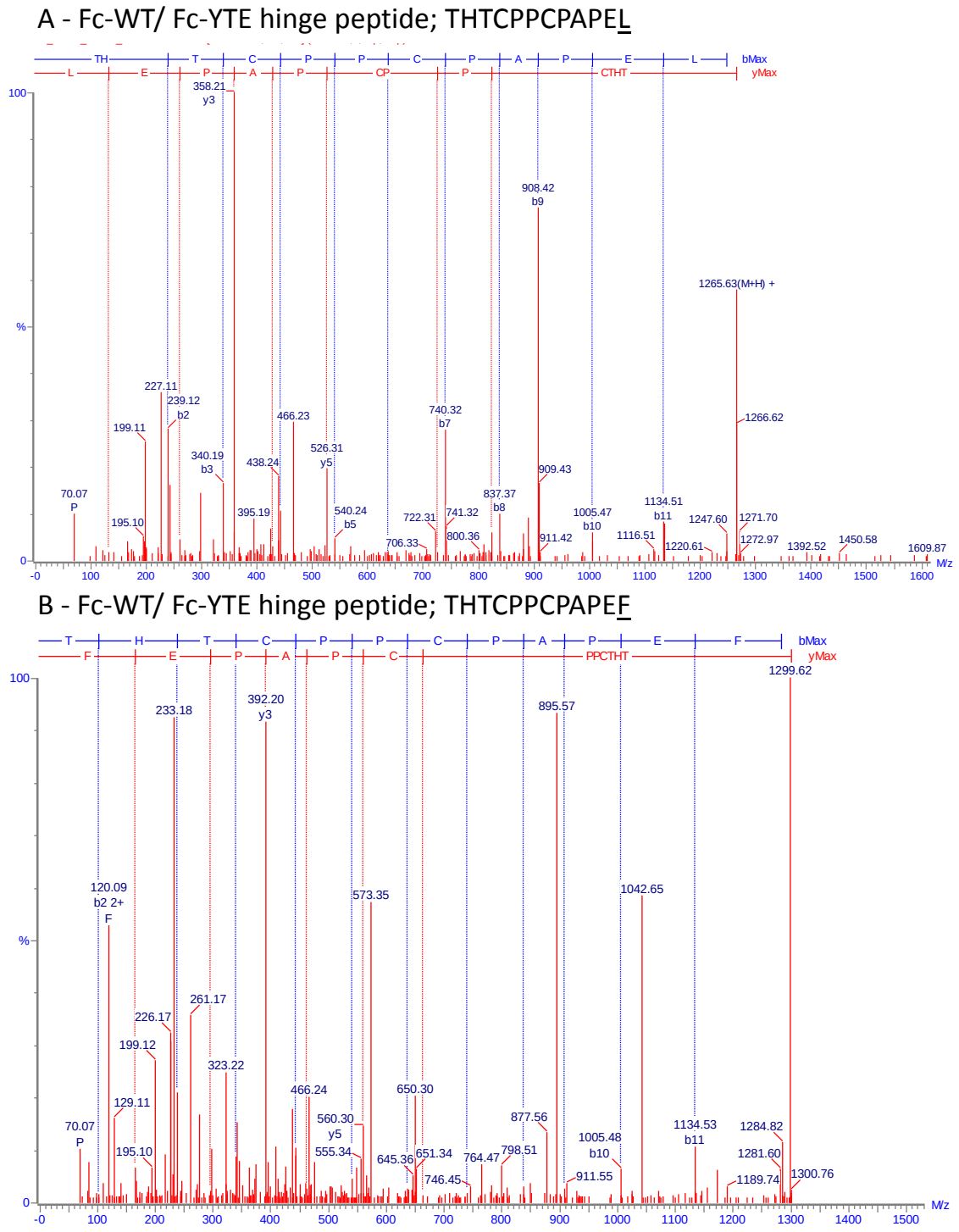


Figure 5-31 – The MS/MS analysis of two important peptides from the hinge region of the Fc constructs. A – The peptide found in the Fc-WT and Fc-YTE sample has a C-terminal leucine, whilst B – shows the same peptide with a C-terminal phenylalanine

Figure 5-31 shows the same peptide found in the Fc-WT/ Fc-YTE (A) and Fc-TM/Fc-TMYTE (B) proteins, based on presence of the L234F mutation. Both

MS/MS spectra show strong b- and y- ions, supporting the assignment made by PLGS and confirming the presence of these low abundance peptides.

Following identification of these key peptides, an HDX time course was run on the peptides and exchange tracked across the Fc. The exchange data was normalised against the Fc-WT and relative protection/ deprotection of Fc-TM, Fc-YTE and Fc-TMYTE is summarised in Figure 5-32.

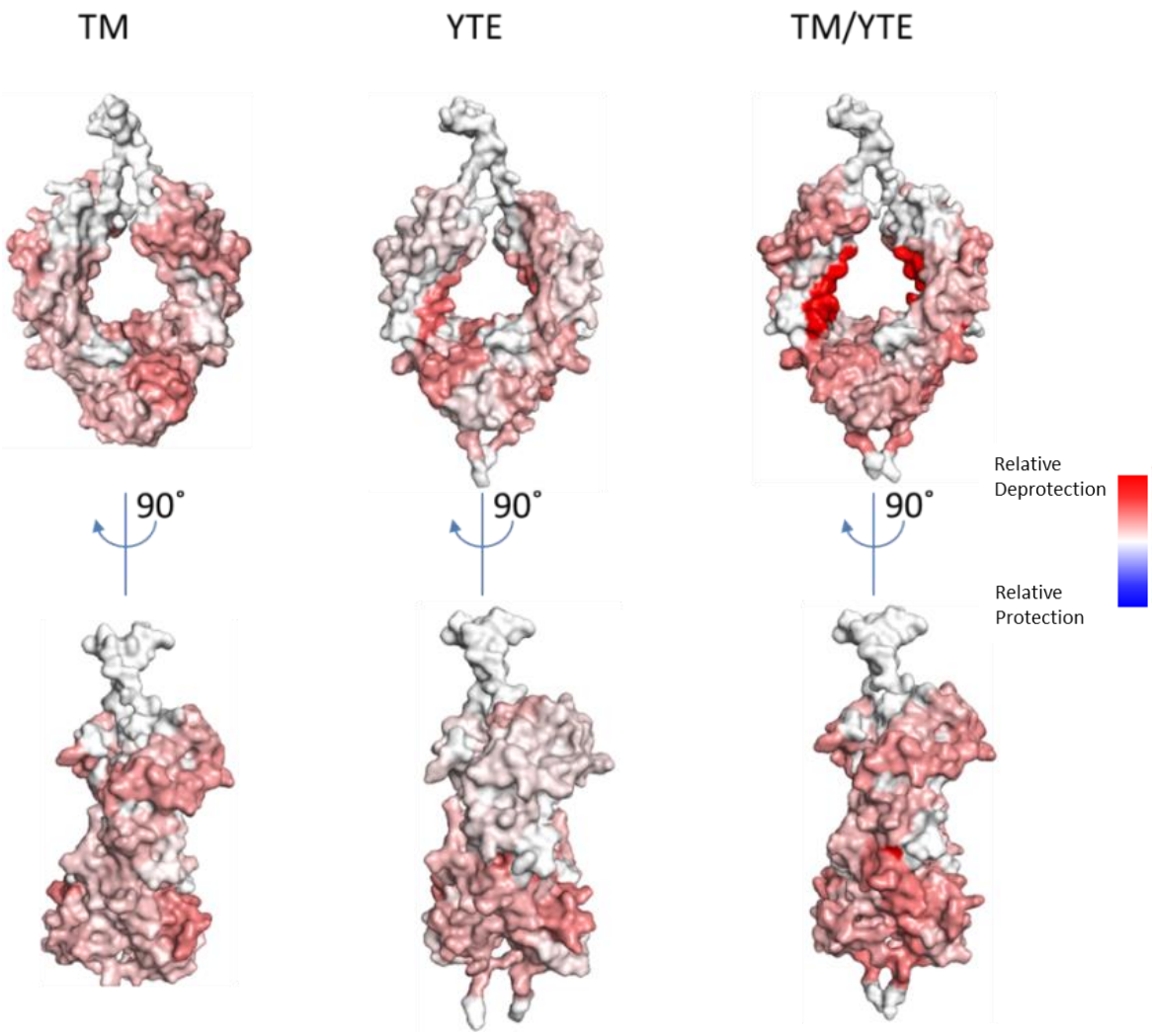


Figure 5-32 – Deuteurim uptake across the fc domain plotted as a heat trace the homology models of the Fc-TM, Fc-YTE and Fc-TMYTE. No data was fit for the hinge region.

No hinge data is modelled in Figure 5-32, as the Fc-WT structure used to normalise the data (crystal structure: 3AVE) had no hinge coverage to map against. The hinge exchange data is shown in Figure 5-33.

Figure 5-32 shows that both TM and YTE mutations introduce extra conformational freedom to the protein, since both mutants show an increase in the levels of

deuteration uptake across the protein (increased deprotection). No region of Fc in either of the Fc-TM or Fc-YTE shows increased protection, indicating that the mutations introduced only serve to destabilise the protein structure and/ or increase its dynamics. This observation is is consistent with DSC and IMS data presented earlier.

Whilst the main site of thermal destabilisation, indicated by DSC is located in the C_H2 domain for Fc-TM, Fc-YTE and Fc-TMYTE HDX shows the mutations have wide ranging effects on the overall dynamics of the protein. Both Fc-TM and Fc-YTE proteins have an observed increase in the deuteration in the C_H3 domain and at sites within the C_H2 both proximal and distal from the sites of mutation.

Fc-TMYTE shows a mostly additive effect in deuteration caused by local conformational changes across the protein , particularly within the C_H3 domain where the uptake plot shows an additive Fc-TM/ Fc-YTE pattern. There is a site within the C_H2 domain, located between the TM mutations L234F and L235E and YTE mutations M252Y, S254T and T256E, which appears to have a significantly increased level of deuterium uptake than any other region in the Fc.

Relative deuterium uptake plots are shown below for the peptides spanning the mutation sites.

One peptide spanning the hinge region (THTCPPCPAPEL) is found in all four constructs, the most C-terminal residue being L234 (the first amino acid of TM mutation site). HDX data shown in Figure 5-33 shows that all four mutants show a high rate of deuterium exchange in the hinge. This is unsurprising given solvent accessibility of the hinge relative to the rest of the protein. Fc-TM and Fc-TMYTE show a slightly elevated initial uptake, suggesting that the two constructs which share the mutated L234F both exhibit a similar increase in deuterium uptake.

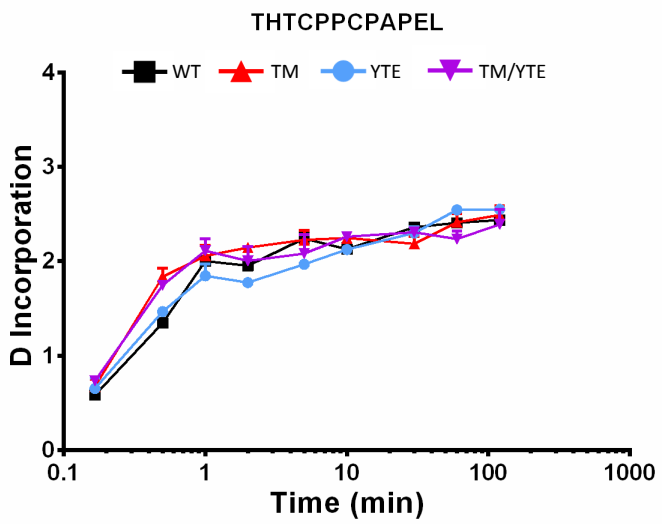


Figure 5-33 – The deuterium incorporation plot for a conserved hinge peptide found across all four Fc constructs. Residue 12, lysine (L), was mutated to phenylalanine (F) in Fc-TM and Fc-TMYTE constructs. Error bars show the standard deviation from repeated data sets

Three peptides were identified spanning the TM mutation site located in the upper CH2, and these are shown in Figure 5-34.

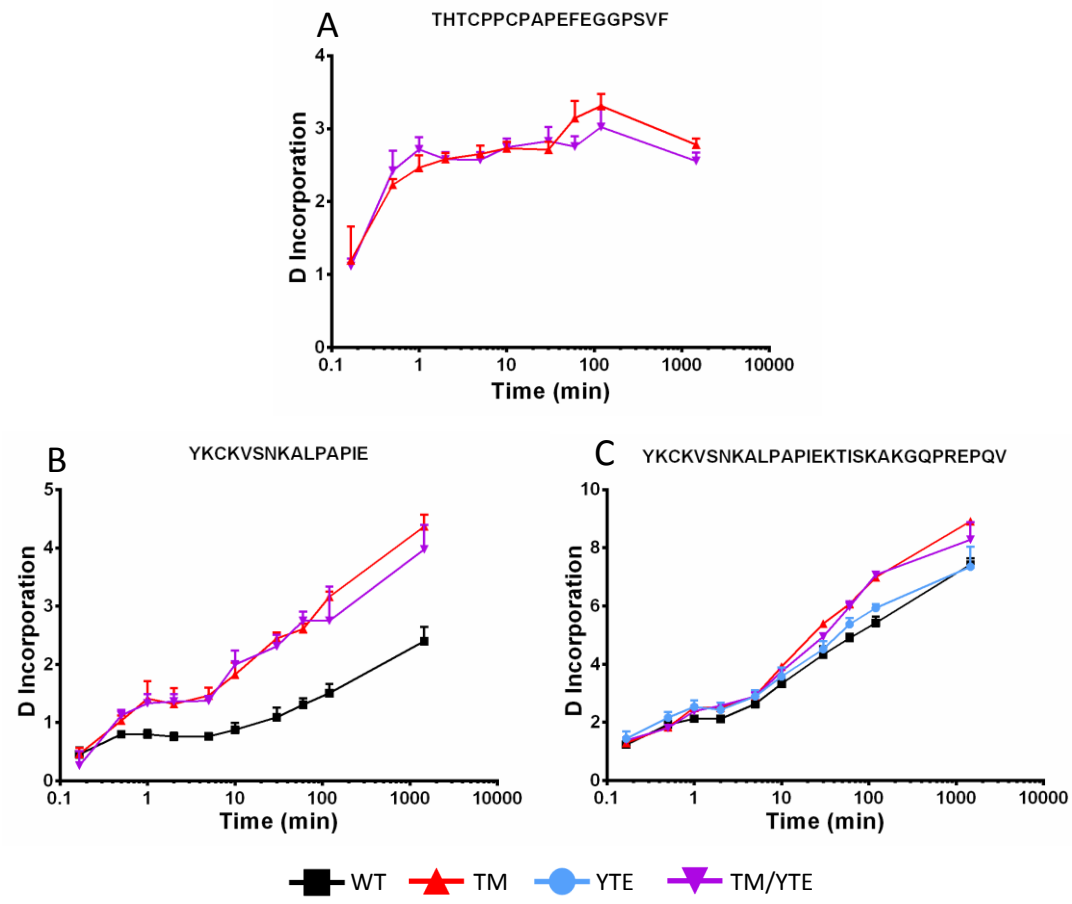


Figure 5-34 – The deuterium incorporation plots for three peptides which span the TM mutation sites. A) A conserved peptide in the TM and TMYTE which covers the L234, and L235 residue, B) and C) are both peptides found which cover the P331. In both B and C residue 13, proline (P), was mutated to serine (S) in Fc-TM and Fc-TMYTE constructs. Error bars show the standard deviation from repeated data sets

Figure 5-34 illustrates similarities between Fc-TM and Fc-TMYTE deuterium uptake in peptides spanning the regions of mutation. This is most clearly seen in Figure 33B, which highlights the differences in Fc-TM and Fc-TMYTE relative to Fc-WT, but can also be seen in the latter time points of Figure 33C where the Fc-TM and Fc-YTE also show increased deuterium uptake relative to Fc-YTE.

A single peptide (ISRTPEVTC) was found that contained residues S254 and T256 from the YTE mutation site. The deuterium uptake is shown in Figure 5-35

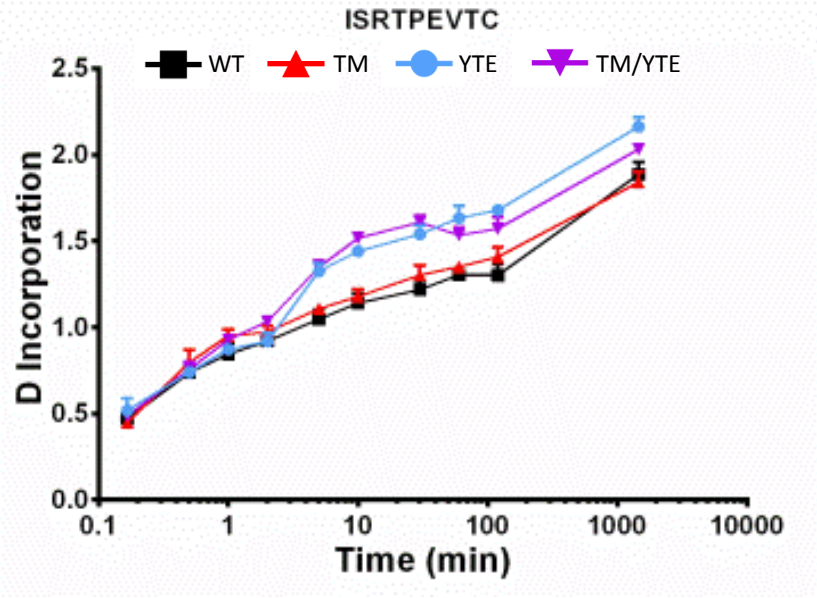


Figure 5-35 – The deuterium incorporation plot for a peptide spanning the YTE mutation site Residues 2, serine (S), and 4, threonine (T) were mutated to threonine (T) and glutamate (E) in Fc-YTE and Fc-TMYTE constructs. Error bars show the standard deviation from repeated data sets

Figure 5-35 illustrates that across the YTE mutation site again the two mutant variants, Fc-YTE and Fc-TMYTE, show a similar deuterium incorporation profile which is different to that observed for Fc-WT and Fc-TM, with the two mutants showing an increase in deuterium uptake.

Figures 4.32-4.34 illustrate potential conformational differences between Fc constructs which contain the amino acid substitutions and those which do not. The substitutions, in all cases, appear to increase the relative uptake of deuterium at the site consistent with an increase in conformational flexibility.

Data presented in Figures 34 and 35 all show a biphasic trend in deuterium uptake; there is an initial increase in deuterium uptake, followed by a plateau and a second increase. Different amino acids, based on the local chemical environment, will exchange at different rates, and the observed trend in deuterium uptake in these figures is suggestive of different exchange rates across the peptides. This suggests possible differences in relative stability of the structural elements across the peptide.

The region between the two sets of mutations, residues 236-251, show an increased degree of deuterium uptake in Figure 5-31, suggesting a possible site of cooperative destabilisation in the Fc-TMYTE protein. The deuterium incorporation plots for the four peptides in this region are shown in Figure 36.

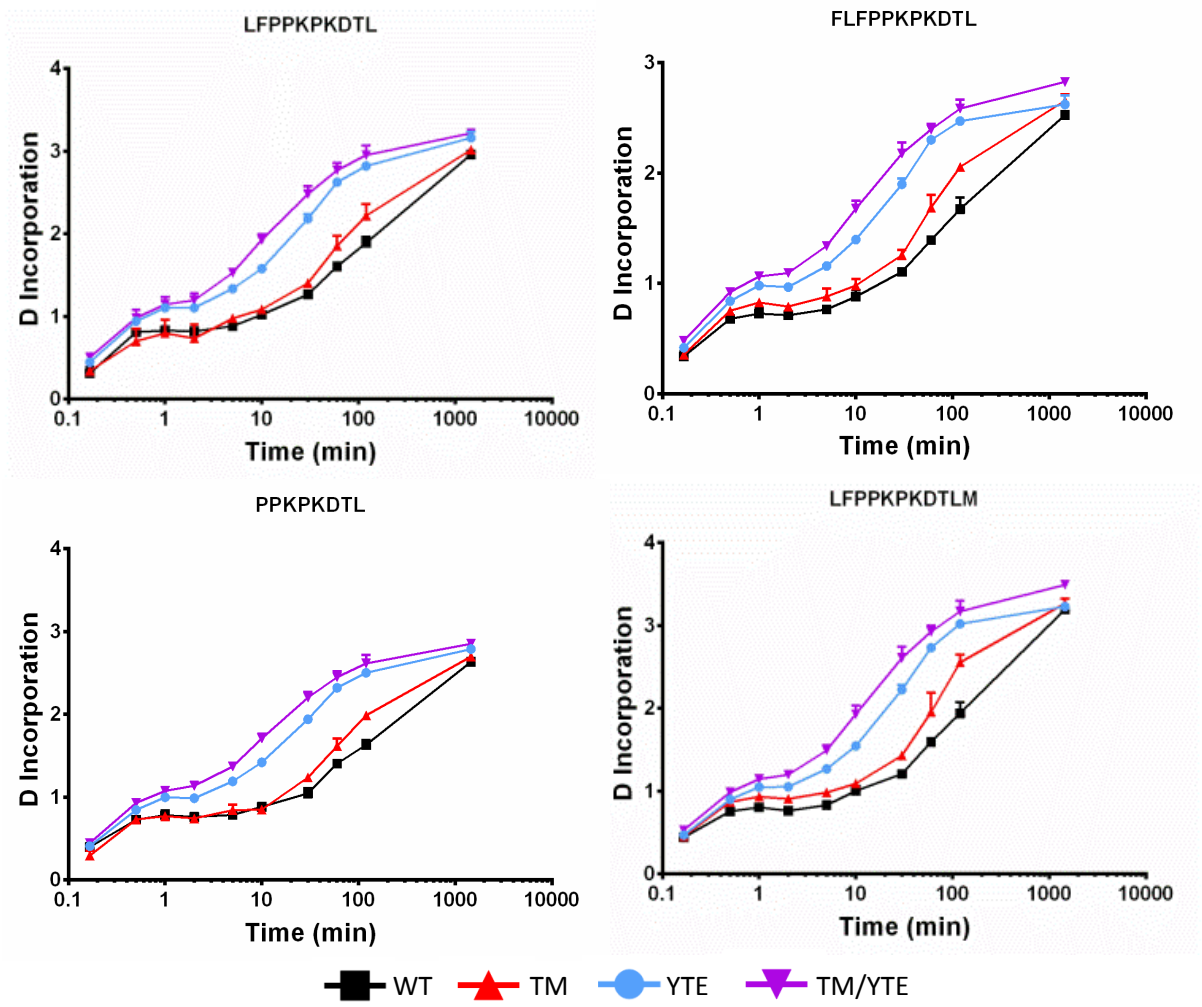
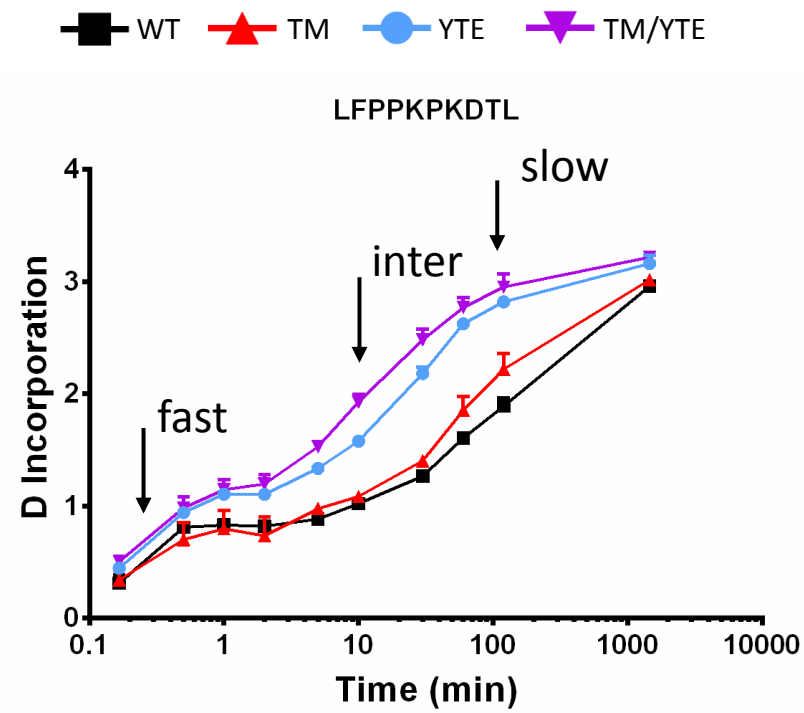


Figure 5-36 – Deuterium uptake plots for the four peptides located at the inner C_H2 interface, between YTE and TM mutation sites. Error bars show the standard deviation from repeated data sets

Each of the four peptides shown in Figure 5-36 show that both Fc-TM and Fc-YTE have increased proton exchange relative to wild type, with Fc-YTE having the greater relative exchange, and thus a potentially more destabilised structure, than Fc-TM. The Fc-TMYTE shows the greatest degree of proton exchange. The degree of proton exchange observed in Fc-TMYTE appears to be greater than the additive effect of individual Fc-TM and Fc-YTE mutations, particularly at intermediate time points (1 minute → 100 minutes) where the combined Fc-TM and Fc-YTE data is less than that observed for Fc-TMYTE. This data suggests that the peptides identified above may contribute to the observed synergistic thermal destabilisation observed in the Fc-TMYTE mutants.

In order attempt to quantify these differences, data from Figure 5-36 was fitted to a number of exponentials. In order to ensure that the simplest model which adequately fitted the data was used, an F-test was applied. Results are shown in Figure 5-37.



	WT	TM	YTE	TM/YTE
k1 (slow)	0.01	0.01	0.03	0.01
k2 (intermediate)	-	-	-	0.09
k3 (fast)	3.55	3.31	3.38	4.27

R^2 all ≥0.99

Figure 5-37 – The rates of deuterium exchange for the fast, intermediate and slow kinetics steps for peptide LFPPKPKDTL

Figure 5-37 indicates that proton exchange in peptide LFPPKPKDTL can be characterised by three distinct rates; fast exchanging amides (k > 3.00), intermediate exchanging amides (k = 0.05- 3.00), and slow exchanging amides (k = < 0.05) (the exchange rates of which are shown in Figure 5-37). Fc-TMYTE data was the only data set, confirmed by F-test, which could be best fitted to three rates of exchange (fast, intermediate and slow) compared to Fc-WT, Fc-TM and Fc-TMYE which were adequately fitted to two rate of exchange. This illustrates that Fc-TMYTE shows not

only differences in proton exchange, but also suggests that the Fc-TMYTE has local unfolding or a reduced degree of structure in this region of the protein.

In order to validate the proposed cooperative destabilisation, Figure 5-38 shows a cooperativity plot, creating by comparing the effects of additive deuterium exchange (Fc-TM + Fc-YTE deuterium uptake) relative to the cooperative deuterium exchange (Fc-TMYTE deuterium uptake. Ω calculated by using the equation ω = ((Fc-TM – Fc-WT) + (Fc-YTE –Fc-WT)) / (Fc-TMYTE – Fc-WT). Deviation from the central line (ω = 1.0) suggests either a greater than additive (ω = 0 - 1.0) or a less than additive (ω = 1.0 – 2.0) effect on deuterium exchange.

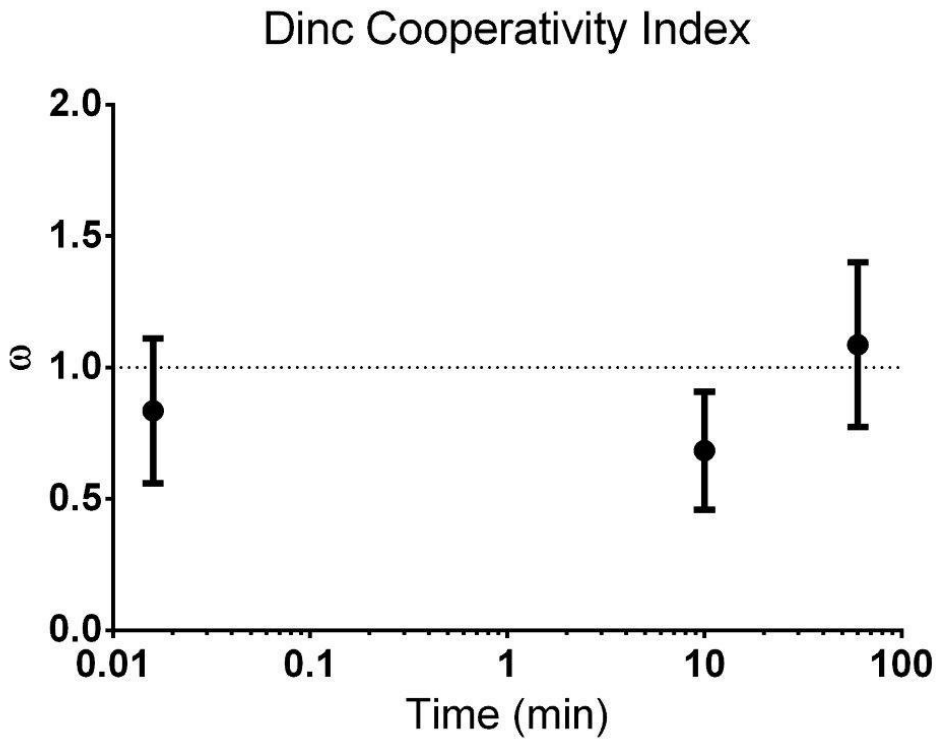


Figure 5-38 – The deuterium incorporation (Dinc) cooperativity index, determined by the ratio of additive (Fc-TM + Fc-YTE) deuterium uptake and cooperative (Fc-TMYTE) deuterium uptake.

Data points are determined by the mid-point of the slow (0.166 minutes), intermediate (10 minutes) and fast (60 minutes) kinetics rates described in Figure 5-37. Figure generated by Dr Phillips.

Figure 5-38 suggests that there is no cooperativity in the Fc-TMYTE construct for the fast and slow exchange rates, since both data points lie within experimental error from the line ω = 1. The intermediate exchange rate shows a deviation from the line ω = 1 greater than error, and falls at approximately 0.7 ω, indicating a greater than cooperative effect on exchange.

5.3 Conclusions

The aim of this study was to rationalise the conformational destabilisation observed by DSC for IgG Fcs engineered with different effector functions/ serum half-life and to identify any other differences in protein structure and/or dynamics using MS based approaches.

Native IM-MS is a powerful tool for the identification of structural changes on a whole protein level. Changes in the charge distribution/ ATD are indicative of changes in the conformation and/or dynamics of a protein.

HDX is a powerful tool for not only looking at global conformational changes, but also identifying local sites within the protein that undergo change. Measuring the rates of exchange on a peptide level provides information on subtle changes to the conformation which cannot be detected using native IM-MS. This approach has in the past been laborious to perform with a high degree of error, but the development of on-line deuteration/ digestion/ analysis systems has helped improve the reproducibility and sensitivity of the technique.

By using a combination of native IM-MS and HDX is has been possible to study different engineered Fc variants in a great degree of detail.

The Fc variants exhibit no observable changes in their global conformation by IMS as a result of the mutations. This suggests that the conformational changes observed are too subtle to be seen on a whole-domain level. It was possible to observe the differences in gas phase stability by studying the arrival time distribution of the different charge states. The IM-MS data supported the DSC observations of a reduction of stability with the introduction of the mutations.

The global pattern of HDX change across the protein also supported the DSC and IMMS data; no regions in the engineered Fcs shows increased protection as a result of mutation indicating the mutations in the Fc-TM, Fc-YTE and Fc-TMYTE destabilised the protein across the molecules.

The deuterium exchange at the site of mutation showed agreement between the Fc-TM and Fc-YTE, and between the Fc-YTE and Fc-TMYTE. This suggests that, at these sites, there is no additional destabilisation as a result of all six mutations.

A site on the inner C_H2 domain surface located between residues 236-251, which is between the two sets of mutations, was shown to have increased deuterium exchange in the Fc-TMYTE compared to the Fc-TM and Fc-YTE. Analysis of the region shows not only an additional exchange rate present in the Fc-TMYE not present in any other construct, but also a quantifiable increase in the exchange greater than the additive effects of the individual Fc-TM and Fc-YTE combined. The combination of the two sets of mutations leads to the synergistic destabilisation of this region which may be responsible for the observed thermal destabilised seen by DSC and IM-MS.

5.4 References

Abès, R. and J.-L. Teillaud (2010). "Impact of Glycosylation on Effector Functions of Therapeutic IgG." Pharmaceuticals **3**(1): 146-157.

Acqua, W. F. D., *et al*. (2002). "Increasing the Affinity of a Human IgG1 for the Neonatal Fc Receptor: Biological Consequences." The Journal of Immunology **169**(9): 5171-5180.

Adamczyk, M., *et al*. (2000). "Papain digestion of different mouse IgG subclasses as studied by electrospray mass spectrometry." Journal of Immunological Methods **237**(1–2): 95-104.

Alegre, M.-L., *et al*. (1994). "A NON-ACTIVATING "HUMANIZED" ANTI-CD3 MONOCLONAL ANTIBODY RETAINS IMMUNOSUPPRESSIVE PROPERTIES IN VIVO." Transplantation **57**(11): 1537-1543.

Armour, K. L., *et al*. (2003). "Differential binding to human FcγRIIa and FcγRIIb receptors by human IgG wildtype and mutant antibodies." Molecular Immunology **40**(9): 585-593.

Ashton, D. S., *et al*. (1995). "Mass Spectrometry of the Humanized Monoclonal Antibody Campath 1H." Analytical Chemistry **67**(5): 835-842.

Augener, W., *et al*. (1971). "The reaction of monomeric and aggregated immunoglobulins with Cl." Immunochemistry **8**(11): 1011-1020.

Bagal, D., *et al*. (2010). "Resolving Disulfide Structural Isoforms of IgG2 Monoclonal Antibodies by Ion Mobility Mass Spectrometry." Analytical Chemistry **82**(16): 6751-6755.

Bell, L. and R. P. Rother (2012). Treatment of paroxysmal nocturnal hemoglobinuria patients by an inhibitor of complement, Google Patents.

Benesch, J. L. P., *et al*. (2003). "Thermal Dissociation of Multimeric Protein Complexes by Using Nanoelectrospray Mass Spectrometry." Analytical Chemistry **75**(10): 2208-2214.

Bolt, S., *et al*. (1993). "The generation of a humanized, non-mitogenic CD3 monoclonal antibody which retains in vitro immunosuppressive properties." European Journal of Immunology **23**(2): 403-411.

Brandt, J. P., *et al*. (2010). "Construction, MD Simulation, and Hydrodynamic Validation of an All-Atom Model of a Monoclonal IgG Antibody." Biophysical Journal **99**(3): 905-913.

Buss, N. A. H., Simon J; McFarlane, Mary; Shenton, Jacintha M; de Haan, Lolke (2012). "Monoclonal antibody therapeutics: history and future." Current Opinion in Pharmacology **12**(5): 615-622.

Canfield, S. M. and S. L. Morrison (1991). "The binding affinity of human IgG for its high affinity Fc receptor is determined by multiple amino acids in the CH2 domain and is modulated by the hinge region." J Exp Med **173**(6): 1483-1491.

Chappel, M. S., *et al*. (1991). "Identification of the Fc gamma receptor class I binding site in human IgG through the use of recombinant IgG1/IgG2 hybrid and point-mutated antibodies." Proceedings of the National Academy of Sciences **88**(20): 9036-9040.

Chen, G. W., Bethanne M.; Goodenough, Angela K.; Wei, Hui; Wang-Iverson, David B.; Tymiak, Adrienne A. (2011). "Characterization of protein therapeutics by mass spectrometry: recent developments and future directions." Drug Discovery Today **16**(1-2): 58-64.

Cole, M. S., *et al*. (1999). "HuM291, A HUMANIZED ANTI-CD3 ANTIBODY, IS IMMUNOSUPPRESSIVE TO T CELLS WHILE EXHIBITING REDUCED MITOGENICITY IN VITRO1." Transplantation **68**(4): 563-571.

Cooper, N. R. (1985). The Classical Complement Pathway: Activation and Regulation of the First Complement Component. Advances in Immunology. J. D. Frank, Academic Press. **Volume 37:** 151-216.

Daëron, M. (1997). "Fc RECEPTOR BIOLOGY." Annual Review of Immunology **15**(1): 203-234.

Dall'Acqua, W. F., *et al*. (2006). "Modulation of the Effector Functions of a Human IgG1 through Engineering of Its Hinge Region." The Journal of Immunology **177**(2): 1129-1138.

Davies, D. R. and H. Metzger (1983). "Structural basis of antibody function." Annu Rev Immunol **1**: 87-117.

Downard, K. M. (2000). "Contributions of mass spectrometry to structural immunology." Journal of Mass Spectrometry **35**(4): 493-503.

Duhamel, R. C., *et al*. (1979). "pH gradient elution of human IgG1, IgG2 and IgG4 from protein A-Sepharose." Journal of Immunological Methods **31**(3–4): 211-217.

Fleischman, J. B., *et al*. (1963). "THE ARRANGEMENT OF THE PEPTIDE CHAINS IN GAMMA-GLOBULIN." Biochem J **88**: 220-228.

Gessner, J. E., *et al*. (1998). "The IgG Fc receptor family." Annals of Hematology **76**(6): 231-248.

Ghetie, V. and E. S. Ward (1997). "FcRn: the MHC class I-related receptor that is more than an IgG transporter." Immunology Today **18**(12): 592-598.

Ghetie, V. and E. S. Ward (2000). "Multiple roles for the major histocompatibility complex class I- related receptor FcRn." Annu Rev Immunol **18**: 739-766.

Green, L. L., *et al*. (1994). "Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs." Nat Genet **7**(1): 13-21.

Harris, R. J., *et al*. (2004). "Commercial manufacturing scale formulation and analytical characterization of therapeutic recombinant antibodies." Drug Development Research **61**(3): 137-154.

Heck, A. J. R. and R. H. H. van den Heuvel (2004). "Investigation of intact protein complexes by mass spectrometry." Mass Spectrometry Reviews **23**(5): 368-389.

Hinton, P. R., *et al*. (2006). "An engineered human IgG1 antibody with longer serum half-life." J Immunol **176**(1): 346-356.

Hober, S., *et al*. (2007). "Protein A chromatography for antibody purification." Journal of Chromatography B **848**(1): 40-47.

Horton, H. M., *et al*. (2008). "Potent in vitro and in vivo activity of an Fc-engineered anti-CD19 monoclonal antibody against lymphoma and leukemia." Cancer Res **68**(19): 8049-8057.

Hutchins, J. T., *et al*. (1995). "Improved biodistribution, tumor targeting, and reduced immunogenicity in mice with a gamma 4 variant of Campath-1H." Proceedings of the National Academy of Sciences **92**(26): 11980-11984.

Idusogie, E. E., *et al*. (2000). "Mapping of the C1q binding site on rituxan, a chimeric antibody with a human IgG1 Fc." J Immunol **164**(8): 4178-4184.

Janeway, C. (2005). Immunobiology: the immune system in health and disease, Garland Science.

Jefferis, R., *et al*. (1990). "A comparative study of the N-linked oligosaccharide structures of human IgG subclass proteins." Biochemical Journal **268**(3): 529-520.

Jekel, P. A., *et al*. (1983). "Use of endoproteinase Lys-C from Lysobacter enzymogenes in protein sequence analysis." Analytical Biochemistry **134**(2): 347-354.

Johnstone, A. P. and L. E. Mole (1977). "Sequence studies on the heavy chain of rabbit immunoglobulin A of different alpha-locus allotypes." Biochem J **167**(1): 255-267.

Jones, P. T., *et al*. (1986). "Replacing the complementarity-determining regions in a human antibody with those from a mouse." Nature **321**(6069): 522-525.

Katta, V. and B. T. Chait (1991). "Conformational-changes in proteins probed by hydrogen-exchange electrospray-ionization mass-spectrometry." Rapid Communications in Mass Spectrometry **5**(4): 214-217.

Kleemann, G. R., *et al*. (2008). "Characterization of IgG1 Immunoglobulins and Peptide−Fc Fusion Proteins by Limited Proteolysis in Conjunction with LC−MS." Analytical Chemistry **80**(6): 2001-2009.

Kohler, G. and C. Milstein (1975). "Continuous cultures of fused cells secreting antibody of predefined specificity." Nature **256**(5517): 495-497.

Konijnenberg, A., *et al*. (2013). "Native ion mobility-mass spectrometry and related methods in structural biology." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(6): 1239-1256.

Krapp, S., *et al*. (2003). "Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity." Journal of Molecular Biology **325**(5): 979-989.

Kukrer, B., *et al*. (2010). "Mass Spectrometric Analysis of Intact Human Monoclonal Antibody Aggregates Fractionated by Size-Exclusion Chromatography." Pharmaceutical Research **27**(10): 2197-2204.

Larrick, J. W. (2012). "Conference scene: progress with promising human antibodies." Immunotherapy **4**(3): 257-261.

Lau, H., *et al*. (2010). "Investigation of degradation processes in IgG1 monoclonal antibodies by limited proteolysis coupled with weak cation-exchange HPLC." J Chromatogr B Analyt Technol Biomed Life Sci **878**(11-12): 868-876.

Lazar, G. A., *et al*. (2006). "Engineered antibody Fc variants with enhanced effector function." Proceedings of the National Academy of Sciences of the United States of America **103**(11): 4005-4010.

Lewis, D. A., *et al*. (1994). "Characterization of Humanized Anti-TAC, an Antibody Directed Against the Interleukin 2 Receptor, Using Electrospray Ionization Mass Spectrometry by Direct Infusion, LC/MS, and MS/MS." Analytical Chemistry **66**(5): 585-595.

Lindsley, C. W. (2013). "The Top Prescription Drugs of 2012 Globally: Biologics Dominate, But Small Molecule CNS Drugs Hold on to Top Spots." ACS Chemical Neuroscience **4**(6): 905-907.

Liu, H. C., *et al*. (2010). "Domain-level stability of an antibody monitored by reduction, differential alkylation, and mass spectrometry analysis." Analytical Biochemistry **400**(2): 244-250.

Liu, Y. D., *et al*. (2009). "Human antibody Fc deamidation in vivo." Biologicals **37**(5): 313-322.

Lonberg, N., *et al*. (1994). "Antigen-specific human antibodies from mice comprising four distinct genetic modifications." Nature **368**(6474): 856-859.

Masuda, K., *et al*. (2000). "Pairing of oligosaccharides in the Fc region of immunoglobulin G." FEBS Letters **473**(3): 349-357.

Morrison, S. L., *et al*. (1984). "Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains." Proceedings of the National Academy of Sciences **81**(21): 6851-6855.

Mostov, K. E. and N. E. Simister (1985). "Transcytosis." Cell **43**(2): 389-390.

Müller-Eberhard, H. J. and M. A. Calcott (1966). "Interaction between C'1q and γG-globulin." Immunochemistry **3**(6): 500.

Nobelprize.org (1972). "The Nobel Prize in Phisiology and Medicine 1972." Retrieved 25/3/15, 2015, from http://www.nobelprize.org/nobel_prizes/medicine/laureates/1972/index.htm.

Nobelprize.org (1984). "The Nobel Prize in Physiology or Medicine 1984." Retrieved 22 Aug 2014, 2014.

O'Donnell, I. J., *et al*. (1970). "The disulphide bonds of the heavy chain of rabbit immunoglobulin G." Biochem J **116**(2): 261-268.

Ober, E. S. W. (2012). Immunoglobulin molecules with improved characteristics, Google Patents.

Ober, R. J., *et al*. (2001). "Differences in promiscuity for antibody-FcRn interactions across species: implications for therapeutic antibodies." Int Immunol **13**(12): 1551-1559.

Oganesyan, V., *et al*. (2009). "Structural characterization of a human Fc fragment engineered for extended serum half-life." Molecular Immunology **46**(8–9): 1750-1755.

Oganesyan, V., *et al*. (2008). "Structural characterization of a human Fc fragment engineered for lack of effector functions." Acta Crystallographica Section D **64**(6): 700-704.

Oudin, J. (1956). "[The allotype of certain blood protein antigens]." C R Hebd Seances Acad Sci **242**(21): 2606-2608.

Pacholarz, K. J., *et al*. (2014). "Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling." Angewandte Chemie International Edition **53**(30): 7765-7769.

Porter, R. R. (1959). "The hydrolysis of rabbit y-globulin and antibodies with crystalline papain." Biochem J **73**: 119-126.

Prabakaran, P., *et al*. (2008). "Structure of an isolated unglycosylated antibody CH2 domain." Acta Crystallographica Section D **64**(10): 1062-1067.

R. L. Hill, *et al*. (1966). Studies on the Amino Acid Sequence of Heavy Chains from Rabbit Immunoglobulin G. Royal Society of London. Series B, Biological Sciences, London.

Raghavan, M. and P. J. Bjorkman (1996). "Fc receptors and their interactions with immunoglobulins." Annu Rev Cell Dev Biol **12**: 181-220.

Reddy, M. P., *et al*. (2000). "Elimination of Fc receptor-dependent effector functions of a modified IgG4 monoclonal antibody to human CD4." J Immunol **164**(4): 1925-1933.

Reid, C. Q., *et al*. (2010). "Rapid Whole Monoclonal Antibody Analysis by Mass Spectrometry: An Ultra Scale-Down Study of the Effect of Harvesting by Centrifugation on the Post-Translational Modification Profile." Biotechnology and Bioengineering **107**(1): 85-95.

Richards, J. O., *et al*. (2008). "Optimization of antibody binding to FcgammaRIIa enhances macrophage phagocytosis of tumor cells." Mol Cancer Ther **7**(8): 2517-2527.

Riechmann, L., *et al*. (1988). "Reshaping human antibodies for therapy." Nature **332**(6162): 323-327.

Riedel, S. (2005). "Edward Jenner and the history of smallpox and vaccination." Proc (Bayl Univ Med Cent) **18**(1): 21-25.

Rosati, S., *et al*. (2012). "Exploring an Orbitrap Analyzer for the Characterization of Intact Antibodies by Native Mass Spectrometry." Angewandte Chemie International Edition **51**(52): 12992-12996.

Rose, R. J., *et al*. (2013). "Mutation of Y407 in the CH3 domain dramatically alters glycosylation and structure of human IgG." MAbs **5**(2): 219-228.

Rother, R. P., *et al*. (2007). "Discovery and development of the complement inhibitor eculizumab for the treatment of paroxysmal nocturnal hemoglobinuria." Nat Biotech **25**(11): 1256-1264.

Saphire, E. O., *et al*. (2002). "Contrasting IgG structures reveal extreme asymmetry and flexibility." Journal of Molecular Biology **319**(1): 9-18.

Sazinsky, S. L., *et al*. (2008). "Aglycosylated immunoglobulin G1 variants productively engage activating Fc receptors." Proceedings of the National Academy of Sciences **105**(51): 20167-20172.

Scarff, C. A., *et al*. (2008). "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 Communications in Mass Spectrometry **22**(20): 3297-3304.

Schumaker, V. N., *et al*. (1976). "Ultracentrifuge studies of the binding of IgG of different subclasses to the Clq subunit of the first component of complement." Biochemistry **15**(23): 5175-5181.

Seiler, F. R., *et al*. (1985). "Monoclonal Antibodies: Their Chemistry, Functions, and Possible Uses." Angewandte Chemie International Edition in English **24**(3): 139-160.

Sheeley, D. M., *et al*. (1997). "Characterization of Monoclonal Antibody Glycosylation: Comparison of Expression Systems and Identification of Terminal [alpha]-Linked Galactose." Analytical Biochemistry **247**(1): 102-110.

Shen, L., *et al*. (1987). "Polymorphonuclear leukocyte function triggered through the high affinity Fc receptor for monomeric IgG." J Immunol **139**(2): 534-538.

Shields, R. L., *et al*. (2001). "High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R." J Biol Chem **276**(9): 6591-6604.

Simister, N. E., *et al*. (1996). "An IgG-transporting Fc receptor expressed in the syncytiotrophoblast of human placenta." European Journal of Immunology **26**(7): 1527-1531.

Stavenhagen, J. B., *et al*. (2007). "Fc optimization of therapeutic antibodies enhances their ability to kill tumor cells in vitro and controls tumor expansion in vivo via low-affinity activating Fcgamma receptors." Cancer Res **67**(18): 8882-8890.

Stern, M. and R. Herrmann (2005). "Overview of monoclonal antibodies in cancer therapy: present and promise." Critical Reviews in Oncology/Hematology **54**(1): 11-29.

Strohl, W. R. (2009). "Optimization of Fc-mediated effector functions of monoclonal antibodies." Current Opinion in Biotechnology **20**(6): 685-691.

Strohl, W. R. (2010). Non-immunostimulatory antibody and compositions containing the same, Google Patents.

Tao, M. H., *et al*. (1993). "Structural features of human immunoglobulin G that determine isotype-specific differences in complement activation." J Exp Med **178**(2): 661-667.

Thalassinos, K., *et al*. (2009). "Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry." Analytical Chemistry **81**(1): 248-254.

Thompson, N. J., *et al*. (2014). "Performing native mass spectrometry analysis on therapeutic antibodies." Methods **65**(1): 11-17.

Tito, M. A., *et al*. (2001). "Probing molecular interactions in intact antibody: Antigen complexes, an electrospray time-of-flight mass spectrometry approach." Biophysical Journal **81**(6): 3503-3509.

Vaughan, T. J., *et al*. (1998). "Human antibodies by design." Nat Biotech **16**(6): 535-539.

Verhoeyen, M., *et al*. (1988). "Reshaping Human Antibodies: Grafting an Antilysozyme Activity." Science **239**(4847): 1534-1536.

Von Behring, E. and S. Kitasato (1890). "On the acquisition of immunity against diptheria and tetanus in animals." Dtsch Med Wochensch **16**: 1145-1148.

von Pawel-Rammingen, U., *et al*. (2002). "IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G." The EMBO Journal **21**(7): 1607-1615.

Wang, L. T., *et al*. (2005). "Structural characterization of a recombinant monoclonal antibody by electrospray time-of-flight mass spectrometry." Pharmaceutical Research **22**(8): 1338-1349.

Wang, W., *et al*. (2007). "Antibody structure, instability, and formulation." Journal of Pharmaceutical Sciences **96**(1): 1-26.

Ward ES and G. V. (1995). "The effector functions of immunoglobulins: implications for therapy. ." Ther Immunol **2**(2): 77-94.

Williams, J. P. and J. H. Scrivens (2008). "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 Communications in Mass Spectrometry **22**(2): 187-196.

Winter, G., *et al*. (1994). "Making Antibodies by Phage Display Technology." Annual Review of Immunology **12**(1): 433-455.

Wright, A. and S. L. Morrison (1997). "Effect of glycosylation on antibody function: implications for genetic engineering." Trends in Biotechnology **15**(1): 26-32.

Wyttenbach, T. and M. T. Bowers (2011). "Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility–Mass Spectrometry Environment." The Journal of Physical Chemistry B **115**(42): 12266-12275.

Xu, D., *et al*. (2000). "In Vitro Characterization of Five Humanized OKT3 Effector Function Variant Antibodies." Cellular Immunology **200**(1): 16-26.

Yeung, Y. A., *et al*. (2009). "Engineering human IgG1 affinity to human neonatal Fc receptor: impact of affinity improvement on pharmacokinetics in primates." J Immunol **182**(12): 7663-7671.

Yu, L., *et al*. (2006). "Identification of N-terminal modification for recombinant monoclonal antibody light chain using partial reduction and quadrupole time-of-flight mass spectrometry." Rapid Communications in Mass Spectrometry **20**(24): 3674-3680.

Zalevsky, J., *et al*. (2009). The impact of Fc engineering on an anti-CD19 antibody: increased Fcγ receptor affinity enhances B-cell clearing in nonhuman primates.

Zhang, H., *et al*. (2014). "Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies." FEBS Letters **588**(2): 308-317.

Zhang, Z. and B. Shah (2007). "Characterization of Variable Regions of Monoclonal Antibodies by Top-Down Mass Spectrometry." Analytical Chemistry **79**(15): 5723-5729.

null

Zhang, Z. Q., *et al*. (2009). "MASS SPECTROMETRY FOR STRUCTURAL CHARACTERIZATION OF THERAPEUTIC ANTIBODIES." Mass Spectrometry Reviews **28**(1): 147-176.

Ziccardi, R. (1983). "The first component of human complement (C1): Activation and control." Springer Seminars in Immunopathology **6**(2-3): 213-230.

Chapter 6 : Conclusions

5.1 Conclusions

The use of mass spectrometry in protein structural biology is becoming increasingly common. Mass spectrometry offers a number of improvements in speed and sensitivity compared to alternative structural techniques. This research made use of mass spectrometry to investigate the structural difference in three different biological systems.

5.1.1 Prolyl oligopeptidase and dipeptidyl peptidase IV

There were a number of theories regarding mechanism of substrate entry to the activate site of PREP. A number of bacterial proteins had been crystalised in an open, domain separated conformation and it was proposed that whole domain separation could occur in the mammalian protein. Using shape selective mass spectrometry it was possible to observe only a single conformation of PREP, suggesting that more subtle conformational changes were occurring to allow substrate entry, which is consistent with MD simulations and electron microscopy data. More extended conformations of PREP were observed at higher charge states, and a comparison between the estimated CCS values of these both open and closed conformations agreed well with those calculated from X-ray crystal structures, suggesting that the open conformation on PREP may represent a non-physiological conformation.

DPPIV, an enzyme in the same family, was also studied using IM-MS. DPP IV is larger than PREP, and it was proposed that small interdomain rearrangements allowed substrate entry. The observations presented in this study were consistent with this theory, as only a single conformation of DPP IV could be observed. DPP IV also showed no evidence of multiple conformations at higher charge states, consistent with the protein being unable undergo domain separation as has been seen in PREP.

5.1.2 Haemoglobin

Data published by Scarff *et al* shortly after the release of the first generation ion mobility instrument was revisited, with a number of experimental changes made in light of recent advances in native MS and IM-MS. The data indicates that, at higher source temperatures, minor differences in the conformation of the protein appear larger than they do at lower temperatures by virtue of a worsening of the strength of

the CCS data calibration curve. Higher source pressures, such as those above 5 mBar, were shown to dramatically affect desolvation. As with source temperature, higher pressures were shown to have a negative impact on the CCS data calibration curve.

Having re-optimised the instrument settings, HbA and HbS CCS estimates were obtained experimentally and compared with those calculated from X-ray crystal data. The values obtained in this study showed an improved agreement with the computational values, illustrating the strength and accuracy that can be obtained in when using native protein standards for CCS estimates, compared with those obtained using denatured proteins. The PSA method, a new computational method developed by the Bowers group, was also used for CCS estimates, and compared well with the TM method. This serves to highlight the strength of the PSA method as a fast and accurate alternative for generating theoretical CCS values to the computationally demanding TM method.

HbA and HbS CCS values were then compared to one another, and to a third Hb variant: HbC. It was found that HbA and HbS have a far more similar CCS that first observed, whilst HbC had a significantly smaller CCS value, suggesting a more compact conformation. With increasing charge, all three proteins showed an increase in CCS, consistent with the protein unfolding, with HbS and HbC adopting a more extended conformation at lower charge states than HbA. This was indicative of stability differences between the proteins, and was further studied by collisionally activating the protein.

The results of the activation showed that all three Hb variants adopt that same extended conformations, but the intensities of the extended conformations varied, with HbA proving to be the most stable.

5.1.3 Fc's engineered for therapeutic antibodies

Monoclonal antibody therapeutics represent a large portion of the new generation of therapeutic agents. Whilst mass spectrometry is used routinely for the characterisation of glycosylation and sample identity, it is becoming more commonly used to investigate higher order structure. In this study, a combination of IM-MS and HDX has been used to characterise the Fc domain, responsible for host cell receptor binding, in four differently engineered Fcs.

The TM mutant, engineered for decreased complement recruitment, and the YTE mutant, engineered for increased serum half-life, were both found to have diminished C_H2 thermal stability by DSC. A TMYTE double mutant was observed to have a further decreased thermal stability in the same domain, and it was believed that the combination of mutants was leading to a synergistic destabilisation greater than the sum of the two sets of mutations.

IM-MS was used to study the global conformation, whilst HDX was used to study the local conformation. The IM-MS results indicated that there were no global conformational changes, nor were there any apparent changes to the global protein dynamics. Given the inclusion of the hinge domain in the Fc samples, a loose stretch of 14 amino acids capable of free motion in solution, it is possible that any changes in conformation were concealed by natural hinge motion.

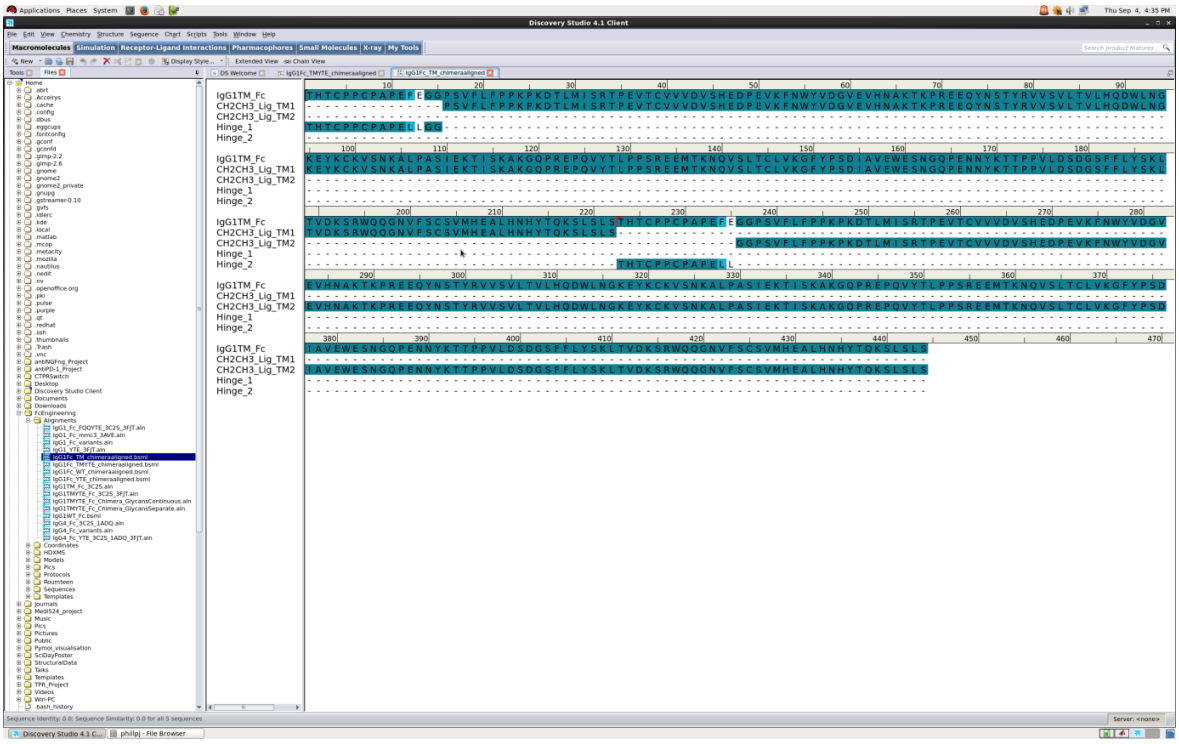
The HDX data showed a number of sites across the Fc domain with increased deuterium uptake, consistent with an increase in the dynamics of the protein. Data at the sites of mutation agreed between the single and double mutants, indicating that mutations at one site did not interfere with the other. HDX also revealed a small region of the Fc domain located between the TM and YTE mutation sites, in the inner C_H2 domain face, which showed a greater than additive degree of deuterium uptake in Fc-TMYTE when compared to Fc-TM and Fc-YTE. Further analysis of this site suggested that there is a different regime of exchange kinetics, consistent with a change in the conformational dynamics in Fc-TMYTE compared to Fc-TM or Fc-YTE. Given the location of the site, it seems possible that this region is responsible for the observed differences in thermal stability.

5.1.4 Final remarks

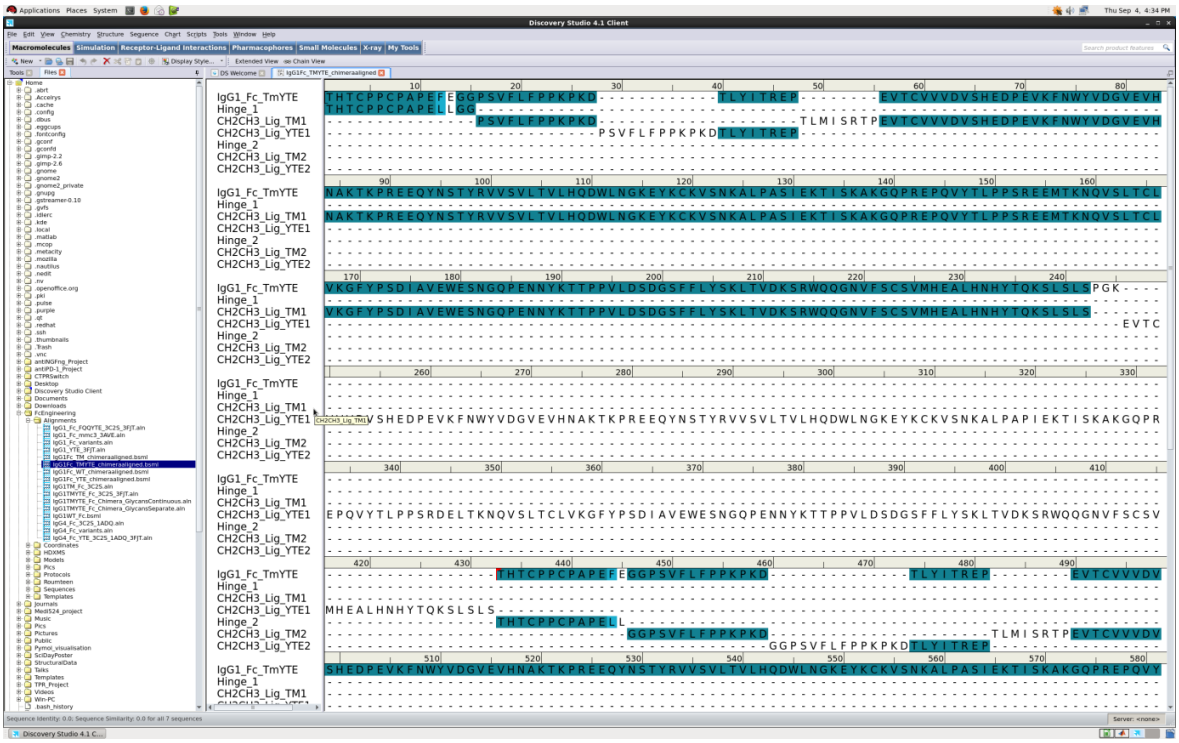
This work has shown the strength of mass spectrometry as a tool for probing subtle, and larger, changes in protein conformation. Using a combination of approaches it has been used to study protein conformation, stability, and dynamics, and has successfully highlighted the differences between proteins differing in only a few amino acids.

Appendix

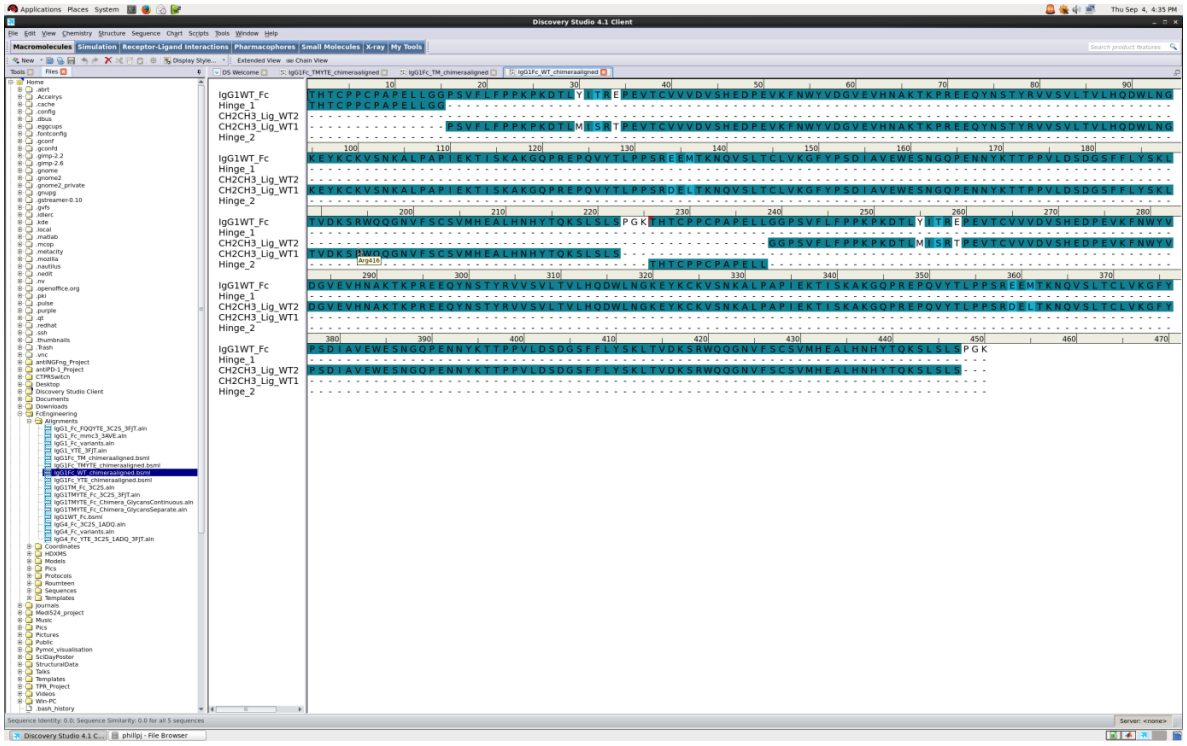
Appendix A



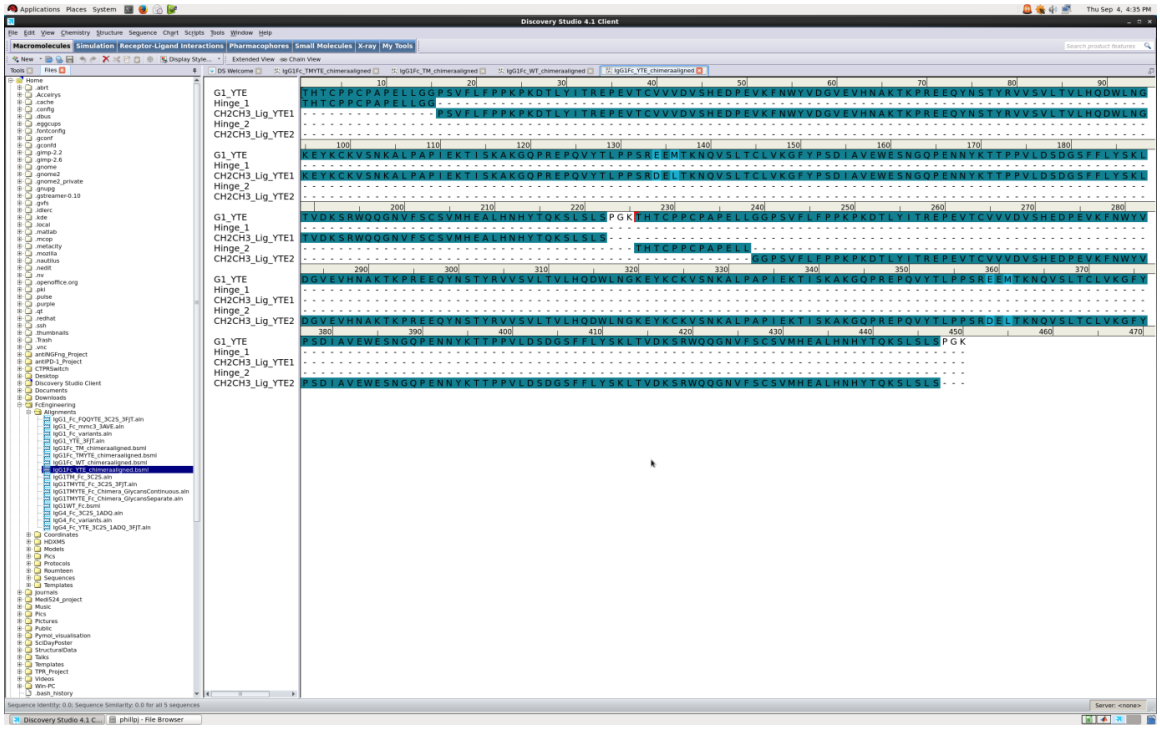
Appendix A-1: The sequence alignment for the homology model of the Fc-TM construct



Appendix A-2: The sequence alignment for the homology model of the Fc-TMYTE construct



Appendix A-3: The sequence alignment for the homology model of the Fc-WT construct



Appendix A-4: The sequence alignment for the homology model of the Fc-YTE construct

Appendix B

In addition to the research described in detail in Chapters 1-5, research has also been carried out on the use of mass spectrometry approaches to characterise glycosylation patterns in proteins. Glycosylation is the enzymatic attachment of carbohydrates to specific sites on a protein. More than 50% of all proteins are thought to be modified in this way. Glycosylation is a key process in the cell and can control the promotion of correct protein folding, the recognition of cell types by the immune system, and primitive protection against viruses and bacteria. Glycans can form large chains and antennary structures, which can give rise to a large amount of heterogeneity in a sample, often due to single sites of glycosylation playing host to a series of different glycans. The characterisation of oligosaccharides presents a significant analytical challenge since it requires the determination of ring size, isomer, linkage position, sequence and anomeric configuration of linkage for each species in what is often a complex mixture of carbohydrate molecules. Existing approaches for characterising glycans are time consuming and labour intensive. Rapid methods are required not only for glycoproteomic research but also for pharmaceutical studies. These involve the recombinant synthesis of glycoproteins using, for example, insect cell lines. Measurement is required for synthesis optimisation, quality control and effect measurement. The Warwick group has enjoyed a successful and productive collaboration with Professor David Harvey (a visiting Professor in the School of Life Sciences and a major international expert in the use of mass spectrometry to characterise glycans) for a number of years. The collaboration has resulted in a number of publications and the approach, based on ion mobility separation and development of improved enrichment and ionisation approaches has been shown to have significant advantages.

I have, during the period of my post graduate research worked closely with Professor Harvey and Professor Scrivens in support of this research carrying out experimental work and providing data interpretation in support of a number of projects. This work has resulted in four publications to date of which I am a co-author. Since these projects, although valuable and interesting, did not form part of my core research program I have not described the work in detail in this thesis. I do however attach copies of the publications in appendix B as a record of the activities.