Ion Mobility & Mass Spectrometric Studies of Macromolecules Required for Organism Viability

A Thesis Presented to University College London in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Chemistry

Richard Alexander Kerr

University College London

August 2013

Declaration

I, Richard Alexander Kerr confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis

Richard Alexander Kerr

August 2013

Abstract

Studies of gas phase protein complexes using mass spectrometry and ion-mobility mass spectrometry are becoming increasingly commonplace in the field of structural biology. These studies apply the combined speed, mass sensitivity and structural resolution of a single instrumental method, that of mass spectrometry. Work presented here has used a range of gas, and solution phase methods. These methods have made it possible to investigate the oligomers and structural conformations of three proteins required, within their respective organisms, to ensure viability.

Mutations of the human serine protease inhibitor, α_1-antitrypsin, are known to promote polymerogenic intermediates under biologically relevant conditions. Using ion-mobility mass spectrometry we have characterised the structure and stability of the K154N slow polymerisation mutant. The results obtained have shown that this mutant populates an increased stability structural intermediate upon incubation at biologically relevant temperatures.

Saccharomyces cerevisiae Sgt1 dimers mediate binding between Hsp90 and Skp1, to initiate chromosome separation. Mutations of the Sgt1 dimerization domain are known to inhibit Sgt1: Skp1 binding, arresting the cell cycle at the G2/M interface. Work here has shown that this dimer is stabilised by an Ascomycota specific structural loop within the dimerisation domain, with potential contributions from other domains. Using tandem mass spectrometry, we have shown that Sgt1 dimers do not represent the structural minimum for Skp1 binding.

The *Escherichia coli* DNA binding protein, CbpA, promotes chromosome compaction that arrests the cell cycle during the stationary phase, and phosphate starvation conditions, producing structural aggregates exceeding 60 nm. The organisation of both CbpA and DNA within these aggregates remains currently unresolved therefore, mass spectrometry combined with ion-mobility mass spectrometry has been applied to resolve these interactions. Although unsuccessful in observing gas phase CbpA-DNA oligomers using a variety of conditions and methods, investigations have produced ion-mobility constraints for computational modelling of the biologically relevant CbpA dimer.

Contents

---

5. Analysis of CbpA's Structure and DNA Binding Interactions

List of Figures

List of Tables

Abbreviations

A

AAT	α_1-Antitrypsin
ADH	Alcohol dehydrogenase
AFM	Atomic Force Microscopy
AmAc	Ammonium Acetate ($C_2H_7NO_2$), PubChem CID; 517165
AUC	Area Under Curve

B

B100	B form, 100 bp dsDNA
bp	DNA base pair length
BSA	Bovine serum albumin

C

CCS	Collision Cross Section ($Å^2$)
CD	Circular Dichromism
cDNA	complementary DNA
CDE	Centromere Determining Element
CID	Collision Induced Dissociation (Synonym; CAD - Collisionally Activated Dissociation)
CIU	Collision Induced Unfolding (Synonym; CAU - Collisionally Activated Unfolding)
CSD	Charge State Distribution
CTD	C-Terminal Domain
CTPR	Consensus Tetratricopeptide Repeat

D

Da	Dalton (Value of 1 is equal to $1/12^{th}$ the mass of a ^{12}C atom)
DNA	Deoxyribonucleic acid

dsDNA	double stranded DNA

E

Eq.	Equation
ESI	Electrospray Ionisation (Synonym; Ion Spray)
EHSS	Exact Hard Sphere Scattering
EMSA	Electrophoretic Mobility Shift Assay

F

Fig.	Figure
FWHM	Full Width at Half (50 %) Maximum (Synonym; FWHH - Full Width at Half (50 %) Height)

G

GC	Guanine/Cytosine

H

HDMS	High Definition Mass Spectrometer (used in context of the Waters Synapt)
His-tag(ged)	Purification tag comprising 6 sequential C-terminal histidine (His) amino acids required for nickel column separation
Hsp#	Heat shock protein, classed in terms of # in kDa mass

I

IM	Ion-Mobility
IM-MS	Ion-Mobility Mass Spectrometry

K

K_a	Association constant
K_d	Dissociation constant
kDa	Kilo Dalton ($1x10^3$ Da)

M

M	Molar concentration, measured as moles per litre
m	Mass (measured in Da)
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MCP	Microchannel Plate
MDa	Mega Dalton ($1x10^6$ Da)
mg/ml	milligrams per millilitre
MIKES	Mass-analysed Ion Kinetic Energy Spectrometry
mM	millimolar ($1x10^{-3}$ M)
ms	milliseconds ($1x10^{-3}$ second)
m/s	meters per second
MSD	Mean Standard Deviation
MSI	Mass Spectrometry Imaging
MS/MS	Tandem Mass Spectrometry
MWCO	Molecular Weight Cut Off
m/*z*	Mass-to-Charge Ratio

N

nESI	nano Electrospray Ionisation
nm	nanometre ($1x10^{-9}$ m)
nM	nanomolar ($1x10^{-9}$ M)
NMR	Nuclear Magnetic Resonance
ns	nanosecond ($1x10^{-9}$ seconds)
NTD	N-terminal Domain

P

PA	Projection Approximation
PAGE	Polyacrylamide Gel Electrophoresis
PDB	Protein Database
Pi	Phenotype Identification

ppm	Parts Per Million
PSA	Projected Superposition Approximation

R

R	Resolving power
R^2	Coefficient of Determination
RF	Radio frequency

S

SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEC-MALS	Size Exclusion Chromatography-Coupled Multi-Angle Light Scattering
Serpin	Serine proteinase inhibitor
SRIG	Stacked ring ion guide
ssDNA	single stranded DNA
SV-AUC	Sedimentation Velocity-Analytical Ultracentrifugation

T

t_D	Drift time derived from FWHM (usually measured in ms)
T_m	Melting temperature
TM	Trajectory Method
ToF	Time-of-flight
TPR	Tetratricopeptide (34 amino acid) repeat
TPRκ	Sgt1 1-150 mutant; TPR plus capping helix (κ)
TPRκλ	Sgt1 1-178 mutant; TPR plus capping helix (κ) and linker (λ)
TTR	Transthyretin
T-Wave	Travelling Wave

V

V	Volt

W

wt	Wild type (i.e. phenotype found in nature)

Z

z	Ion charge value
z/m	Charge-to-Mass ratio

Other

50[50]	50 bp [50 % GC content] dsDNA
Å	Angstrom ($1x10^{-10}$ meter)
Δm	Resolution (m_2 - m_1)
ΔSGS	Sgt1 1-278 mutant
μM	micromolar ($1x10^{-6}$ M)
$^{+}/_{-}$	plus/minus

Acknowledgements

Try not. Do, or do not. There is no try.

Yoda (896 BBY - 4 ABY)

Star Wars Episode V: The Empire Strikes Back

Throughout any life there are times that the individual must persevere and remain determined to achieve their ambitions. On many of these occasions, they will rely on the support of others to aid them. I am therefore proud to acknowledge the following people, all of whom have made a positive impact on my life as I progressed through my PhD studies.

I would first of all like to thank Mikie Kukwikila and Rachael Morgan who, having shared an office with me for a notable period of time, have both born the burden of listening to my many tirades. I would like to further extend my gratitude the other members of office 334, both past and present, that I have come to know over the past four years: Jon, Lana, Isabelle, and Husna to name but a few.

To the newly engaged Samantha Chalker, thank you for all your support and putting up with me on a near-daily basis over the past few years. I'm simply sorry that I will not be able to be there when you get married next year.

To the numerous past, present, and affiliated individuals of office 320 who I have shared more than a fair share of my Friday night drinks with, including: Rachael Hazael, Richard Briggs, Michael Warwick, and the many others who I am unable to list here. I thank you for those many times we shared drinking with only that one golden rule of thumb; 'No science after six!'

To my secondary supervisor, Stefan Howorka, I would like note my gratitude for all his support throughout my studies at UCL. Additionally I would like to acknowledge my former supervisor, Adam McKay, for his contribution in introducing me to the field of biological mass spectrometry.

For their key roles in leading the Sgt1 project I would like to extend my thanks to Cara Vaughan, and Oliver Willhöft. For leading the α_1-Antitrypsin project I would like to

extend my thanks to Bibek Gooptu and Peg Nyon. Additionally, I would like to thank David Grainger for supplying all of the CbpA samples.

I would like to thank all of the postgraduate members of the Thalassinos research group: Adam Cryar, Harpal Sahota, David Sutherland, Ganesh Sivalingam, and Jun Yan. Your support and alternative perspectives of mass spectrometry related issues have proved to be continually insightful.

I could not finish however without acknowledging my little brother Daniel (a.k.a. 'Brat'), and my Auntie Lucie, for their love and support in all of my endeavours. And to the feather and fur kids: Sparky (who having sadly passed away shortly after this thesis was submitted for examination will be sorely missed), Biggles, Chester, Barbosa and Oliver, for their warm and welcoming fuzzy faces at the end of every day.

As they say, however, always leave the best to last. The fact is if it wasn't for the following three people, the chances are I would not be here writing these words. I therefore owe them my eternal gratitude.

To my supervisor Konstantinos Thalassinos, I would like extend my gratitude for all his support and advice, both in science and issues unrelated.

To Jade Hamilton, I thank you for being there through thick, thin, and often at your own expense. Also, as there is a 50 % chance I'll be right; 'B.'

And to my ever supportive Mother, you've been there since day one, and have acted as the metaphorical single set of foot prints in the sand on more than several occasions since. I don't know what I would have done without you.

To you all, I wish you the best in all of your future endeavours, and I hope that our paths will cross again soon.

1. Of Mountains and Molehills: the Development of Mass Spectrometry and Ion Mobility for the Study of Noncovalent Complexes

---

As mass spectrometry continues past its recent centenary, the development of the field continues to grow at an ever expanding rate, driving the high mass range further, mass resolution ever greater, and analysis capabilities in continuously branching directions. Indeed if the father of the field, Joseph J. Thomson, could compare the mass of ions and level of information that can be obtained from instruments he used a little over 100 years ago, to what can be acquired now, chances are that we may indeed cite his words as 'like comparing mountains to molehills.'

1.1. The Birth of the Field

Whilst investigating cathode rays by deflection in electric and magnetic fields, Joseph J. Thomson produced results that enabled measurement of the *z*/*m* (charge-to-mass ratio) of an electron (Thomson 1897). For his work, Thomson, won the 1906 Nobel Prize in Physics (The Nobel Prize in Physics 1906). As his work and interests progressed, Thomson initiated studies of anode rays, constructing his first mass spectrograph instrument in 1907 (Thomson 1907). This instrumental set up used a fluorescent screen to detect and record the trajectory of hydrogen ions, under the influence of magnetic deflection proportional to the *z/m* value (Budzikiewicz & Grigsby 2006). Resulting from these, and additional, observations Thomson would subsequently conclude that there was a "different type of positive ion for each element" (Thomson 1913).

In collaboration with Francis W. Aston, the 1907 design was updated in 1910 to improve the operating vacuum providing increases in mass resolution (Budzikiewicz & Grigsby 2006). This updated instrument was used to progress the studies of anode rays, in addition to initiating analysis of the inert gasses. Accompanying the expected mass value for Neon at *m/z* (mass-to-charge ratio) 20, their results indicated a further mass line at

m/z 22, in addition to evidence of doubly charged species. Summarising these data in a lecture given in 1913 (Thomson 1914), Thomson discussed the potential application of this value stating "that was has been called neon is not a simple gas but a mixture of two gases." The unexpected mass value at *m/z* 22 was, however, unresolved and resulted in a level of discussion, with several hypotheses proposed by both Thomson and Aston to explain this unexpected result (Audi 2006). Due to the low mass resolving power of the instruments available at the time, no proposed hypothesis was conclusively proven.

In the years following Thomson's mass spectrograph studies of Ne and H ions, Thomson would develop what can be argued to represent the first mass spectrometer (Thomson 1912). Ion rays where deflected under the influence of a constant electric field. This was followed in space by a magnetic field that could be modified to direct the trajectory of ions proportional to the analyte *m/z.* Magnetic deflection directs the charged analyte towards a physical slit in front of a faraday cup, which detects the charge of the ion. Using this instrumentation, Thomson produced the first mass spectrum of CO_2^+, with evidence of CO^+, O^+, C^+, and H^+ species (Fig. 1.1) (Budzikiewicz & Grigsby 2006).

As instrumental advancements continued, Aston developed the velocity focusing mass spectrograph providing further increases in resolving power (Aston 1919) that would finally settle the discussion regarding the unknown mass at *m/z* 22 started almost a decade earlier. Using this updated instrument, Aston determined that the previously inconclusive value was the isotope ^{22}Ne, and represents the first conclusive evidence of stable atom isotopes. This is not excluding a further weaker line (0.3 %), indicating the existence a further isotope of neon, ^{21}Ne (Aston 1920a). In combination with further work for his development of the whole number rule (Aston 1920b), Aston would win the 1922 Nobel Prize in Chemistry (The Nobel Prize in Chemistry 1922).

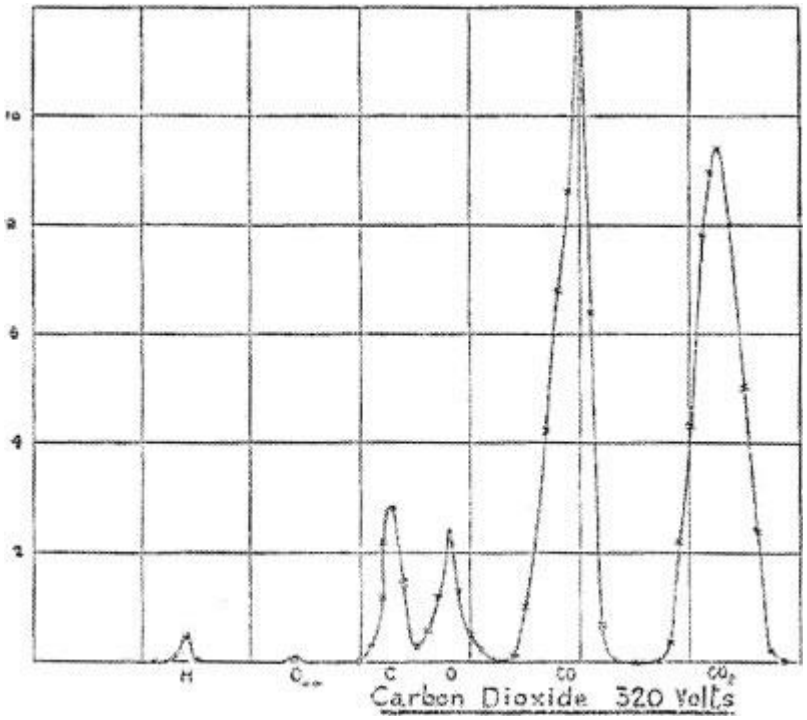

Figure 1.1. The First Mass Spectrum of CO_2

The first mass spectrum of CO_2 obtained by Joseph J. Thomson (1912). Peaks (heights normalised to the most intense peak) represent ion masses corresponding to (left to right) H^+ (4 %), C^+ (20 %), O^+ (18 %), CO^+ (100 %), and CO_2^+ (80 %). Figure reproduced from Budzikiewicz & Grigsby 2006

1.1.1. Progression of Mass Spectrometry to a Higher Prominence

For the vast majority of its early applications, mass spectrometry was confined within the field of physics as investigators applied the technique to answer a number of questions regarding isotopes and the fundamental nature of the atom. Studies such as these subsequently drove the development of mass spectrometers with greater resolving power, and extended mass ranges.

These studies and separations of isotopes would later propel mass spectrometry to a higher prominence with the out break of World War II, and initiation of the Manhattan Project[1] by the UK, USA, and Canadian governments (Griffiths 2008). In 1940 Alfred Nier had shown, using mass spectrometry, that ^{235}U was responsible for the slow neutron fissionable component of natural uranium, with the method used outlined in Nier (1989). Mass spectrometry was applied within the Manhattan Project to measure the stable and non-stable isotope abundances of uranium (and later plutonium), as a post processing tool

[1] The Manhattan Project was initiated with the aim to utilise fissionable materials that would lead to the development the world's first atomic weapons (Nier 1989).

to analyse the effectiveness of various enrichment methods. It has been said that without Nier or, the availability of mass spectrometry, the Manhattan Project would never have been completed within the required time-scale (De Laeter 1996, and Nier 1989).

As an individual, Nier was one of the first individuals to heavily promote the technique outside of the physics community. This includes the first biological application of mass spectrometry, preparing enriched ^{13}C samples for tracer studies. It has been suggested that it is because of Niers' tireless efforts that mass spectrometry was propelled to the tool that it has become today, used within the fields of all the classical sciences (Griffiths 2008).

1.2. Electrospray Ionisation and MALDI: 'Wings for Molecular Elephants'

Other than the mass analyser itself, arguably the most important component of any mass spectrometer is its ionisation source. These sources impart a level of charge on the analyte permitting detection of the sample within the instrument. Prior to the mid-to-late nineteen eighties however, ionisation methods required for mass analysis had been too harsh for larger biological macromolecules and complexes to be transferred into the gas phase intact and therefore remained primarily a tool of physicists and chemists.

Upon the development of Matrix-Assisted Laser Desorption/Ionisation (MALDI, section 1.2.1.) in 1988 by Karas & Hillenkamp (1988), with similar methods developed by Tanaka and co-workers (1988),[2] and the electrospray ionisation (ESI, section 1.2.2.) source for biological applications in 1989 by Fenn *et al.* (1989), the potential to study these macro molecules and complexes was suddenly possible. These methods would open the doors to an almost completely new field of structural and biological mass spectrometry (Hilton & Benesch 2012). Combination of these two techniques "[conferring] the greatest benefit on mankind," (Nobel 1895) attributed Fenn and Tanaka, the 2002 Nobel Prize in Chemistry[3] (The Nobel Prize in Chemistry 2002).

[2] Laser desorption ionization of "ultra fine metal plus liquid [glycerol + organic solvents] matrix method" within which the sample was diluted (Tanaka *et al.* 1988).

[3] Awarded in combination with Kurt Wüthrich for his development of NMR techniques for determining 3D structures of biological macromolecules in solution (The Nobel Prize in Chemistry 2002).

1.2.1. The Fundamentals of Matrix-Assisted Laser Desorption/Ionisation

The development of MALDI by Karas & Hillenkamp (1988), provided a drastic improvement in the potential to study high mass ions, combined with improved sensitivity compared to previously developed methods such as fast atom bombardment and secondary ion mass spectrometry. This mass range increased from species of 10 kDa using these previous methods, to proteins of 35 kDa (Karas & Hillenkamp 1988, Karas & Krüger 2003, and Tanaka *et al.* 1988).

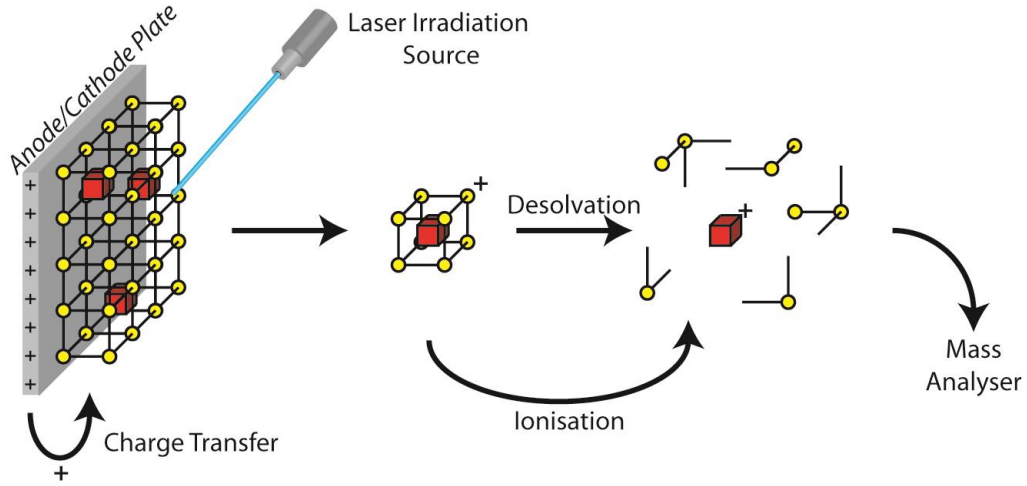

Figure 1.2. Principles of the MALDI Ionisation Source

The analyte of interest (red squares) is mixed with a suitable organic matrix which absorbs energy at certain wavelengths. This mixture is affixed to a metal plate and dried, embedding the analyte throughout the matrix. The analyte/matrix loaded plate is then placed within a vacuum, a charge applied to the plate which acts as the anode/cathode for ion formation (depicted as the anode), and irradiated using a pulsed laser (e.g. UV laser) (Karas & Hillenkamp 1988). Irradiation heats the matrix promoting localised sublimation, allowing entry of charge laden analyte/matrix clusters into the gas phase under the influence of a charge gradient (Dreisewerd 2003). The analyte sequentially undergoes desolvation, allowing secondary ionisation mechanisms to occur (see Fig. 1.3) (Karas & Krüger 2003).

Prior to ionisation, the analyte is first dissolved in a solvent suspended small organic matrix, such as, 2,5-dihydroxy benzoic acid (Jackson *et al.* 2007), or α-cyano-4-hydroxy-cinnamic acid (Pittenauer & Allmaier 2009). This mixture is subsequently dried, and ionised via the action of a pulsed laser (outlined in greater detail in

Fig. 1.2), and sequestered into the gas phase through charge potential differences. The matrix functions to absorb the majority of the laser energy, whilst permitting transfer of analyte/matrix clusters into the gas phase, without degradation of the analyte (Chughtai & Heeren 2010).

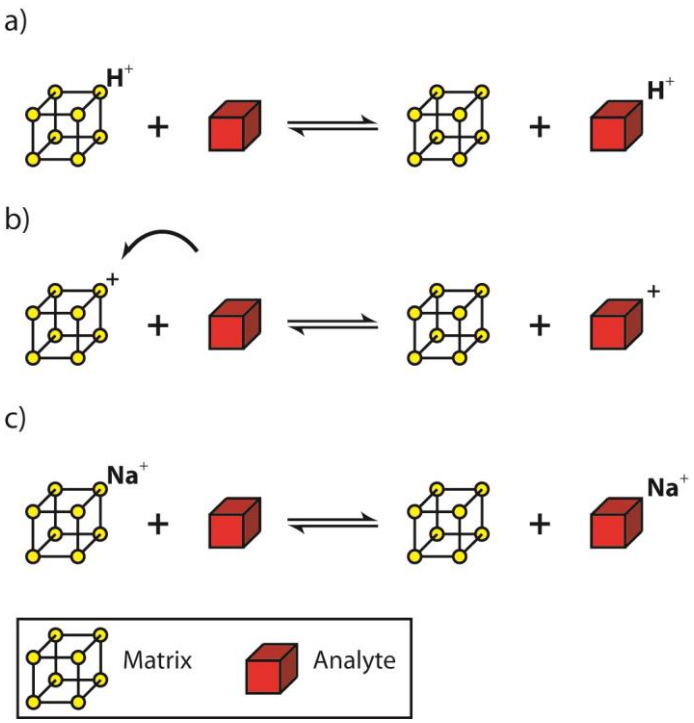

Figure 1.3. Positive Mode MALDI Secondary Ionisation Mechanisms MALDI supports a range of secondary ion formation mechanisms, which occur within the laser induced plume i.e. the solid to gas phase transition (Zenobi & Knochenmuss 1998). a) Ionisation by the proton transfer mechanism, progresses by matrix proton (H^+) donation to the analyte. This occurs due to the analyte having a higher proton affinity than the matrix (Knochenmuss & Zenobi 2003). b) Under suitable analyte/matrix combinations the electron transfer mechanism supports a model where the matrix sequesters an electron from the analyte, producing a charged radical analyte. This method however is not widely observed using UV laser source (Knochenmuss & Zenobi 2003, Macha *et al.* 1999, and McCarley *et al.* 1998). c) The cation transfer mechanism, supports a model whereby a charged adduct (in this case Na^+) is transferred from the matrix to the analyte, imparting its charge signal. Comparatively proteins and peptides have a lower affinity for adducts such as Na^+ compared to H^+. Therefore when the matrix has a higher affinity for H^+ than the analyte, Na^+ derived charge can become the predominant species (Knochenmuss & Zenobi 2003).

Ionisation of the analyte is not conclusively understood, and can be explained by several different means, separated in terms of primary (direct) or secondary (gas phase) ionization mechanisms. Primary ionization mechanisms are understood to occur as a result of two methods, photochemical processes (i.e. direct influence of the laser resulting in radical formation), or an anion/cation excess that directly ionises the analyte

(Karas & Krüger 2003, and Zenobi & Knochenmuss 1998). Mechanisms of secondary ionisation are outlined in figure 1.3.

As extended understanding of the MALDI source is of little relevance to the investigation presented here, for a larger overarching review on the method the author directs the reader to the following review articles by Karas & Krüger (2003), and Zenobi & Knochenmuss (1998).

1.2.2. The Development and Fundamentals of Electrospray Ionisation

The fundamental basis of ESI was not a new one in the 1980's with Dole and co-workers (1968) having previously demonstrated that a negatively charged solution (60 % Benzene, 2 % Acetone) could be used to spray a dilute aromatic polymer solution into the gas phase for analysis (Dole *et al.* 1968).

John Fenn, to whom the development of ESI for biological applications is attributed, started working with the technique in the early/mid 1980's, hypothesising its use for large organic molecules in Yamashita & Fenn (1984). Evidence supporting Fenn's hypotheses quickly followed, and in 1985 he presented the first evidence of gas phase peptide analysis (all less than 1500 Da) by ESI (Whitehouse *et al.* 1985). By 1989 further developments had allowed Fenn and co-workers to clearly demonstrate the use of ESI for the analysis of large non-native proteins that exceeded 130 kDa (Bovine albumin; 133 kDa) (Fenn *et al.* 1989).

The principle of ESI for mass spectrometric analysis relies on several key factors. First and foremost is the use of a suitable, low vapour pressure solvent such as: acetonitrile, isopropanol or, methanol (Fenn *et al.* 1989). This buffer is used to suspend a dilute concentration of the analyte of interest (mM to nM concentrations) prior to injection into the ESI capillary source (inner diameter 100 nm, Yamashita & Fenn 1984). Upon injection (Fig. 1.4a), a solvent dependent voltage is applied to the capillary (Kebarle & Verkerk 2009), inducing an electrospray aerosol through the formation of a Taylor cone. This Taylor cone dependent electrospray allows the solvent suspended analyte to enter the gas phase for mass spectrometric analysis via a charge gradient.

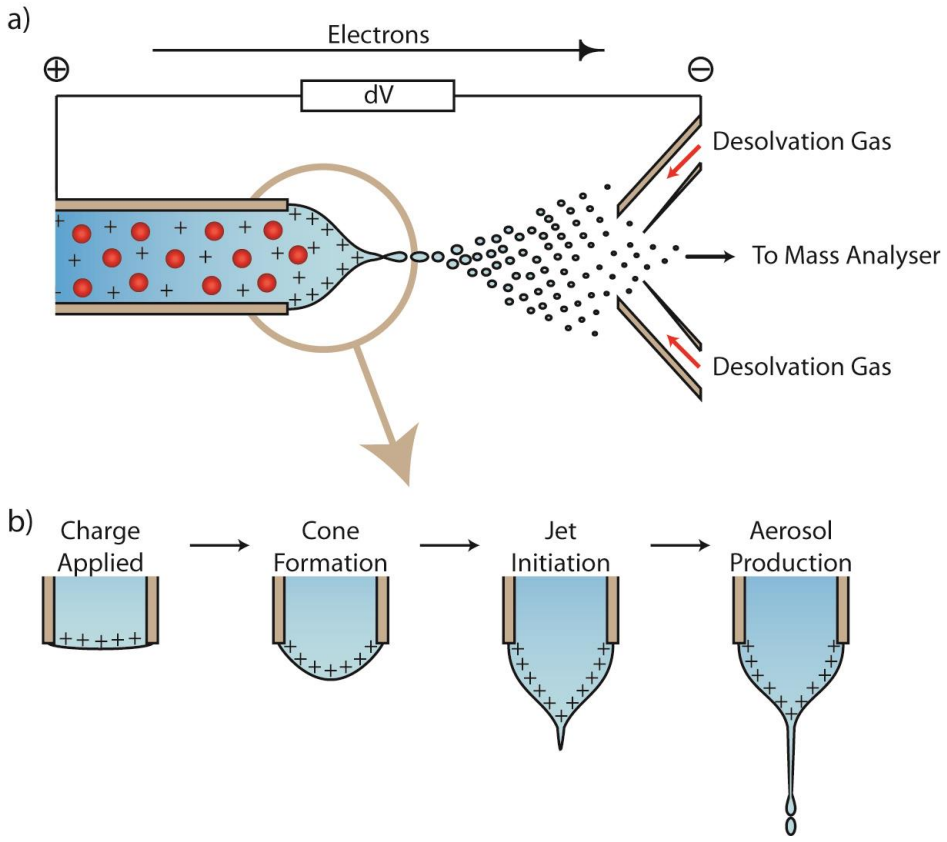

Figure 1.4. Principles of Electrospray Ionisation and Taylor Cone Theory

a) A dilute sample of analyte (red spheres) is injected (few nL/min in nESI to, ml/min values in liquid chromatography-mass spectrometry, Karas *et al.* 2000) into the capillary source. A voltage difference between the capillary and counter-electrode, at 3 to 6 kV over a distance of 0.5-2 cm (Smith *et al.* 1990), induces formation of the Taylor cone (b). This enables production of a fine aerosol spray which undergoes desolvation (Fig. 1.5), aided by a heated desolvation gas (initial temperatures 46-76 °C increased to vaporise 90-95 % of the solvent, Fenn *et al.* 1989) as it transfers into the vacuum of the gas phase.

b) Taylor cone formation occurs due to the application of a current to a solvent, resulting in electrostatic repulsion that deforms the upper surface producing a convex meniscus. As the electrostatic charge accumulates, the angle of the meniscus increases and above a critical field (i.e. electrostatic repulsion > surface tension) becomes unstable and the jet is initiated. Subsequently an analyte laden, charged aerosol is produced (Luedtke *et al.* 2008, and Wilm & Mann 1994).

Deformation of liquid droplets under the force of electric fields, was first demonstrated in 1964 by Sir Geoffrey Taylor (Taylor 1964), and is explained in terms of the cone theory that bears his name. The basis of this theory is that solvent deformation at the ESI source cone results in an aerosol production (Fig. 1.4b), providing a suitable

surface area for ion desolvation to occur. Production of desolvated ions from the Taylor cone aerosol is believed to manifest under one of two proposed models: the charged residue model (Fig. 1.5a), and the ion desorption model (Fig. 1.5b). Both of these models are dependent on the Rayleigh limit (Eq. 1.1) (Rayleigh 1882) that defines the limit value on droplet total surface charge (Gomez & Tang 1994, and Kebarle & Verkerk 2009).

$$q^2 = 8\pi^2 \varepsilon_0 \gamma D^3 \quad \text{(Eq. 1.1)}$$

q charge on droplet surface, ε_0 the permittivity of medium surrounding the droplet, γ the liquid surface tension, and D the droplet diameter.

The key components of this definition are the surface tension force of the droplet, and the coulombic repulsion of like-like charges on the droplet surface. When the former exceeds the latter, the droplet remains intact. As the droplet loses mass (and diameter) through solvent evaporation, the coulombic repulsion value equals and eventually overcomes the surface tension, causing droplet fission (Gomez & Tang 1994, and Last *et al.* 2002).

The charged residue model (Fig. 1.5a), originally proposed by Dole *et al.* (1968) upon the development of ESI, explains charged analyte production through repeated solvent evaporation and coulombic fission events. Current evidence (De La Mora 2000) supports that this is the mechanism by which native-like, macroions such as globular proteins are generated within the gas phase.

The ion desorption model (also known as the ion evaporation model, Fig. 1.5b), was originally proposed by Iribarne & Thomson (1976) and later supported by Hager and co-workers (1994). This model states that the production of ions occurs through the direct desorption of charged analyte once the radii of the electrospray aerosol droplet drops below 10 nm. This model also requires that the droplet maintains a suitable surface charge density at a value less than the Rayleigh stability limit. Current reports suggest that this is the mechanism by which small organic and inorganic ions are generated within the gas phase (Cole 2000, Felitsyn *et al.* 2002, and Kebarle & Peschke 2000).

It has been suggested however that both methods may occur in tandem. Whilst initial desolvation events occur by the charged residue model, once the mass of droplet reaches the 10 nm threshold size, direct desorption of the ion is able to occur (Cole 2000).

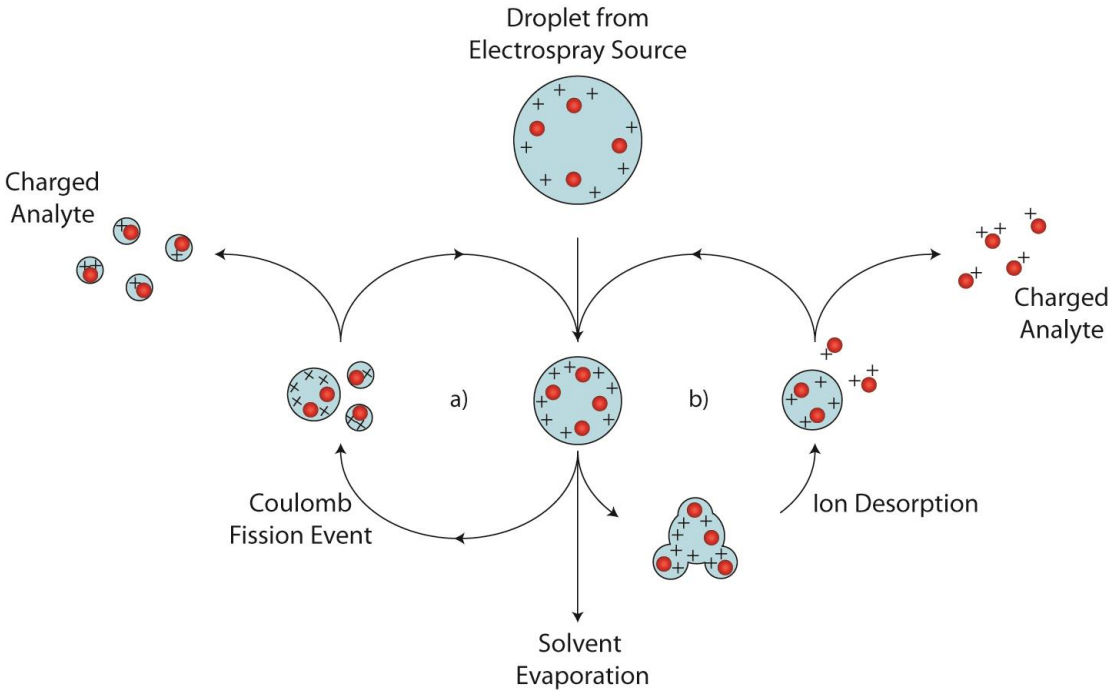

Figure 1.5. Mechanisms of ESI Analyte Desolvation and Ion Formation

Both the (a) charged residue, and (b) ion desorption/evaporation models are initiated upon ejection of charged solvent (blue spheres) suspended analyte (red spheres) from the Taylor cone aerosol (Fig. 1.4b). This suspended analyte loses mass via solvent evaporation, whilst maintaining a similar level of charge to the initial droplet. a) The charged residue model occurs due to increasing internal coulombic repulsion resulting from decreasing solvent levels. Below a critical mass the charge repulsion exceeds the surface tension holding the droplet together. Passage of this limit promotes droplet fission. This evaporation/fission cycle is repeated several times between aerosol formation and mass analysis. This cycle typically results in a single molecule (dependent on starting concentration, Sun *et al.* 2007), of one or more charges, within each solvent droplet (Cole 2000, and Heck & van den Heuvel 2004). b) The ion desorption/evaporation model dictates that as a droplet loses mass from desolvation the increased electrostatic charge drives expulsion of analyte from the solvent. The analyte is expelled with added charge, and little excess buffer bound. Desorption has little influence on the overall solvent mass (Iribarne & Thomson 1976).

ESI however, is not without its limitations for the study of large native-like biomolecules and complexes, including protein, DNA and RNA. These are typically cited as the use of organic solvents and high temperatures to aid desolvation, which reduce the native like oligomeric and conformational state of the analyte. The required high flow rates can further limit the application of ESI when available sample is limited (Hilton & Benesch 2012, and Karas *et al.* 2000).

There are however, observations of ESI applied to the study of native-like proteins. In the early 1990s scientists within the Chait research group produced the first evidence for native-like proteins in the gas phase. Studying bovine cytochrome C (11.7 kDa, Chowdhury, Katta & Chait 1990) and equine myoglobin (17 kDa, Katta & Chait 1991), using a variety of pH values, investigations highlighted the appearance of charge state distributions (CSD) at higher *m/z* values compared to the non-native. These high *m/z* CSD values are indicative of the native-like state due to the reduced solvent accessible area that inhibits higher charge levels.

1.2.2.1. The Continued Development and Miniaturisation of Electrospray Ionisation

Miniaturisation of the ESI source in the early-to-mid nineteen nineties, led to the development of micro-electrospray (inner diameter 5-50 μm) by Emmet & Caprioli (1994) and Gale & Smith (1993), and sequentially removed certain limitations imposed by standard ESI. Further refinements of the design by Wilm & Mann (1994) resulted in the development of the nano-electrospray ionisation (nESI) source.

In contrast to previous techniques, nESI provided a reduced capillary inner diameter of 1-2 μm (from 100 μm, Yamashita & Fenn 1984) enabling a drastically reduced flow rate of ~20 nL/min (from 1-20 μL/min, Fenn *et al.* 1989). This reduced capillary size additionally produced droplets with a smaller diameter from the Taylor cone aerosol, removing the requirement for harsh desolvation conditions (e.g. desolvation gas temperature) that can affect the native state of large biomolecules/complexes (Wilm & Mann 1996).

The lower sample flow would typically be identified as a limiting factor, expected to result in decreased ion signal intensity due to the reduced effective sample concentration (i.e. sample reaching mass detector). In contrast the opposite phenomenon is observed, with improved counts and peak resolution. Juraschek and co-workers (1999) showed that these improvements resulted from the improved salt tolerance exhibited by the smaller droplet size, by comparing mass spectra of NaCl containing solutions of peptides using both ESI and nESI. With droplet size as the primary cause of the improved tolerance, the work highlighted that the reduced diameter prevented large increases in salt concentration

during desolvation. These decreases additionally improved sensitivity by reducing singly-charged specie sample loss (Juraschek *et al.* 1999).

With the reduced droplet diameter and improved salt tolerance of nESI, the requirement to use organic solvents and acids, in tandem with high source temperatures to permit suitable sample desolvation were removed (Hilton & Benesch 2012). In lieu neutral aqueous volatile buffers such as, ammonium acetate (AmAc) and ammonium bicarbonate could be used (Felitsyn *et al.* 2002). Buffers such as these aid the preservation of the native like conformation of proteins in the gas phase, required for the study of noncovalent oligomeric complexes and fundamental for the study of conformational dynamics of ions using mobility separation (section 1.4) (Hilton & Benesch 2012, van den Heuvel & Heck 2004, and Winston & Fitzgerald 1997).

Typical nESI source emitter tips[4] are currently produced using borosilicate glass capillaries that are shaped to produce the desired exit diameter (1-2 μm). *In-house* production of these nESI emitter tips is becoming increasingly common, with instruments such as the P-97 (Sutter Instruments, CA, USA) commonplace within mass spectrometry research groups.

To ionise samples these nESI emitter tips classically used a platinum wire housed within the capillary to ionise the analyte laden solvent. Subsequent observations by Wang and co-workers (2003) indicated that this method of sample ionisation combined with the low sample flow typical of nESI, results in the electrochemical breakdown of water (Oxidation in positive mode, reduction in negative mode) (Kitova *et al.* 2012, and Wang *et al.* 2003). Operation in positive mode (typical of native protein ionisation) therefore leads to decreases in the pH over time (as much as 1 pH unit within 30 minutes), resultant from the production of H_3O^+ ions (Wang *et al.* 2003). These changes in pH can adversely influence electrostatic and ligand binding interactions, which are undesirable in native mass spectrometry studies.

Today, the most common method for nESI emitter tip production is to coat shaped capillaries with a thin layer of gold, or a gold/palladium mix (such as those used by Tito *et al.* 2000). This coat conducts the charge required for production of the Taylor cone and sequential analyte ionisation. Current literature however, fails to indicate if nESI

[4] Also referenced within the literature as 'needles' as early as 1996 (Wilm & Mann 1996)

capillaries produced using this method suffer the same electrochemical oxidation of water properties that the platinum wire-in-capillary source tips have been shown to suffer.

Additionally the use of chip based nESI ionisation sources, such as those developed by Advion (NY, USA) have also become common-place. These chip based sources permit high throughput sample analysis using a programmable automated system. Early designs, much like the wire in capillary methods, were identified to suffer from pH increases resulting from the oxidation of water due to electrical conductance of the source tips used to deliver the sample. Since these issues were highlighted however, internal wall coated source tips have been developed that no longer suffer from this problem due to a reduced ionisation surface area (van Berkel & Kertesz 2012).

1.3. Mass Analysis

Arguably as important as the ionisation source, the mass analyser is required to obtain information regarding mass from the charged analyte. A variety of principles exist to separate ions based on their *m/z* value, including: the quadrupole (section 1.3.2.) that uses *m/z* dependent trajectory stability, ion traps which use resonance frequency (Douglas 2009, and March 2009), or the time-of-flight (ToF) analyser (section 1.3.3.) that separates ions according to mass dependent velocity over a set distance through a field free region (Guilhaus 1995). Each of these mass analysers (along with those not listed), differ with respect to the following definitions of suitability:

Mass accuracy; difference between measured and theoretical *m/z*.

Mass range; limit of *m/z* mass detection where $z = 1$.

Resolving power; ability to distinguish between two peaks at similar *m/z* value (covered more extensively in section 1.3.1).

Speed; rate at which the analyser acquires results over a set mass range (units per time period stated).

No mass analyser however provides a suitable mix of the above characteristics for all situations. Whilst instruments such as Fourier transform-ion cyclotron resonance mass spectrometers provide excellent mass accuracy and resolution at low mass values, they lack

the mass range required for high mass studies. In contrast, ToF instruments exhibit this suitable mass range, but lack the resolution of the former (Domon and Aebersold 2006).

1.3.1. Mass Resolution and Resolving Power

Experimental method and design has in the past century driven the development of mass spectrometers with greatly improved mass resolving power. Since the development of the first mass spectrometers around the turn of the twentieth century, the resolving power (*R*) of mass spectrometers has shown rapid growth, and a select example of this progress is briefly highlighted below:

Parabola mass spectrograph, $R = 13$ (Thomson 1913)

Velocity focusing mass spectrograph, $R = 130$ (Aston 1919)

Second-order focusing mass spectrograph, $R = 2000$ (Aston 1937)

Fourier transform-ion cyclotron resonance-mass spectrometer, $R = 8 \times 10^6$ (Shi *et al.* 1998).

The *R* value can be described using equation 1.2 (Gross 2004, and Marshall *et al.* 2002):

$$R = \frac{m}{\Delta m} \qquad \text{(Eq. 1.2)}$$

m, represents the mass of the peak of interest on the *m/z* scale, *Δm*, represents the resolution defined by Eq. 1.3

The *Δm* of a mass spectrum is described in the terms of the difference in mass between two points, measured in Da (Eq. 1.3, Fig. 1.6) (de Hoffmann & Stroobant 2007):

$$\Delta m = m_2 - m_1 \qquad \text{(Eq. 1.3)}$$

Therefore at a constant *R* value, *Δm* will decrease as the *m/z* value increases (Fig 1.7). Consequently, manufacturers of mass spectrometers usually quote their instrument values as 1 in *R*, at a defined *m/z* range (Gross 2004). This definition is somewhat limited, as instrumental set ups that utilise the ToF (section 1.3.3.) for mass analysis which have fixed *R* values across the entire mass range, and therefore do not suffer from the phenomena outlined in figure 1.7.

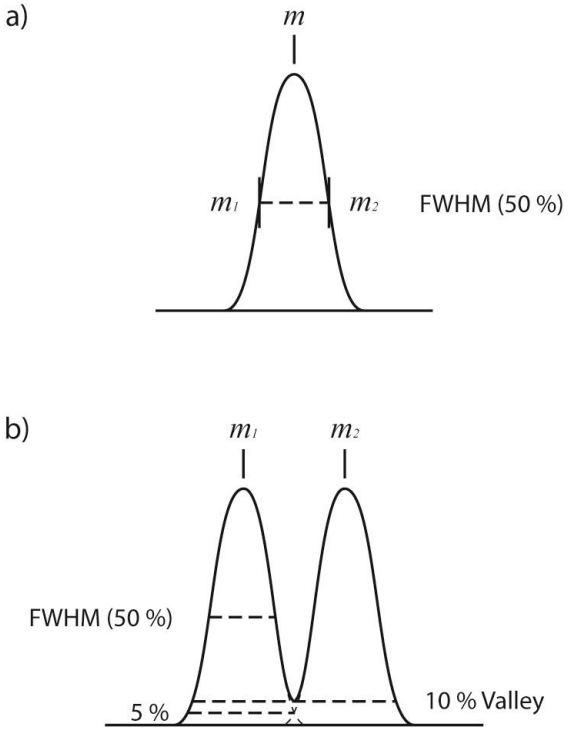

Figure 1.6. Determining Δm Using The Modern ($R_{50\%}$) and Classical ($R_{10\%}$) Methods

a) The modern determination of the R value, especially since the advent of the linear quadrupole, ToF, and ion-trap mass analysers, calculates Δm from the values m_1 and m_2 at the full width at half maximum (FWHM) of a single peak. b) The classical method measures Δm when the valley between two peaks is sufficient to exist at a level proportional to 10 % of the relative height, and termed $R_{10\%}$. Additionally the $R_{10\%}$ is only fulfilled if the peak width at $R_{5\%}$ equals the mass difference of the corresponding ion (Gross 2004).

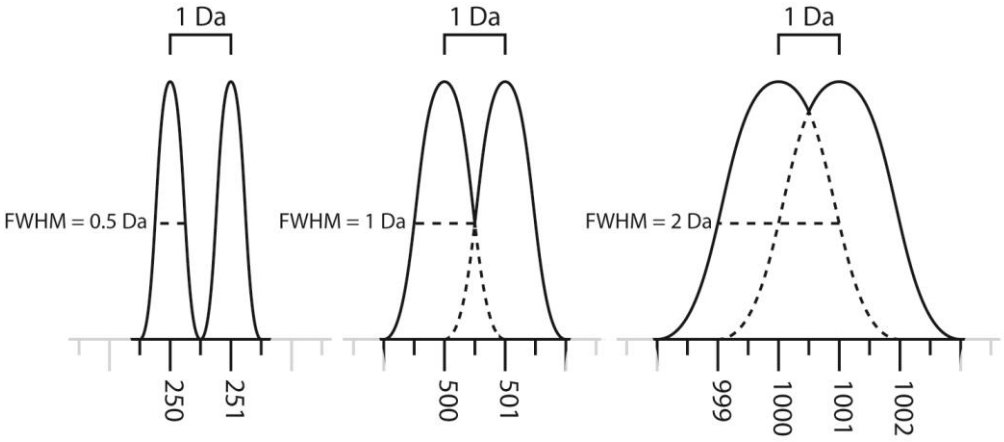

Figure 1.7. Effect of the *m/z* Value on the Ability to Distinguish Between Two Neighbour Ion Peaks at a Constant $R_{50\%}$

The $R_{50\%}$ at a set value of 500 supports that as the *m/z* value increases from 250, to 500, and 1000, the mass difference at FWHM value increases, proportional to Δm. This leads to increased levels of peak superimposition at higher *m/z* values. Therefore ion peaks at *m/z* 500 require an $R_{50\%}$ value of 1000 to resolve down to the baseline, if $\Delta m_{50\%} = \Delta m_{0\%} / 2$.

1.3.2. The Quadrupole Mass Analyser

The linear quadrupole for mass spectrometric applications was first described by Paul[5] & Steinwedel (1953) and is currently applied to a range of applications, including ion guide, filter, and mass analyser (Douglas 2009). The principle design of the quadrupole centres around four parallel rods of circular (Fig. 1.8), or hyperbolic section, through which direct currents are applied with alternating radio potentials (Douglas 2009, and El-Aneed *et al.* 2009).

Opposing rods share the same charge state, whilst neighbour rods the alternate, with the opposing charges labelled $+\Phi_0$ and $-\Phi_0$ (Fig. 1.8). These charges are defined using equation 1.4 in the positive, and negative respectively (Collings & Douglas 1997, and Douglas 2009).

$$\phi_0 = (U - V\cos\omega t) \qquad \text{(Eq. 1.4)}$$

Φ_0 is the rod charge potential, ω the angular frequency, U is the maximum DC voltage applied between pole pairs, and V the zero to peak amplitude (AC voltage).

The combination of U and V, whilst having no influence on the acceleration along the z axis, which remains constant, can be used to produce a region of stable ion trajectory potentials. This effectively filters ions based on the x/y acceleration potential, connected with mass and charge (Fig. 1.9) (Blaum 2006, and Douglas 2009).

By correctly filtering ions based on their stability, using optimal U and V values, the quadrupole can select species based on their m/z value which are sequentially detected by the ion detector. Application of a scanning range that comprises a range of U and V values can be applied used to produce a full mass spectrum within the quadrupole mass analyser. In contrast, quadrupole analysis lacking modification of these values to scan across the desired mass range produces a limited mass range that is able to filter ions (Chernuschevich *et al.* 2001, and Collings & Douglas 1997).

[5] An individual who would later help develop the 3D ion-trap (Paul & Steinwedel 1960), and combined with the development of the penning ion-trap by Dehmelt (1968) would be jointly awarded the 1989 Nobel Prize in Physics "for the development of the ion-trap technique" (The Nobel Prize in Physics 1989).

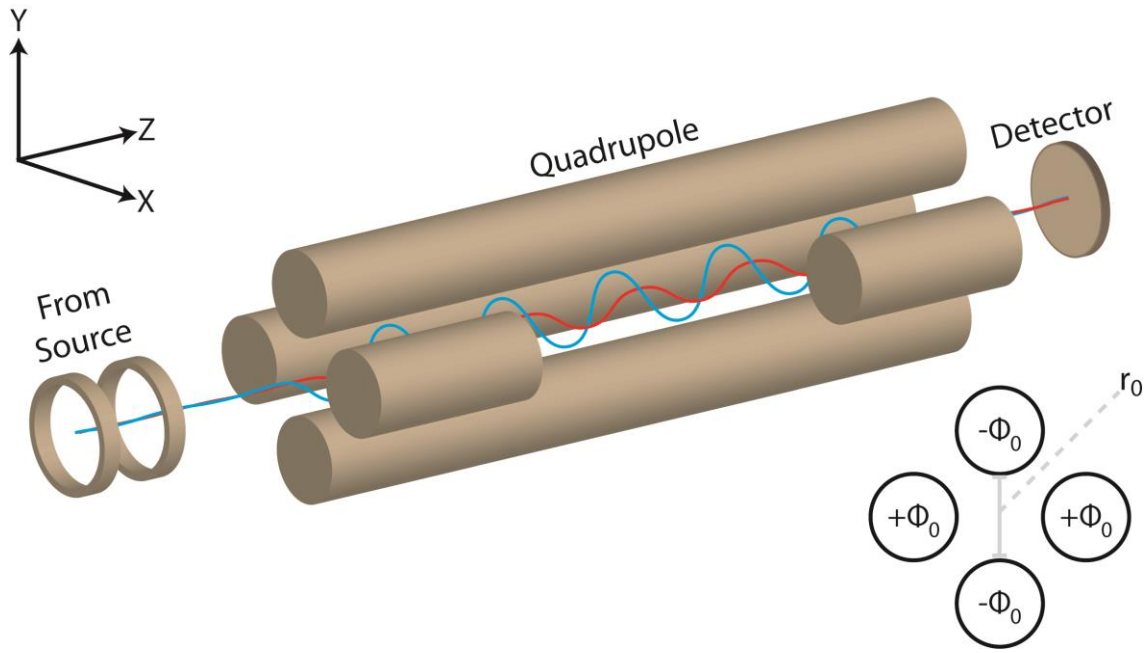

Figure 1.8. Basic Representation of the Linear Quadrupole

"Ions travelling along the z axis are subjected to the influence of a total electric field made up of a quadrupolar alternative field superimposed on a constant field resulting from the application of the potentials on the rods" defined using equation 1.4 (de Hoffmann & Stroobant 2007). With reference to the *U* and *V* values that are used to determine Φ_0, changes allow ions with different *m/z* values to traverse the quadrupole (See Fig. 1.9). Furthermore changes to the strength and frequency of the alternate field, in addition to fixed changes in the field radius (r_0) of the x/y axis can be applied to modify the mass range and ion transferral window capabilities (along the z axis) (Chernuschevich *et al.* 2001, and Douglas 2009).

The mass range limits of the quadrupole can be calculated using the Mathieu Parameter equations (see Collings & Douglas 1997). These equations ascertain that to extend the mass range of the quadrupole for high mass studies, an individual has to either extend the maximum DC and RF (Radio frequency) voltages, or decrease the field radius (r_0) and/or frequency of the RF (Chernuschevich *et al.* 2001, and Collings & Douglas 1997).

The quadrupole can be further combined in line with a collision cell, and a second mass analyser such as another quadrupole (e.g. triple quadrupole, see section 1.3.2.1), or ToF (section 1.3.3), for tandem mass spectrometry (MS/MS) in-space studies

(section 1.3.2.1).[6] The key difference between quadrupoles used for normal mass analysis (narrow select window) and high mass filtering (i.e. acting as an ion guide) however, is that non mass analysis quadrupoles run in RF only mode. The RF is then modified to allow a broader or narrower *m/z* range through the region (Chernuschevich *et al.* 2001).

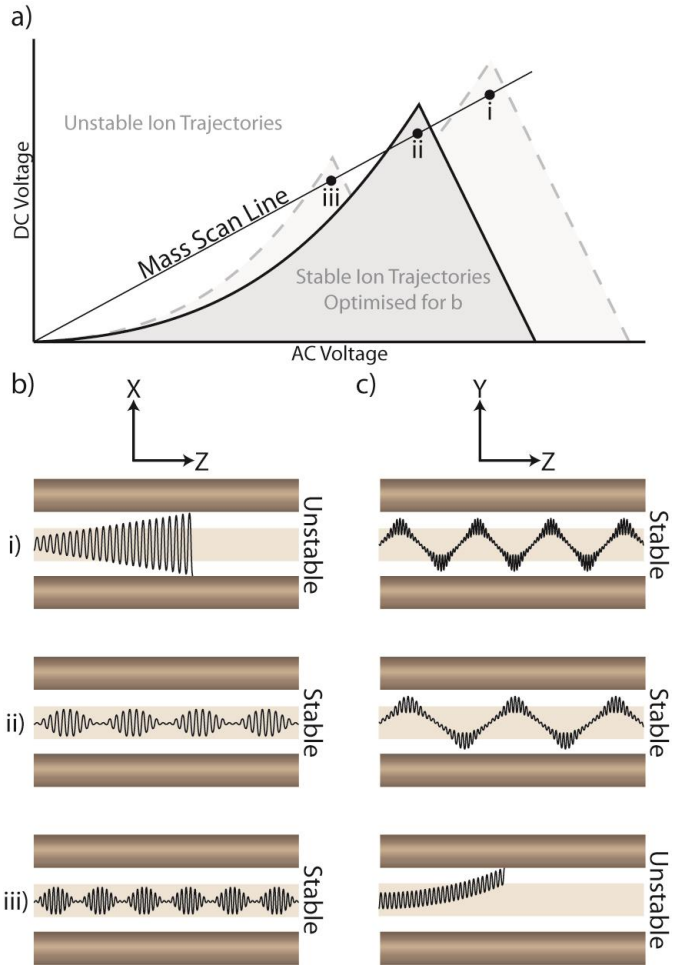

Figure 1.9. Trajectory Stabilisation of Ions Through a Linear Quadrupole

a) Stability diagram (dark grey area with axis numerical values not stated) where the shaded area represent the AC/DC voltage combinations corresponding to stable ion trajectories in the x-z (b, ii and iii) and x-y (c, i and ii) planes. Trajectory of ions i-iii, (set mass/increasing charge, or decreasing mass/set charge) highlights the trajectory along the (b) x-z or, (c) y-z axis. Ions therefore are only able to traverse the z axis, when the ion is stable in both x and y axis (Baum 2006, and Miller & Denton 1986). In practice the AC (*V*) and DC (*U*) voltages can be modified to shift the stability area, i.e. the quadrupole transmission window that allows i and iii to pass stably through both the x and y axis (lighter grey areas) (Chernuschevich *et al.* 2001).

1.3.2.1. Collision Induced Dissociation of Isolated Species

Development of collision induced dissociation (CID) of gas phase species by McLafferty & Bryce (1967), and Jennings (1968), had a major impact on mass spectrometry by extending the range of studies that could be performed (Hayes & Gross 1990). Using CID an

[6] MS/MS-in-time studies decomposition of ions as a series of sequential events, requiring the use of an ion storage device such as an Ion Trap (Johnson & Yost 1991).

investigator is able to induce the unimolecular decay of an analyte ion through collisions with inert neutral gases, such as helium and argon. Energy transfer between the neutral backing gas and analyte ion is generally influenced by several factors, including: collisional gas pressure, the molecular weight of the neutral gas (larger masses deposit more energy), collisional cross section of the analyte ion, in addition to their charge, injection energy and the loss of this energy as they traverse the collisional cell (Daniel *et al.* 2002, and Pittenauer & Allmaier 2009).

The development of the MIKES (Mass-analysed Ion Kinetic Energy Spectrometry), instrumentation by Beynon and co-workers (1973) allowed the mass selection of precursor ions using magnetic sector instrumentation and a variable width aperture. Subsequently the decomposition of mass selected ions was induced using a collision cell, prior to analysis of the fragmentation patterns using a second magnetic sector mass analyser (Beynon *et al.* 1973, and Kruger *et al.* 1976). This instrumentation found early applications in ion chemistry kinetics studies, such as those presented by Cooks and co-workers in the structural analysis of benzene and bromobenzene ions (Cooks, Beynon & Litton 1975).

Subsequent work presented by Kruger and co-workers (1976) applied the mass selection capabilities of the MIKES instrumentation for the analysis of a complex mixture of dialkyl ketones. Using the mass selection capabilities available, combined with analyte fragmentation induced by electron impact and chemical ionization, the dissociation patterns of the precursor ions could be studied to determine the component dialkyl ketones from the original sample (Kruger *et al.* 1976). This work represents the first implementation of the MS/MS technique, whereby the instrumentation has been applied for sample analysis in lieu of the fundamental studies of ions and molecules previously commonplace.

Instrumental developments by Yost & Enke (1978) would later lead to the development of the triple quadrupole mass spectrometer for MS/MS applications. This instrumentation incorporated three sequential in-space quadrupoles. The first and last operated in mass filter modes, permitting the selection and analysis of ions. In contrast the central quadrupole, operating in the RF only mode, was used to fragment ions for CID studies (Yost & Enke 1978). The application of quadrupoles for CID studies has since become commonplace, with many commercial and *in-house* built instruments incorporating at least one to mass filter ions for MS/MS studies.

Differences in MS/MS can be performed as a function of the energy levels (high or low) used to induce fragmentation. High energy (keV range) CID utilise a unimolecular beam, directed toward an outlet pump, perpendicular to a focal point of an accelerated ion beam. Interaction between the ion and unimolecular beam corresponding to gains in electronic energy, that induces decomposition of the analyte ion into fragment ions resultant from subtle differences in structural composition (e.g. peptide side chains) (Biemann 1990, and Pittenauer & Allmaier 2009).

Comparatively, dissociation of ions studied using low energy (< 100 eV, Pittenauer & Allmaier 2009) studies occurs under the influence of an unfocussed higher pressure backing gas. Interactions between the backing gas and analyte permit transmissions of energy from the backing gas to analyte that correspond to gains in vibrational energy. Although these energy transmissions are lower relative to the high energy, the unfocussed nature of the backing gas allows multiple collisions to occur (de Hoffmann 1996, and Jennings 2000).

Applied to structural biology analysis, using low-energy MS/MS methods an investigator is able to quadrupole mass select the ion of interest allowing studies of the structural organisation of macromolecular assemblies. Under increasing collision energies the outer, more weakly bound substrates will unfold and dissociate from the complex more rapidly. This provides an indication of the structural organisation and stability, with examples readily within the literature (Benesch *et al.* 2006, and Hernandez *et al.* 2006).

CID of large multimeric complexes typically results in a non-uniform redistribution of the ion charge. An early example of this non-uniform redistribution was presented by Sobott & Robinson (2004) using the isolated +65 ion of the 14 subunit *E. coli* GroEL oligomer. Fragmentation of ions, over an increasing collision voltage (4 to 200 V), highlighted the dissociation of a single GroEL subunit that carried ~50 % (z = 25-38) of the total charge of the isolated species. The remaining charge was distributed amongst the intact 13 subunit oligomer (z = 26-39).

Subsequent work performed by Benesch and co-workers (2006) would refine the explanation for this asymmetry. The misconception is that each subunit carries a level of charge proportionate to its mass. One would therefore expect an ion distribution functional to the ratio of the mass of the individual subunit, divided by the multimeric oligomer. In

actuality, the charge distribution occurs as a function of the surface area, typically confined to subunits with an increased solvent accessible area. This work supported the theory that removal of sequential subunits requires lower dissociation energies, resulting from coulombically favourable forces occurring due to changes in surface area and reduced charge density (Benesch *et al.* 2006).

1.3.3. The Time-of-Flight Mass Analyser

In terms of mass analysis, the ToF mass analyser is quickly becoming the default tool of choice, especially within native structural biology studies that rely on its superior high mass range. Combined with a high level of adaptability, compatibility with an extensive range soft ionisation sources (e.g. nESI, and MALDI), and other components (e.g. quadrupole, and ion mobility separator), the ToF mass analyser has been applied to an expanding number of native structural biology based studies.

The original concept for ToF analysis was proposed by Stephens (1946), and later demonstrated by Cameron & Eggers (1948). ToF analysers measure the physical properties of an analyte using velocity as a function of mass, timed over a set distance. As velocity and mass correlate negatively at a constant charge, two equally charged ions (e.g. R and B) pulsed into the ToF component at the same time, with B of greater mass than R, will dictate that R reaches the detector first.

Measurement of the *m/z* value can then be calculated from the time required for ions to traverse from the source, through the field free region of the ToF, to the ion detector (section 1.3.3.1). To calculate the *m/z* of an ion, using the linear free flight tube, the following equations (Eq. 1.5 to Eq. 1.10) are used.

Prior to entering the free-flight region, ions of specific mass (m) and total charge ($q = ze$) are accelerated by a potential (V_S). Its electric potential energy (E_{el}) is converted into kinetic energy (E_k), using equation 1.5:

$$E_k = \frac{mv^2}{2} = qV_S = zeV_S = E_{el} \quad \text{(Eq. 1.5)}$$

Rearrangement of this equation can be used to calculate the velocity (v, in ms^{-1}) of an ion through a simple free-flight tube using equation 1.6 below:

$$v = \left(\frac{2zeV_S}{m} \right)^{1/2} \qquad \text{(Eq. 1.6)}$$

Post acceleration, an ion travelling in a straight line at the velocity defined above, the time (t_f, in s) required to cover the distance of the free flight tube (L, in m) is given by equation 1.7:

$$t_f = \frac{L}{v} \qquad \text{(Eq. 1.7)}$$

Replacing v by its value in equation 1.6 gives equation 1.8:

$$t_f{}^2 = \frac{m}{z}\left(\frac{L^2}{2eV_S} \right) \qquad \text{(Eq. 1.8)}$$

e, represents the electronic charge in coulombs.

Equation 1.8 can therefore be rearranged to determine *m/z* from the t_f producing equation 1.9:

$$(m/z)^{1/2} = \left(\frac{\sqrt{2eV_s}}{L} \right) t_f \qquad \text{(Eq. 1.9)}$$

In practice, as L and V_S represent fixed values, applying the known *m/z* values of at least two suitably spaced calibration points (represented by *A* and *B*), equation 1.9 can be simplified to equation 1.10:

$$(m/z)^{1/2} = At_f + B \qquad \text{(Eq. 1.10)}$$

(Eq. 1.5 to 1.10 are taken from Guilhaus 1995, and de Hoffmann & Stroobant 2007)

To calculate m we require the need to calculate the charge value (z) of the ion. In the case of smaller ions that remain singly charged, i.e. z equals + or -1, using ESI protonation/deprotonation ($M+H^+/M-H^+$) methods, the mass gain/loss of the single added charge can be simply taken into account. Larger proteins/peptides (generally in excess of

1200 Da), are increasingly likely to undergo multiple protonation/deprotonation events during ESI, producing a Gaussian-type distribution. Assuming that two neighbouring charge peaks in a mass spectrum are part of the same Gaussian CSD the higher *m/z* value ion will carry the charge z_n, and the lower *m/z* z_{n+1} (n = number of charges). Determination of the z_n can be calculated using equation 1.11 below (Mann, Meng & Fenn 1989):

$$z_n = \frac{M_{z+1} - 1}{M_z - M_{z+1}} \qquad \text{(Eq. 1.11)}$$

M represents the m/z value of the parent ion at z and z+1 respectively.

Equation 1.11 allows us to calculate the charge number of z_n, and therefore given the mass of the added/lost H^+ ion using equation 1.12, we can calculate the mass in Da using the ion *m/z* at z_n:

$$Da = (M_z z_n) - (z_n H^+) \qquad \text{(Eq. 1.12)}$$

H^+ represents the mass of a proton using the C^{12} rule. Please note that the equation here represents mass calculation using the protonation model. With the deprotonation model, subtraction of the added mass is replaced with addition of the lost mass.

As the ToF mass resolving limitations are a function of the length of the free flight region, theoretically a mass spectrometer with a near-infinite length could hypothetically have a near limitless mass resolving power.[7] The key limitation of the linear ToF analyser however, is the failure to account for differences in ion kinetics that can modulate the peak broadness of mass spectra proportional to the length of the free flight tube (Fig. 1.10a part i and ii). As a result of this phenomenon, Mamyrin and co-workers (1973) combined the ToF with an electrostatic mirror (commonly known as a reflectron, Fig. 1.10b part i) to lessen the influence of kinetic differences, that modulate spectrum resolution (Fig. 1.10b part ii). Advantageously this instrumental set up effectively doubled the length of the free flight tube, within the same area of the linear tube (Mamyrin *et al.* 1973).

[7] Typically ToF cells are also coupled to quadrupoles (section 1.3.2.), and/or other ion guides, which in practice tend to limit the mass range of acquired spectra more than the length of the free flight tube.

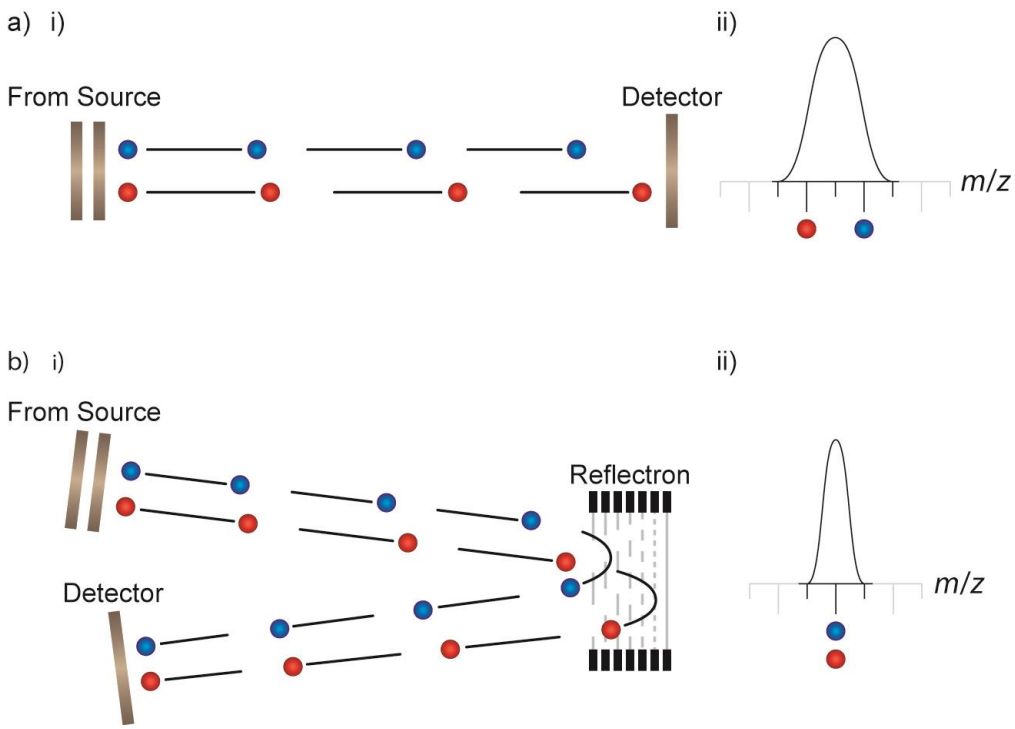

Figure 1.10. Linear and Reflectron Time of Flight Mass Analysers

Both red (R) and blue (B) components represent the same theoretical *m/z*, with B exhibiting a lower kinetic energy value. a) i) Differences in the kinetic energy promote difference in the arrival times of R and B reaching the detector. ii) These differences sequentially modulate the spectrum peak width and mass accuracy accordingly. b) i) Addition of an electrostatic mirror (reflectron) counteracts the influence of kinetic differences between R and B. ii) In addition to doubling the length of the free flight tube, compared to a linear ToF of comparable size, this electrostatic mirror reduces peak broadening.

1.3.3.1. Detecting Ions at the End of the Free Flight Tube

The earliest method to record ions, by Aston and Thomson, were photographic plates that allowed them to calculate the *m/z* or *z/m* values as a function of the deflection exhibited under the influence of a magnetic field. In contrast, ion detection at the end point of the ToF free flight tube is typically performed using a microchannel plate (MCP). These are a flat conversion surface with micro channels orientated 82 degrees from the parallel (Chernuschevich *et al.* 2001, and Wiza 1979). Ion strike events at the MCP produce electrostatic pulse ~1-2 ns in length positively correlated with the momentum of the analyte (Chernuschevich *et al.* 2001, and Jungmann *et al.* 2013). This pulse is amplified, and used

to the record t_f. Variations in MCP instrumentation exist, including the transient recorder, and time-to-digital recorder (Chernuschevich *et al.* 2001).

Application of these MCPs is dependent on the typical ionisation source at the front end of the mass spectrometer (Chernuschevich *et al.* 2001). The transient recorder is an analogue system, typically used for MALDI-ToF instruments due its wide dynamic range, and digitises output ion current from the MCP. Typically these are of limited use for single ion or MS/MS analysis due to the background noise associated with analogue systems. The limitation of the transient recorder is that the width of the ion pulse window, translates directly to the resultant mass spectrum to determine the t_f.

The time-to-digital converter MCP is commonly found on nESI/ESI source instruments due to lower the intensity ion signal magnitudes that reach the detector. Despite the lower dynamic range compared to the transient recorder, this variation is typically of greater use in MS/MS experiments that requires the detection of individual ion events to improve the signal to noise ratio (Chernuschevich *et al.* 2001). In further contrast to the transient recorder, the time-to-digital converter uses only the leading edge of the ion pulse to record the t_f (Chernuschevich *et al.* 2001).

The general limitation of the MCP overall is that it can record only one ion pulse window at any incidence time. This includes the dead time following ion impact, where the MCP is unable to register another ion. If further ions strike during this dead time MCP saturation can occur (Chernuschevich *et al.* 2001). Accounting for all ions in the ToF exhibiting kinetic energy values equivalent to their mass, high mass ions will impinge the detector at a lower momentum than low mass ions. As ion detection and recording events are dependent on secondary ion generation, which is positively correlated to ion momentum, the MCP can suffer from 'high-mass roll-off' reducing high mass analyte detection (Jungmann *et al.* 2013).

1.4. Ion Mobility: The Next Dimension

The theoretical basis for the calculation of the mobility of ions through the gaseous phase is was originally proposed by Langevin (1903). This theory ultimately resulted in the development of the ion-mobility (IM) method, with the first mobility measurements

presented by Bradbury (1931) (Uetrecht *et al.* 2010). The principle of the IM method, using the basic drift tube/drift cell mobility separator (Fig. 1.11), is that a pulsed packet of heterogeneous gas phase ions can be separated on the basis of their velocity through a low pressurised drift cell superimposed with an electric field (section 1.4.1) (Uetrecht *et al.* 2010).

The earliest example of a combined IM-MS instrument was first cited within the literature in the early nineteen sixties (McDaniel, Martin & Barnes 1962). Combining the classical drift cell with a Nier type 60° magnetic sector mass analyser, the instrument was used to study the kinetics of ion-molecule reactions. Later modifications of the IM-MS instrumental method would replace the magnetic sector with both ToF (McAfee, Sipler & Edelson 1967, and Edelson *et al.* 1967), and quadrupole mass analysers (Albritton *et al.* 1968). The focus of the instrumental method over the course of the next two decades however, remained on using the mass spectrometer component of the instrument to identify IM analysed components (Kanu *et al.* 2008).

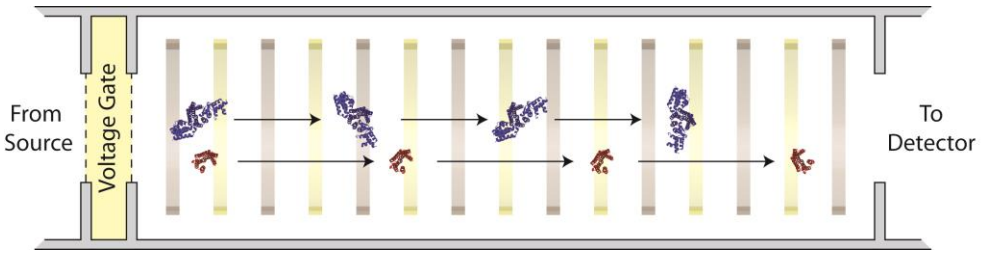

Figure 1.11. The Ion Mobility Drift Cell

The fundamental operating procedure of the drift cell is that packets of ions are pulsed into the mobility separator at time 0, controlled through the action of a voltage/ion gate, at a constant pressure (1-15 mBar) of a neutral backing gas (e.g. He or N_2) (Dugourd *et al.* 1997, and Uetrecht *et al.* 2010). Ions traverse the drift cell under the influence of a uniform weak electric field (10 V cm^{-1}) originating from metal (e.g. copper beryllium) drift guard rings (grey), typically interspersed with ceramic spacers (yellow) (Dugourd *et al.* 1997). Once ions have migrated across the mobility separator, these pass into subsequent mechanical components allowing the instrument to record drift time and mass. Weakly charged ions traverse the region less quickly than those of a higher charge state (if we maintain consistent mass and diameter), due to reduced influence of the static field. Additionally ions of increased size and shape (linear or radial) traverse the drift cell more slowly due to increased interactions with the backing gas (Thalassinos & Scrivens 2009, and Uetrecht *et al.* 2010).

Work presented by Kuk and co-workers (1989), as well as Kemper & Bowers (1990), combined an IM drift cell with a mass spectrometer to obtain shape information for a series of analyte ions, presenting information on the structure of silicon clusters and cobalt respectively. Despite the early nature of this work, it provided the cornerstone for work on far larger complexes, including: peptides, proteins, and protein complexes (section 1.4.3).

1.4.1. Calculating Velocity and Size in the Drift Cell

The velocity of an ion through a mobility region such as the drift cell (Fig. 1.11) is dependent on several factors, including: ion mass, charge, shape and density. This is in addition to drift cell conditions including the elevated pressure neutral backing gas used and the electric field. Components of smaller and/or more compact shape, at a constant ion charge state interact with the backing gas less. Subsequently these traverse the drift cell faster than ions exhibiting larger less compact shapes of equal charge (Ruotolo *et al.* 2008, and Uetrecht *et al.* 2010). The IM (K, $cm^2\ V^{-1}\ sec^{-1}$) of an analyte ion is defined in its simplest form using equation 1.13 outlined below (Kemper & Bowers 1990):

$$K = \frac{V_d}{E} \quad \text{(Eq. 1.13)}$$

V_d, represents the velocity, and E, the electric field strength (V/cm)

In practice it is customary to work with the reduced mobility (K_0) which expresses the mobility at standard temperature (T_0) in Kelvin, and pressure (P_0) in Torr (Eq. 1.14):

$$K_0 = K\frac{PT_0}{TP_0} \quad \text{(Eq. 1.14)}$$

P and T represent effective pressure (Torr) and temperature (Kelvin) respectively

Substituting T_0 and P_0 for their defined values, 273.15 Kelvin and 760 Torr respectively gives equation 1.15:

$$K_0 = K\frac{P*273.15}{T*760} \quad \text{(Eq. 1.15)}$$

The K_0 of a species can then be related to the collision cross section (CCS), by applying equation 1.16:

$$K_0 = \frac{3ze}{16N_0}\frac{1}{\Omega}\sqrt{\frac{2\pi}{\mu k_B T}} \quad \text{(Eq. 1.16)}$$

z, is the charge number, e, the electronic charge, N_0 the buffer gas number density at T_0 and P_0, μ, the reduced mass of the buffer gas and ion (Da), k_B, the Boltzmann constant, and Ω, the CCS (Å^2)

(Eq. 1.13 to 1.16 are taken from Thalassinos & Scrivens 2009)

Coupling the drift cell to a mass spectrometer subsequently enabled direct determination of the mass and charge values of an ion, required for equation 1.12. Using these values the average Ω (Ω_{avg}) can then be calculated, by relating it to the drift time (t_D), in ms, through the following equation (Eq. 1.17) (Thalassinos *et al.* 2004):

$$\Omega_{avg} = \frac{(18\pi)^{1/2}}{16}\left[\frac{1}{m}+\frac{1}{m_B}\right]^{1/2}\frac{ze}{(k_B T)^{1/2}}\frac{t_D E}{L}\frac{1}{\rho} \quad \text{(Eq. 1.17)}$$

m, is the mass of the ion in Da, m_B, the mass of the backing gas (Da) with its density represented by ρ, L the length of the cell (cm), and E, the electric field strength (V/cm).

1.4.2. The Travelling Wave Mobility Separator

The development of the travelling wave (T-wave) ion guide (Fig. 1.12) by the Waters Corporation (Manchester, UK) and outlined by Giles and co-workers (2004), allowed new IM techniques to be developed. The basic principle was to use a sequential RF phase stacked ring ion guide (SRIG) (Fig. 1.12a) to confine analyte ions. The charged analyte is then propelled through the central aperture of the ion guide using a DC potential superimposed on the RF of a single ring electrode. This potential sequentially migrates to the next ring electrode and all subsequent ring electrodes in the cell, at a user defined velocity (Giles *et al.* 2004).

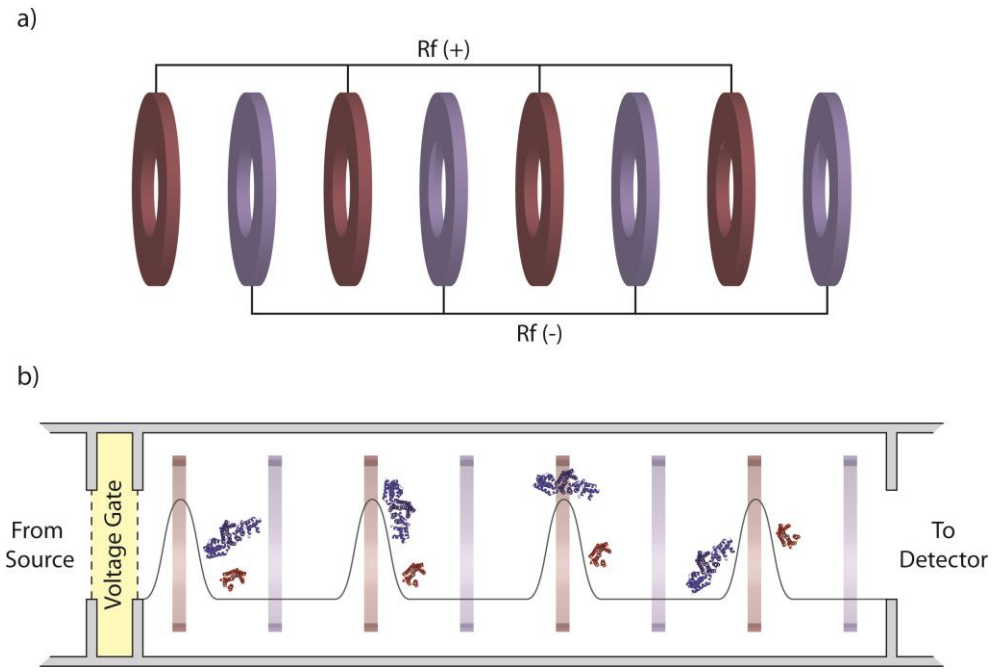

Figure 1.12. The Travelling Wave Ion Guide

The T-wave mobility separator differs from the drift cell by using a non-uniform voltage distribution across the length of the cell, in addition to a RF confining field, to separate ions. a) The non-uniform nature of the field results from a DC voltage applied to individual electrodes of the alternating RF phase SRIG. b) This voltage is applied to sequential electrodes producing a 'travelling wave' that propels ions through the mobility separator. As ions ride the crest of the DC wave, larger, more extended analytes (linear or radial), undergo increased interactions with the backing gas and are resultantly more likely to pass over this crest exiting the T-wave later (Thalassinos & Scrivens 2009, and Utrecht *et al.* 2010).

Although the idea of the RF confined T-wave ion guide to propel and separate ions was not new, having previously been applied by Lampel *et al.* (1998) and Wangler (1998), this was the first time that the SRIG setup was applied to this function (Giles *et al.* 2004). Experimentally the T-wave SRIG is a diverse experimental set up in comparison to the drift cell, and is able to act as a collision cell required for MS/MS in addition to IM separation (Giles *et al.* 2004).

Following on from the development of the T-wave, in mid-2006, the Waters corporation (Manchester UK), announced the Synapt High Definition Mass Spectrometer (HDMS, Fig. 1.13). This instrument combined the T-wave IM separator with the widely used Q-ToF Premier (Waters, Manchester, UK), producing the world's first commercially available mass analysis equipment capable of IM measurements for CCS analysis

(Waters Corporation 2006, and Pringle *et al.* 2007). Functionally, the Synapt combined three neighbouring T-wave cells, marketed as the 'TriWave.' The first and last functioned as collision cells (termed the Trap, and Transfer respectively), whilst the central under the influence of increased operating pressures provided functional T-wave IM separation (Pringle *et al.* 2007, and Thalassinos & Scrivens 2009).[8]

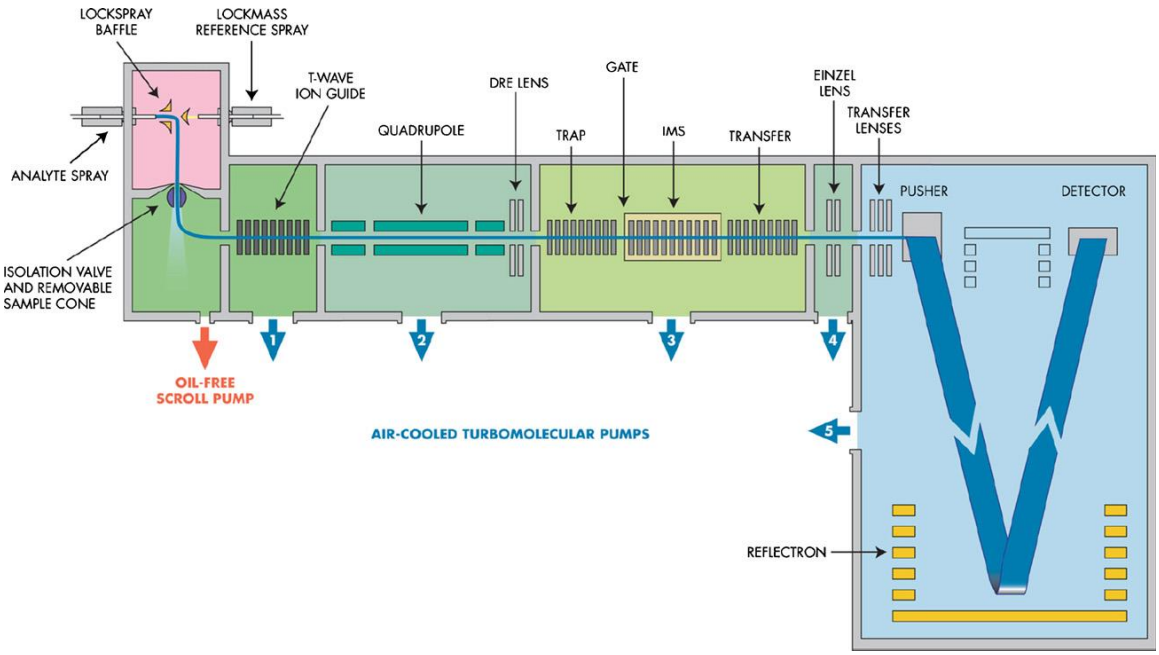

Figure 1.13. The Waters Synapt HDMS

The above schematic represents the mechanical component layout of the first generation Waters Synapt HDMS. IM t_D measurements occur as a sum of the time between the release of a packet of ions via the voltage gate (labelled 'GATE' above) and arrival at the detector within the ToF (Giles *et al.* 2004, and Pringle *et al.* 2007). For each packet of ions released by the voltage gate the orthogonal acceleration pusher (labelled 'PUSHER') 'pushes' 200 times, totalling one mass spectrum acquisition. Arrival of an ion is recorded at the detector, and its t_D expressed as the pusher number (also known as 'scan') that the ion is recorded, multiplied by the pusher period (time between pushes in ms). It should be noted however, that due to multiple packets of ions traversing the T-wave at any incidence in time, any ion(s) taking longer than the total pushes x pusher period, effectively 'roll over' to the next spectrum (Pringle et al. 2007).

Figure reproduced from Pringle *et al.* 2007

[8] Please note that since the development of the second generation Synapt HDMS, the TriWave has been modified to include a helium filled entry cell at the entrance to the mobility separator permitting increased operating pressures, suitable for improved drift time resolution.

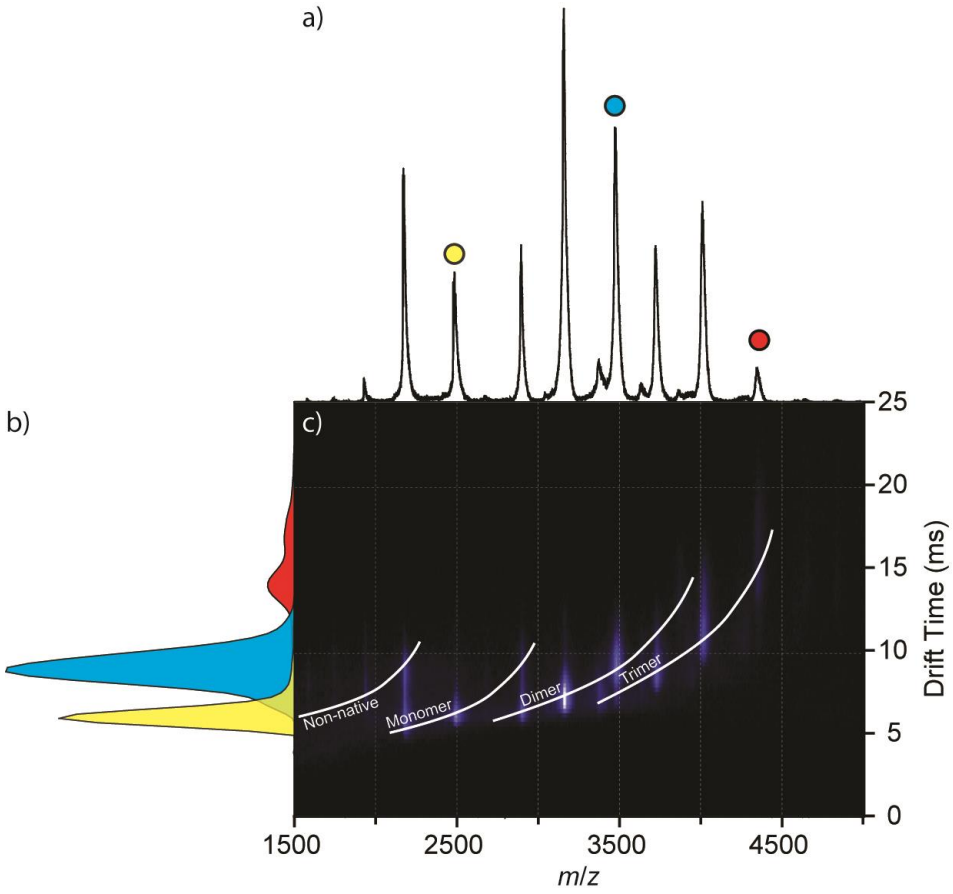

Figure 1.14. Example of the Multi-Layered Data Acquired Using the Waters Synapt HDMS

Data obtained using IM-MS methods, such as that from the Synapt HDMS, is inherently three dimensional comprising; (a) mass, (b) relative size in terms of the t_D function (converted from scan number), and intensity as a function of the dynamic range (shown as a function of relative height to the most intense peak). c) Heat map data (white = highest intensity, black = lowest) output from Driftscope (Waters, Manchester, UK), is superimposed with lines that indicate the relationship between t_D, and the oligomeric/structural state. Data depicted shows a single IM-MS spectrum of the multimeric nature of the truncated *S. cerevisiae* protein Sgt1 (amino acids 1-150) (Chapter 4). Each colour (a, b) represents the extracted t_D of the nearest native state (lowest charged ion of oligomer) of each oligomer present: yellow = monomer, blue = dimer, red = trimer.

The design of the Synapt[9] provides a number of advantages over prior drift cell systems. The TriWave is more versatile and easier to use, and the availability of a commercialised product derived from a popular and widely used instrument, is a clear improvement over the *in-house* developed instrumentation which had previously been the norm. This commercial instrumentation was further aided by an updated version of the

[9] For an in depth description of the Waters Synapt HDMS please see Pringle *et al.* (2007)

industry standard mass analysis software, MassLynx, combined with the newer IM analysis program Driftscope. Combined, these programmes enabled easy analysis of the multilayered data obtained using the Synapt instrumentation (Fig. 1.14). This is in addition to improvements in sensitivity along with, a high level of t_D reproducibility over the previous method (Thalassinos & Scrivens 2009).

In the reverse comparison, drift cells have a comparatively better t_D resolution, due to their typically higher operating pressures when compared to the T-wave. Arguably, the largest advantage of the drift cell compared to the T-wave is the ability to directly determine the CCS of analysed components (Bush *et al.* 2010, and Thalassinos & Scrivens 2009). Therefore as the T-wave is unable to directly correlate the arrival time of an ion to determine the CCS, analyte t_D requires calibration against standards of experimentally derived known cross sections.

Known standard CCS values, required for T-wave calibration, are calculated using drift cell IM-MS and the equations previously highlighted in section 1.4.1. These drift cell CCS values are presented in a range of literature, including (but not limited to): Bush *et al.* (2010), Campuzano *et al.* (2012), and Valentine *et al.* (1997). Publications of these types typically feature a range of different native and non-native proteins and peptides, over a series of ionic charge states, along with variations in mobility gas used ($N_{2(g)}$ and $He_{(g)}$) (Bush *et al.* 2010). Additionally these values are typically collated and featured in a number of open access CCS databases, including those hosted by the Bush,[10] Clemmer[11], and McLean[12] (MALDI generated ions only) research groups.

Calibration of T-wave derived experimental ion t_D to calculate the CCS requires, as with any scientific investigation, equipment parameters consistent with the analyte. This factor is essential in terms of the mobility T-wave wave height and velocity parameters, as modification of either will drastically influence the validity of the calibration plot for determining CCS from the unknown. Various methods to plot calibration curves from

[10] http://depts.washington.edu/bushlab/ccsdatabase/ (correct as of 7th August 2013)
[11] http://www.indiana.edu/~clemmer/Research/Cross%20Section%20Database/cs_database.php (correct as of 7th August 2013)
[12] http://www.vanderbilt.edu/AnS/Chemistry/groups/mcleanlab/ccs.html (correct as of 7th August 2013)

standard derived t_D values are outlined by publications, including: Ruotolo *et al.* (2008), Thalassinos *et al.* (2009), and Williams & Scrivens (2008).[13]

Work presented by Leary and co-workers (2009) has shown that these calibrations can provide comparable CCS values between two individual and independently located instruments operating under similar conditions (Leary *et al.* 2009). Current estimates of propagated errors for calculated average CCS, using careful T-wave calibration strategies, are estimated to be less than 5 % (Bush *et al.* 2010).

Additional sources of CCS error have been considered and include the collisional heating of analyte ions under the influence of increasing T-wave velocities required for mobility separation. Data presented by Morsa and co-workers (2011) indicated however that the $He_{(g)}$ and $N_{(g)}$ commonly used for mobility separation are unable to promote significant levels of ion heating that typically correspond to gains in CCS (see collision induced unfolding, section 1.4.3) (Morsa *et al.* 2011).

Due to their relevance in work presented throughout subsequent chapters, the calibration methods for calculating the CCS of multiply charged proteins from T-wave t_D values will be examined here. Taking the drift cell derived calibrant CCS (Ω) in $Å^2$, these are corrected for both the charge state and reduced mass (μ, Eq. 1.18), to generate corrected CCS (Ω') values using equation 1.19 (Ruotolo *et al.* 2008, and Williams & Scrivens 2008):

$$\mu = \frac{m_{ion} m_{gas}}{m_{ion} + m_{gas}} \qquad \text{(Eq. 1.18)}$$

$$\Omega' = \frac{\Omega}{\left[z \left(\frac{1}{\mu} \right)^{1/2} \right]} \qquad \text{(Eq. 1.19)}$$

This corrected CCS is then plotted against the $t"_D$ (*m/z* dependent corrected t_D), which is calculated using equation 1.21. In order to resolve this value however, calculation

[13] Facilities exist on the internet to allow rapid calibration curve plotting and CCS determination of analyte ions, including those hosted by the Scrivens, and Thalassinos research group websites: http://www2.warwick.ac.uk/fac/sci/lifesci/research/jscrivens/synapt_calibration/ and http://www.homepages.ucl.ac.uk/~ucbtkth/resources.html respectively (correct as of 1st November 2013, Thalassinos *et al.* 2009)

of the *m/z* independent t_D (t'_D) is required, and outlined by equation 1.20 below (Thalassinos & Scrivens 2009):

$$t'_D = t_D - (61t_m + 31t_t) \quad \text{(Eq. 1.20)}$$

t_D, is the experimental drift time extracted from the FHWM of the ion of interest in ms. t_m and t_t, are the time (ms) spent in the mobility and transfer regions respectively, calculated as the time between pairs of ring electrodes in each region (61 and 31 respectively in the Synapt HDMS). These are a function of the experimental wave velocity parameter and T-wave electrode spacing (3 mm in the Synapt HDMS).

$$t''_D = t'_D - \sqrt{\left(\frac{m/z}{1000}\right)} * 0.085 \quad \text{(Eq. 1.21)}$$

0.085, is the sum of the ToF flight time (44 μs), and transit time (41 μs) in ms at *m/z* 1000.

The resultant plot of Ω' against t''_D is subsequently used to calculate the CCS of the unknown analyte ion applying a suitable trend line (Fig. 1.15). Choice of trend line function is typically based on research presented by Shvartsburg & Smith (2008) with multiply charged calibrant protein ions (relevant here) plotted using a non-linear or power fit trend line. In contrast singly or doubly charged peptide calibrants are plotted using a linear fit trend line. This plot can be additionally applied to calculate the difference between the CCS derived from the fitted trend line against the published standard value, allowing for improved calibration accuracy by removing outlying points (Thalassinos *et al.* 2009).

Further experimental refinements require that the selection of calibrant(s) closely bracket the resultant analyte t_D acquired using the T-wave, as well apparatus conditions maximising the coefficient of determination (R^2) fitting of the calibration curve, to provide relevant, high accuracy CCS values (Salbo *et al.* 2012). Failure to adapt to either of these pre-requisites may result in large calibration errors. This is especially true if extrapolating curve fitting from calibrants that bracket t_D values well below that of the experimental analyte, impairing the experimental validity (Bush *et al.* 2010, Shvartsburg & Smith 2008, and Thalassinos & Scrivens 2009).

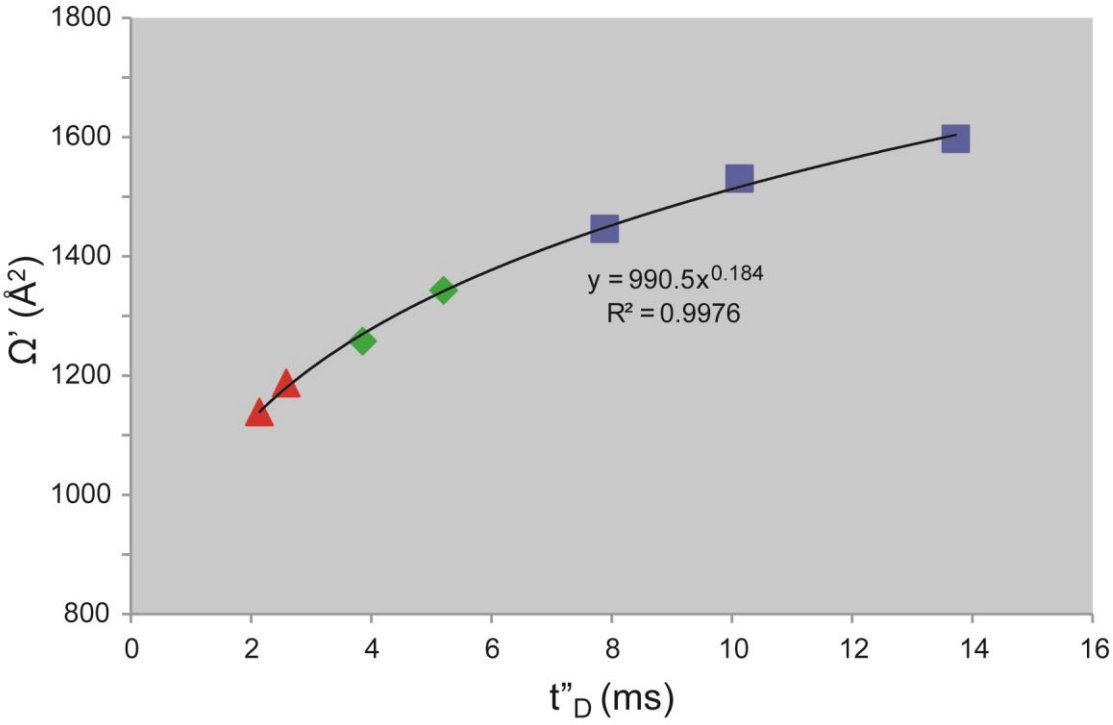

Figure 1.15. Example Calibration Curve Required for T-wave CCS Analysis

Example calibration curve drawn using denatured equine myoglobin (red), in combination with native bovine serum albumin (green), and *S. cerevisiae* alcohol dehydrogenase (blue) (all calibrants purchased from Sigma-Aldrich Ltd, Dorset UK). These are calibrated using CCS values derived from Bush *et al.* (2010), at a constant wave height of 9 V, travelling at 300 m/s. R^2 represents the agreement between the calibration curve and all of the calibrant points contributing to its gradient value (y).

1.4.3. Applying the "Next Dimension"

To date, IM-MS has been applied in a broad range of experimental methods including studies of conformation stability and gas phase separation. Proteomic and metabolomic studies have applied the mobility separator to decongest spectra peaks. Peptide ions between 500 to 2500 Da typically produce linear mass/ion mobility plots at a constant charge state, H^+ or H_2^+ (Shvartsburg & Smith 2008, Valentine *et al.* 1999). Therefore IM-MS allows separation of ions that share the same *m/z*, but exhibit different structural conformations, including other molecular classes that may be present such as lipids or MALDI matrices, due to differences in linear mass/mobility plots. This method allows multidimensional separations of complex analyte mixtures, including *in situ* tissue analysis

using mass spectrometry imaging methods (Chughtai & Heeren 2010, and Jackson *et al.* 2007).

In addition to separating ions of the same *m/z*, but different structural conformations, instrumental set-ups such as the TriWave promoted the development of new experimental methods that allowed direct analysis of protein folding and stability dynamics. One commonly applied method is that of collision induced unfolding (CIU), first used to study structural folding intermediates by Ruotolo and co-workers (2007). Combining T-wave IM-MS with increasing injection energies, this early CIU investigation was used to study the folding stability and conformations along the unfolding pathway of human transthyretin (TTR).

Subsequent refinements to the CIU method applied the quadrupole to isolate a precursor ion, prior to collisional activation using the TriWave Trap region (Fig. 1.13) to induce unfolding. Analysis of the structural state is subsequently performed using the mobility separator, to determine changes in CCS (Fig. 1.16) (Hopper & Oldham 2009). The advantage of this method is that an investigator can rapidly target and analyse the desired ion (typically the highest charged ion in the native CSD), studying its gas phase stability and conformational states with direct relationship to the collision energies (Han *et al.* 2011). Applications of the method within the literature include the structural modulation, and conformational influences of binding small ligands (Hopper & Oldham 2009), and anions (Han *et al.* 2011).

What is undoubtedly the largest advantage provided by IM-MS analysis, is the ability to use the calculated CCS values, to aid *de novo* structural determination of an analyte using computational methods (see chapter 5). This is as well as providing corroboratory evidence for X-ray crystallography (see chapter 4) and nuclear magnetic resonance (NMR, see chapter 3) results, without the imposed limitations of these respective methods.

To allow comparison of IM calculated cross section data with protein database (PDB)[14] structures, derived using X-ray and NMR based techniques, software such as MobCal (Mesleh *et al.* 1996, and Shvartsburg & Jarrold 1996) is required to convert these structures into compatible CCS values. To produce these values MobCal applies a variety

[14] http://www.rcsb.org/pdb/home/home.do (correct as of 30th May 2013)

of models to calculate structure derived CCS, including: exact hard sphere scattering, (EHSS, section 1.4.3.3), projection approximation (PA, section 1.4.3.1), and the trajectory method (TM, section 1.4.3.2). Each of these methods, that have their own strengths and weaknesses, are explored in brief below along with the newly developed projection superposition approximation (PSA) model.

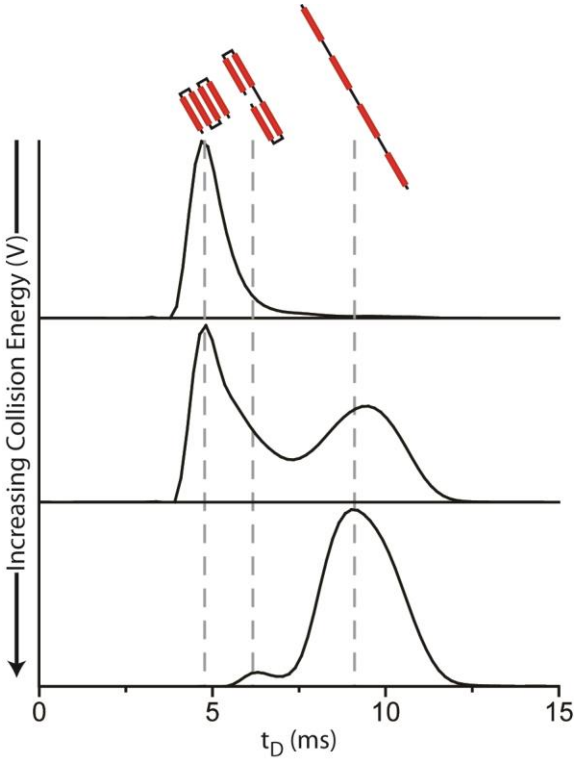

Figure 1.16. Example of Collision Induced Unfolding to Study Protein Stability

Study of the structural conformation of the monomeric +8 ion of a truncated *S. cerevisiae* protein (Sgt1, amino acids 1-178), using CIU and increasing collision energies (18, 27, and 36 V). Image indicates a shift in t_D indicative of protein unfolding, with at least two stable, partially unfolded, conformations. A cartoon representation of the protein structure is depicted in red, to represent the unfolding that contributes to these increases in t_D. Please note that these cartoons are not representative of the structural nature of the Sgt1 protein, and are for illustrative purposes only.

1.4.3.1. Calculating Collision Cross Sections Using the Projection Approximation Model

The simplest method to calculate CCS from PDB files is the PA (Fig. 1.17) (Mesleh *et al.* 1996). This model provides an average cross section based on the sum of all 'shadows' or 'projections' created by a protein orientated about the x and y axis (Scarff *et al.* 2008, and Scarff *et al.* 2009). Typically Monte Carlo scoring is applied to allow determination of the projected area in any given x/y orientation (Bleiholder *et al.* 2011).

Limitations of the PA occur at the equation level as it does not account for the long range electrostatic interactions, convex surface features and scattering of the backing gas.

Consequently the PA underestimates the CCS of larger analytes with complex collision geometries (Mesleh *et al.* 1996, Scarff *et al.* 2008, and Thalassinos & Scrivens 2009). Observations, performed by Scarff *et al.* (2008), on experimental T-wave derived data compared against theoretical CCS calculations obtained using PDB structures support these conclusions.

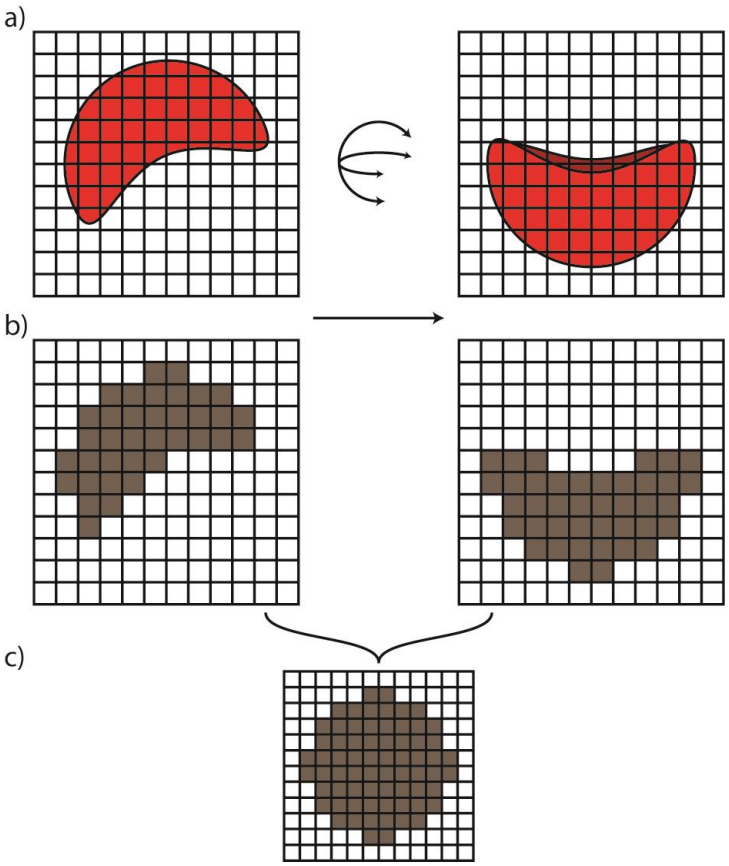

Figure 1.17. Representation of the Projection Approximation Model for CCS Calculation Using PDB Data

The PA calculates CCS values from NMR and X-ray derived PDB data. a) The three dimensional PDB structure is rotated about the x and y axis. b) At each change in x and y, a new Monte Carlo hit test is performed producing an effective shadow of the structure that is equal to the sum of radii between the backing gas and analyte structure. c) All resultant hit test scores are summed to produce an overall CCS shadow.

1.4.3.2. Calculating Collision Cross Sections Using the Trajectory Method Model

The TM (Fig. 1.18), has been shown to produce the most accurate PDB derived CCS estimates, especially for larger analytes (Scarff *et al.* 2008). This model, first outlined by Mesleh *et al.* (1996), takes into account all interactions between the analyte and backing gas ($He_{(g)}$) across the orientationally-averaged CCS (Campuzano *et al.* 2012). These analyte-backing gas interactions include: collision geometries (i.e. angle of deflection), and the effect of convex features, on calculated CCS values (Fig. 1.18a). Additionally the TM

calculates the influence of long range electrostatic interactions, which do not result in direct 'hits' under the orientationally averaged CCS Monte Carlo method (Fig. 1.18b) (Hopper & Oldham 2009, Mesleh *et al.* 1996, and Campuzano *et al.* 2012). Calculating the contributions of these long range interactions limits the application of the TM for CCS studies due to the computational demands required to perform them.

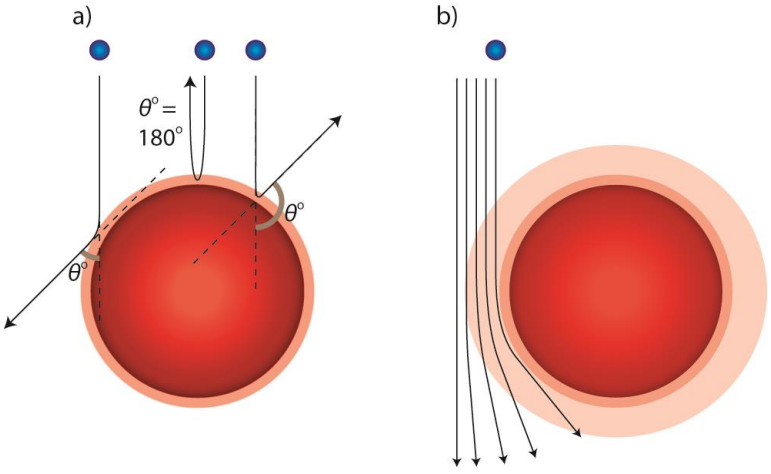

Figure 1.18. Representation of the Trajectory Method for CCS Calculation Using PDB Data

a) Applying the hard sphere collision model, the TM calculates for the angle of deflection (θ°) resultant from direct backing gas (blue sphere) interactions with the analyte (red sphere) superimposed with a 'hard sphere collision limit' (pale red). b) The computational drain of the TM occurs as a result of calculating long range electrostatic interactions between the analyte charge cloud on both the analyte and backing gas (pink). The EHSS enhances computational efficiency by removing this step from calculations, whilst providing accurate CCS measurements within a few percent of the TM (Scarff *et al.* 2008).

1.4.3.3. Calculating Collision Cross Sections Using the Exact Hard Sphere Scattering Model

Representing a compromise between the accuracy of the TM, and the speed of the PA models, the EHSS (Shvartsburg & Jarrold 1996) ignores the long range electrostatic interaction calculations of the TM (highlighted in Fig. 1.18b).[15] Comparatively, the removal of these calculations from the model reduces accuracy only a few percent when compared to the TM (Jarrold 1999, and Scarff *et al.* 2008).

[15] These interactions are quoted to contribute "less than 10 %," to the sum of all interactions of the fullerenes studied by Mesleh *et al.* (1996) when outlining the development of the TM.

The EHSS model superimposes a 'hard sphere collision limit,' equal to the combined radii of the analyte and backing gas (Fig. 1.18a) on the structure. When this limit is surpassed by the buffer gas it is treated as a contact moment, with measurements of the scattering/deflection angle (Fig. 1.18a), counting towards the CCS value calculated by MobCal (Jarrold 1999). Furthermore, this model accounts for structural features such as convex geometries which increase drag and reduce gas phase mobility (Bleiholder, Wyttenbach & Bowers 2011).

1.4.3.4. Calculating Collision Cross Sections Using the Projection Superposition Approximation Model

The recently developed PSA (Bleiholder, Wyttenbach & Bowers 2011, Bleiholder et al. 2013, and Anderson *et al.* 2012), acts as a modified version of the PA highlighted previously. Compared to the PA the PSA additionally accounts for the collective shape and size effects of analyte and backing gas, whilst remaining computationally efficient (Bleiholder, Wyttenbach & Bowers 2011). In contrast to the other models previously highlighted, the PSA is only available in the modelling program Sigma (von Helden *et al.* 1993, Wyttenbach *et al.* 1997, and Wyttenbach, Witt & Bowers 2000).

Unlike the models above, the PSA replaces the idea of a hard sphere shell with a collision probability value, dependent on distance as well as the size of both the analyte and selected collision gas (Fig. 1.19). Combined with calculations of shape modulation influences on the t_D (also accounted for in the TM and EHSS), the PSA provides a greater accuracy compared to the PA from which it is derived (Bleiholder, Wyttenbach & Bowers 2011).

In-house analysis of the model, by the Bowers research group (Anderson *et al.* 2012), has found that application of the PSA to study small supramolecular assemblies with complex shape factors was in high agreement with experimental CCS values. Compared to PA and TM, the PSA has been shown to provide improved agreement between the theoretical and experimentally calculated CCS values. This is in addition to being almost a "factor of 10 faster" than the TM (Anderson *et al.* 2012).

Follow up work by Wyttenbach, Bleiholder & Bowers (2013), applied the PSA to study a range of ions of greater mass (130 Da to 800 kDa) and size. Research concluded with several observations outlined, at both the level of the PSA itself and buffer gas interactions. With regards to the latter, data agreed with previous observations by von Helden and co-workers (1991) that relative contributions of buffer gas to the obtained t_D decrease with ions of increasing size (nearly negated by $1x10^6$ Da), and is of importance when studying small peptides (Wyttenbach, Bleiholder & Bowers 2013).

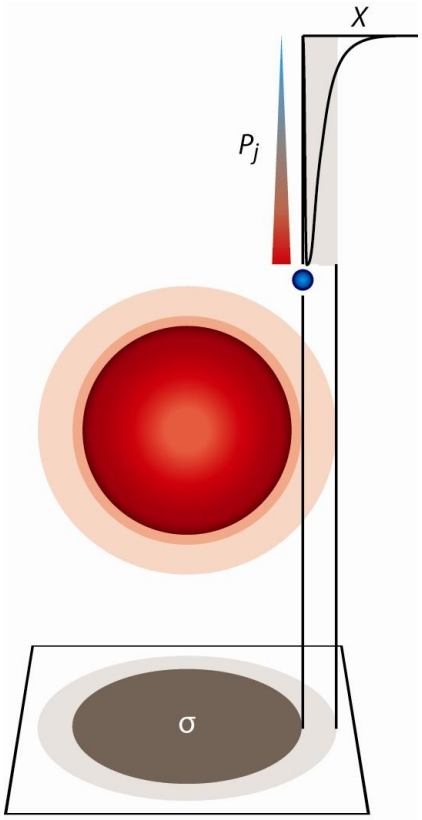

Figure 1.19. Representation of the Projection Superposition Approximation Model for CCS Calculation Using PDB Data

Image represents a visualised depiction of the collision probability (*Pj*) for a chosen random point on the analyte, perpendicular to the plane of interception. Increasing temperatures (*T*), and distances (*X*) decrease the probability of interaction between the backing gas (blue) and the analyte (red, with hard sphere collision limit). This probability decreases exponentially with distance, as long range electrostatic interactions (pink) are gradually negated and the influence of the backing gas lessens. Below a certain distance however ($X < 0$) the hard sphere collision limit is surpassed and results in direct interactions that count directly toward to the projection cross section (σ).

Of significant importance in future IM-MS studies are the observations that Wyttenbach, Bleiholder & Bowers (2013) highlight between the He acquired t_D values, compared to those obtained using N_2. These are of concern due to an increased interaction between the analyte and $N_{2(g)}$ occurring due to its increased diameter compared to $He_{(g)}$ (~4x increase in interaction effects). Therefore ions of larger size, exhibiting increased levels of surface concaveness, cannot be accurately predicted based on their formula whilst using N_2 to correct for the experimental CCS (Wyttenbach, Bleiholder & Bowers 2013).

The wider applications of results presented using this model in the overall field of IM analysis, are however too early to dictate until further observations are published.

1.5. Confirming the Biological Relevance of Gas Phase Structures

Prior to the inception of the soft ionisation sources (section 1.2) publications of native-like mass spectra from biological materials were rare occurrences, these resulted due the non-favourable conditions required for ionisation. Upon the development of the MALDI and ESI soft ionisation sources, large multimeric complexes could be successfully transferred intact into the gas phase for mass analysis. The combination of mass spectrometers with IM instrumentation for structural analysis subsequently meant that the validation of gas phase biological species took a massive step forward. Before the turn of the 21st century however the viability of native mass spectrometry and IM-MS to produce biologically relevant results had been questioned. Some individuals doubted or remained undecided on the ability of mass spectrometry to render the appropriate conditions required for gas-phase analysis (Loo 1997).

Work performed by Pace and co-workers (2004), supported by earlier non-mass spectrometric based observations performed by Wolynes (1995), indicated that in a vacuum intermolecular interactions such as electrostatics were sufficient to stabilise the fold of the small globular protein, RNase Sa (*S. aureofaciens*). In the absence of water however, hydrogen bonds and van der Waals interactions are weakened, with the structure primarily stabilised through electrostatic interactions. Net protein charge is therefore a factor in gas phase stability, and exceeding a threshold value stabilising electrostatic interactions are compromised and protein unfolding occurs (Pace *et al.* 2004).

Later work by other individuals would expand upon these observations and support them at the level of the multimeric protein assemblies. Ruotolo and co-workers (2005), provided evidence to support the existence of native macromolecular structures in the gas phase. Studying the trp RNA binding protein, TRAP (large multimeric ring structure), using drift cell IM analysis, experimenters were able to indicate that the lowest charged ion in the CSD produced a CCS in strong agreement with previous X-ray derived structural data. Increases in charge state showed a transition state between native macromolecular

states, to a non-native collapsed state (Ruotolo *et al.* 2005), and supports the observations by Pace and co-workers (2004) above.

As these higher charge states are increasingly likely to provide biologically irrelevant CCS values, the work supports the use of the 'nearest native' charge state for IM t_D analysis, i.e. lowest charged analyte ion in the CSD of interest. This is further supported by extended work performed by Scarff and co-workers (2009) who presented a positive correlation between the proteins charge state and increasing, less native, CCS. These changes in CCS are concluded to occur as an influence of increasing Coulomb repulsion at the higher charge states (Scarff *et al.* 2009).

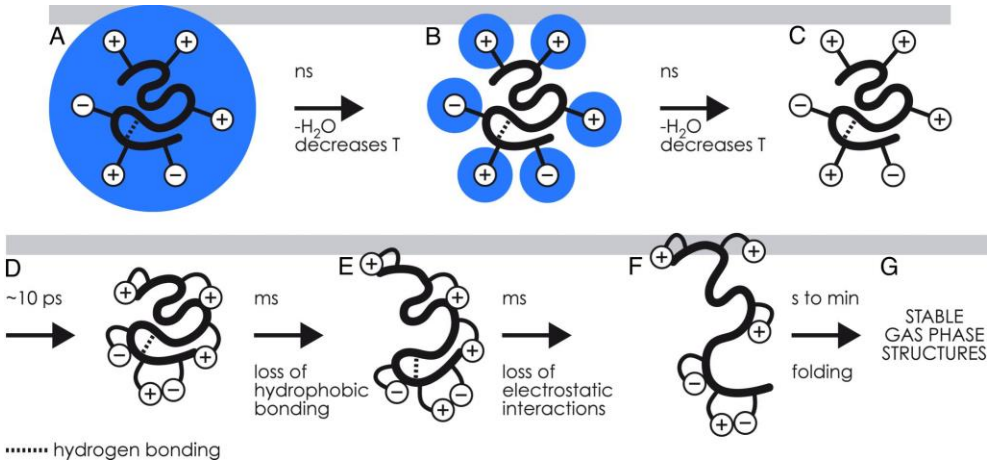

Figure 1.20. Stepwise Gas Phase Evolution: Desolvation to Gas Phase Stabilization

Post ESI globular proteins undergo a series of structural evolution events dependent on desolvation and intramolecular binding loss/gain events. A) The ionised sample enters the gas phase within a solvent shell. B) This ionised sample desolvates over the ns timescale until only exterior ionic functional groups remain encapsulated within the solvent. C) Further desolvation leads to an effectively dry sample allowing the functional ionic groups' to collapse over the picosecond timescale. D) This results in a 'near-native' state that over an extended ms timescale suffers a loss of (E) hydrophobic and (F) electrostatic bonding, producing a transient unfolded state. G) Over an extended period of time, not typically of IM-MS analysis in instrumentation such as the Synapt HDMS, these transient, unfolded states may refold to produce stable gas phase structures (Brueker & McLafferty 2008). 'T' represents temperature.

Figure reproduced from Brueker & McLafferty 2008

Using questions highlighted by Loo (1997), with regards to the stability and relevance of biological materials in the gas phase, Breuker & McLafferty (2008) proceeded

to ask; "for how long, under what conditions, and to what extent can solution structure be retained without solvent?" Combination of experimental and computational analysis allowed the authors to produce a time frame evolution of sample desolvation and gas phase stability modifications (Fig. 1.20). The work performed concluded that with regards to the original question, asked above, solution structure is eventually lost on the post ms time scale. This therefore confirms the biological validity of IM derived CCS values which occur on the sub ms time scale.

Additional considerations with regards to the stability of gas phase ions and their native-like nature are further outlined by Tahallah and co-workers (2001) on their studies of oligomeric species using increasing source pressures. Increasing source pressures showed a direct correlation with the detection of higher *m/z* ions. This is accounted as a feature of differential collisional cooling effects on ions of increasing mass, in tandem with reduced desolvation represented by decreasing peak resolution. This decreased desolvation would hypothetically stabilise hydrophilic interactions within the gas phase for an extended period of time. Additionally their observations further outline the use of suitable ionisation conditions to prevent excessive source dissociation, stabilising the biologically relevant state (Tahallah *et al.* 2001).

Subsequently, with the above inferring biological validity of results obtained under suitable conditions, studies of gas phase protein complexes using mass spectrometry and IM-MS are becoming increasingly commonplace in the field of structural biology. These studies apply the combined speed, mass sensitivity, and structural resolution, of a single instrumental method. Work presented in the following chapters has applied a range of both gas and solution phase methods, to probe the oligomeric and structural conformations of three proteins required within their respective organisms, to ensure viability.

1.6. References

Albritton, D., *et al.*, 1968 "Mobilities of Mass-Identified H_3^+ and H^+ Ions in Hydrogen," Physical Review, Vol.171 pp.94-102

Anderson, S., *et al.*, 2012 "A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (III): Application to Supramolecular Coordination-Driven Assemblies with Complex Shapes," International Journal of Mass Spectrometry, Special Issue, Vol.330-332 pp.78-84

Aston, F.W., 1919 "A Positive Ray Spectrograph," Philosophical Magazine, Series 6, Vol.37 pp.707-714

Aston, F.W., 1920a "The Constitution of Atmospheric Neon," Philosophical Magazine, Series 6, Vol.39 pp.449-455

Aston, F.W., 1920b "The Mass-Spectra of Chemical Elements," Philosophical Magazine, Series 6, Vol.39 pp.611-625

Aston, F.W., 1937 "A Second-Order Focusing Mass Spectrograph and Isotopic Weights by the Doublet Method" Proceedings of the Royal Society, Vol.163A pp.391-404

Audi, G., 2006 "The History of Nuclidic Masses and of the their Evaluation," International Journal of Mass Spectrometry, Vol.251 pp.85-94

Benesch, J.L.P., *et al.* 2006 "Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies," Chemistry & Biology, Vol.13 pp.597-605

Beynon, J.H., *et al.*, 1973 "Instrumentation Design and Performance of a Mass-analyzed Ion Kinetic Energy (MIKE) Spectrometer," Analytical Chemistry, Vol.45 pp.1023-1031

Biemann, K., 1990 "Sequencing Peptides by Tandem Mass Spectrometry and High-Energy Collision-Induced Dissociation," Methods in Enzymology, Vol.193 pp.455-479

Blaum, K., 2006 "High-Accuracy Mass Spectrometry with Stored Ions," Physics Reports, Vol.425 pp.1-78

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, Vol.308 pp.1-10

Bleiholder, C., *et al.*, 2013 "A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (II): Model Parameterization and Definition of Empirical Shape Factors for Proteins," International Journal of Mass Spectrometry Special Edition, Vol.345-347 pp.89-96

Bradbury, N., 1931 "The Mobility of Aged Ions in Air in Relation to the Nature of Gaseous Ions," Physical Review, Vol.37 pp.1311-1319

Breuker, K., & McLafferty, F.W., 2008 "Stepwise Evolution of Protein Native Structure with Electrospray into the Gas Phase, 10^{-12} to 10^{-2} s" Publications of the National Academy of Sciences, Vol.105 pp.18145-18152

Budzikiewicz, H., & Grigsby, R., 2006 "Mass Spectrometry and Isotopes: A Century of Research and Discussion," Mass Spectrometry Reviews, Vol.25 pp.146-157

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, Vol.82 pp.9557-9565

Cameron, A.E., & Eggers, D.F., 1948 "An Ion 'Velocitron,'" The Review of Scientific Instruments, Vol.19 pp.605-607

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, Vol.84 pp.1026-1033

Chowdhury, S., Katta, V., & Chait, B.C., 1990 "Probing Conformational Changes in Proteins by Mass Spectrometry," Journal of the American Chemical Society, Vol.112 pp.9012-9013

Chernuschevich, I.V., Loboda, A.V., & Thomson, B.A., 2001 "An Introduction to Quadrupole-Time-of-Flight Mass Spectrometry," Journal of Mass Spectrometry, Vol.36 pp.849-865

Chughtai, K., & Heeren, R., 2010 "Mass Spectrometric Imaging for Biomedical Tissue Analysis," Chemical Reviews, Vol.110 pp.3237-3277

Cole, R.B., 2000 "Some Tenets Pertaining to Electrospray Ionization Mass Spectrometry," Journal of Mass Spectrometry, Vol.35 pp.763-772

Collings, D., & Douglas, D., 1997 "An Extended Mass Range Quadrupole for Electrospray Mass Spectrometry," International Journal of Mass Spectrometry, Vol.162 pp.121-127

Cooks, R.G., Beynon, J.H., & Litton, J.F., 1975 "Ion structure determination through collisions of ions of high kinetic energy," Organic Mass Spectrometry, Vol.10 pp.503-506

Daniel, J.M., *et al.* 2002 "Quantitative Determination of Noncovalent Binding Interactions Using Soft Ionization Mass Spectrometry," International Journal of Mass Spectrometry, Vol.216 pp.1-27

Dehmelt, H.G., 1968 "Radiofrequency Spectroscopy of Stored Ions I: Storage," Advances in Atomic and Molecular Physics, Vol.3 pp.53-72

de Hoffmann, E., 1996 "Tandem Mass Spectrometry: a Primer," Journal of Mass Spectrometry, Vol.31 pp.129-137

de Hoffmann, E., & Stroobant, V., 2007 "Mass Spectrometry: Principles & Applications," Third Edition, John Wiley & Sons Ltd, (Chichester, UK)

De Laeter, J.R., 1996 "The Role of Off-Line Mass Spectrometry in Nuclear Fission," Mass Spectrometry Reviews, Vol.15 pp.261-281

De La Mora, J.F., 2000 "Electrospray Ionisation of Large Multiply Charged Species Proceeds via Dole's Charged Residue Mechanism," Analytica Chimica Acta, Vol.406 pp.93-104

Dole, M., at. al., 1968 "Molecular Beams of Macroions," The Journal of Chemical Physics, Vol.49 pp.2240-2249

Domon, B., & Aebersold, R., 2006 "Mass Spectrometry and Protein Analysis," Science, Vol.312 pp.212-217

Douglas, D., 2009 "Linear Quadrupoles in Mass Spectrometry," Mass Spectrometry Reviews, Vol.28 pp.937-960

Dreisewerd, K., 2003 "The Desorption Process in MALDI," Chemical Reviews, Vol.103 pp.395-426

Dugourd, Ph., *et al.*, 1997 "High-Resolution Ion Mobility Measurements," Review of Scientific Instruments, Vol.68 pp.1122-1129

Edelson, D., *et al.*, 1967 "Interpretation of Ion-Mobility Experiments in Reacting Systems," Physical Review, Vol.164 pp.71-75

El-Aneed, A., Cohen, A., & Banoub, J., 2009 "Mass Spectrometry, Review of the Basics: Electrospray, MALDI and Commonly Used Mass Analysers," Applied Spectroscopy Reviews, Vol.44 pp.210-230

Emmet, M., & Caprioli, R., 1994 "Micro-Electrospray Mass Spectrometry: Ultra-High-Sensitivity Analysis of Peptides and Proteins," American Society for Mass Spectrometry, Vol.5 pp.605-613

Felitsyn, N., Peschke, M., & Kebarle, P., 2002 "Origin and Number of Charges Observed on Multiply-Protonated Native Proteins Produced by ESI," International Journal of Mass Spectrometry, Vol.219 pp.39-62

Fenn, J.B., *et al.*, 1989 "Electrospray Ionization for Mass Spectrometry of Large Biomolecules," Science, Vol.246 pp.64-71

Gale, D.C., & Smith, R.D., 1993 "Small Volume and Low Flow-Rate Electrospray Ionisation Mass Spectrometry of Aqueous Samples," Rapid Communications in Mass Spectrometry, Vol.7 pp.1017-1021

Giles, K., *et al.*, 2004, "Applications of a travelling wave-based radio-frequency-only stacked ring ion guide," Rapid Communication is Mass Spectrometry, Vol.18 pp.2401-2414

Gomez, A., & Tang, K., 1994 "Charge & Fission of Droplets in Electrostatic Sprays," Physics Fluids, Vol.6 pp.404-414

Gross, J.H., 2004 "Mass Spectrometry: A Textbook," Springer-Verlag, Berlin, Germany

Griffiths, J., 2008 "A Brief History of Mass Spectrometry," Analytical Chemistry, Vol.80 pp.5678-5683

Guilhaus, M., 1995 "Principles and Instrumentation in Time-of-Flight Mass Spectrometry Physical and Instrumental Concepts," Journal of Mass Spectrometry, Vol.30 pp.1519-1532

Hager, D.B., *et al.*, 1994, "Droplet Electrospray Mass Spectrometry," Analytical Chemistry, Vol.66 pp.3944-3949

Han, L., *et al.*, 2011 "Bound Anions Differentially Stabilize Multiprotein Complexes in the Absence of Bulk Solvent," Journal of the American Chemical Society, Vol.133 pp.11358-11367

Hayes, R.N., & Gross, M.L., 1990 "Collision-Induced Dissociation," Methods in Enzymology, Vol.193 pp.237-263

Heck, A.J.R., & van den Heuvel, R.H.H., 2004 "Investigation of Intact Protein Complexes by Mass Spectrometry," Mass Spectrometry Reviews, Vol.23 pp.368-389

Hernández, H., *et al.*, 2006 "Subunit Architecture of Multimeric Complexes Isolated Directly from Cells," EMBO Reports, Vol.7 pp.605-610

Hilton, G.R., & Benesch, J.L.P., 2012 "Two Decades of Studying Non-Covalent Biomolecular Assemblies by Means of Electrospray Ionization Mass Spectrometry," Journal of the Royal Society Interface, Vol.9 pp.801-816

Hopper, J., & Oldham, N., 2009 "Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability," Journal of the American Society for Mass Spectrometry, Vol.20 pp.1851-1858

Iribarne, J.V., & Thomson, B.A., 1976 "On the Evaporation of Small Ions from Charged Droplets," The Journal of Chemical Physics, Vol.64 pp.2287-2294

Jackson, S.N., *et al.* 2007 "MALDI-Ion-Mobility-TOFMS Imaging of Lipids in Rat Brain Tissue," Journal of Mass Spectrometry, Vol.42 pp.1093-1098

Jarrold, M.F., 1999 "Unfolding, Refolding and Hydration of Proteins in the Gas Phase," Accounts of Chemical Research, Vol.32 pp.360-367

Jennings, K.R., 1968 "Collision-Induced Decompositions of Aromatic Molecular Ions," International Journal of Mass Spectrometry Ion Physics, Vol.1 pp.227-235

Jennings, K.R., 2000 "The changing impact of the collision-induced decomposition of ions on mass spectrometry," International Journal of Mass Spectrometry, Vol.200 pp.479-493

Johnson, J., & Yost, R., 1991 "Tandem-in-Space and Tandem-in-Time Mass Spectrometry: Triple Quadrupoles and Quadrupole," Analytical Chemistry, Vol.62 pp.2162-2172

Juraschek, R., Dülcks, T., & Karas, M., "Nanoelectrospray - More than just a Minimized-Flow Electrospray Ionization Source," American Society for Mass Spectrometry, Vol.10 pp.300-308

Kanu, A., *et al.* 2008 "Ion Mobility-Mass Spectrometry," Journal of Mass Spectrometry, Vol.43 pp.1-22

Karas, M., & Hillenkamp, F., 1988 "Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons," Analytical Chemistry, Vol.60 pp.2299-2301

Karas, M., Bahr, U., & Dülcks, T., 2000 "Nano-Electrospray Ionization Mass Spectrometry: Addressing Analytical Problems Beyond Routine," Fresenius Journal of Analytical Chemistry, Vol.366 pp.669-676

Karas, M., & Krüger, R., 2003 "Ion Formation in MALDI: The Cluster Ionization Mechanism," Chemical Reviews, Vol.103 pp.427-440

Katta, V., & Chait, B., 1991 "Observation of the Haemoglobin Complex in Native Myoglobin by Electrospray-Ionization Mass Spectrometry," Journal of the American Chemical Society, Vol.113 pp.8534-8535

Kebarle, P., & Peschke, M., 2000 "On the Mechanisms by which the Charged Droplets Produced by Electrospray Lead to Gas Phase Ions," Analytica Chimica Acta, Vol.406 pp.11-35

Kebarle, P., & Verkerk, U.H., 2009 "Electrospray: From Ions in Solution to Ions in the Gas Phase, What we Know Now," Mass Spectrometry Reviews, Vol.28 pp.898-917

Kemper, P.R., & Bowers, M.T., 1990 "A Hybrid Double-focusing mass spectrometer - High pressure drift reaction cell to study thermal energy reactions of mass-selected ions," Journal for the American Society for Mass Spectrometry, Vol.1 pp.197-207

Kitova, E., *et al.*, 2012 "Reliable Determinations of Protein-Ligand Interactions by Direct ESI-MS Measurements. Are we there yet?" Journal of the American Society for Mass Spectrometry, Vol.23 pp.431-441

Knochenmuss, R., & Zenobi, R., 2003 "MALDI Ionisation: the Role of In-Plume Processes," Chemical Reviews, Vol.103 pp.441-452

Kruger, T., *et al.*, 1976 "Mixture Analysis by Mass-Analyzed Ian Kinetic Energy Spectrometry," Analytical Chemistry, Vol.48 pp.2113-2119

Kuk, Y., *et al.* 1989 "Preparation and Observation of Silo Clusters on a Au(001) -(5 x 20) surface," Physical Review B, Vol.39 pp.168-171

Lampel, M.C., 1998 "A Self-Consistent Beam Loaded Travelling Wave Accelerator Model for use in Trace 3-D," Proceedings of the 1997 Particle Accelerator Conference, Vol.1-3 pp.2615-2617

Langevin, P., 1903 "The Ionisation of Gases," Annales De Chimie Et De Physique, Vol.28 pp.289-384

Last, I., Levy, Y., & Jortner, J., 2002 "Beyond the Rayleigh Instability Limit for Multicharged Finite Systems: from Fission to Column Explosion," Proceedings of the National Academy of Sciences, Vol.99 pp.9107-9112

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, Vol.20 pp.1699-1709

Loo, J.A., 1997 "Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry," Mass Spectrometry Reviews, Vol.16 pp.1-23

Luedtke, W., *et al.*, 2008 "Nanojets, Electrospray, and Ion Field Formation: Molecular Dynamics Simulators and Laboratory Experiments," Journal of Physical Chemistry, Vol.112 pp.9628-9649

Macha, S.F., McCarley, T.D., & Limbach, P.A., 1999 "Influence of Ionization Energy on Charge-Transfer Ionization in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry," Analytica Chimica Acta, Vol.397 pp.235-245

Mamyrin, B.A., *et al.*, 1973, "Mass-Reflectron a New Non-Magnetic Time-of-Flight High-Resolution Mass-Spectrometer," Zhurnal Eksperimentalnoi I Teoreticheskoi Fiziki, Vol.64 pp.82-89

Mann, M., Meng, C.K., & Fenn, J.B., 1989 "Interpreting Mass Spectra of Multiply Charged Ions," Analytical Chemistry, Vol.61 pp.1702-1708

March, R., 2009 "Quadrupole Ion Traps," Mass Spectrometry Reviews, Vol. 28 pp.961-989

Marshall, A., Hendrickson, S., & Shi, S.D.-H, "High-Resolution FT-ICRMS," Analytical Chemistry, Vol.74 pp.252-259

McAfee, K.B., Sipler, D., & Edelson, D., 1967 "Mobilities and Reactions of Ions in Argon," Physical Review, Vol.160 pp.130-135

McCarley, T., McCarley, R., & Limbach, P., 1998 "Electron-Transfer Ionization in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry" Analytical Chemistry, Vol.70 pp.4376-4379

McDaniel, E.W., Martin, D.W., & Barnes, W.S., 1962 "Drift Tube-Mass Spectrometer for Studies of Low-Energy Ion-Molecule Reactions," Review of Scientific Instruments, Vol.33 pp.2-7

McLafferty, F.W., & Bryce, T.A., 1967 "Metastable-ion Characteristics: Characterization of Isomeric Molecules," Chemical Communications, Issue 23 pp,1215-1217

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

Miller, P., & Denton, M., 1986 "The Quadrupole Mass Filter: Basic Operating Concepts," Journal of Chemical Education, Vol.63 pp.617-623

Morsa, D., 2011 "Effective Temperature of Ions in Travelling Wave Ion Mobility Spectrometry," Analytical Chemistry, Vol.83 pp.5775-5782

Nier, A., 1989 "Some Reminiscences of Mass Spectrometry and the Manhattan Project," Journal of Chemical Education, VOl.66 pp.385-388

Nobel, A. B., 1895; "Full Text of Alfred Nobel's Will". Nobelprize.org. Access Date; 7th Aug 2013 http://www.nobelprize.org/alfred_nobel/will/will-full.html

Pace, C.N., *et al.* 2994 "Protein Structure, Stability and Solubility in Water and Other Solvents," Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, Vol.359 pp.1225-1235

Paul, W., Steinwedel, H., 1953 "*Ein Neues Massenspektrometer Ohne Magenfeld" Zeitschrift Fur Naturforschung, Vol.8 pp.448-450

Paul, W., & Steinwedel, H., 1960 "Apparatus for Separating Charged Particles of Different Specific Charges," US Patent 2939952, filed; 21 Dec. 1954, patented; 7 June 1960

Pittenauer, E., & Allmaier, G., 2009 "High-Energy Collision Induced Dissociation of Biomolecules: MALDI-TOF/RTOF Mass Spectrometry in Comparison to Tandem Sector Mass Spectrometry," Combinatorial Chemistry & High Throughput screening, Vol.12 pp.137-155

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, Vol.261 pp.1-12

Rayleigh, Lord, 1882 "On the Equilibrium of Liquid Conducting Masses Charged with Electricity," Philosophical Magazine Series 5, Vol.14 pp.184-186

Ruotolo, B.T., *et al.*, 2005 "Evidence for Macromolecular Protein Rings in the Absence of Bulk Water," Science, Vol.310 pp.1658-1661

Ruotolo, B.T., *et al.*, 2007 "Ion Mobility-Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes," Angewandte Chemie-International Edition, Vol.46 pp.8001-8004

Ruotolo, B., *et al.* 2008 "Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes," Nature Protocols, Vol.3 pp.1139-1152

Salbo, R., *et al.*, 2012 "Travelling-Wave Ion Mobility Mass Spectrometry of Protein Complexes: Accurate Calibrated Collision Cross-Sections of Human Insulin Oligomers," Rapid Communications in Mass Spectrometry, Vol.26 pp.1181-1193

Scarff, C., *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 Spectrometry," Rapid Communications in Mass Spectrometry, Vol.22 pp.3297-3304

Scarff, C., *et al.*, 2009 "Probing Haemoglobin Structure by Means of Travelling-Wave Ion Mobility Mass Spectrometry," Journal of the American Society for Mass Spectrometry, Vol.20 pp.625-631

Shi, S.D.-H., Hendrickson, C.L., Marshall, A.G., "Counting Individual Sulphur Atoms in a Protein by Ultrahigh-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Experimental Resolution of Isotopic Fine Structure in Proteins," Proceedings of the National Academy of Sciences, Vol.95 pp.11532-11537

Shvartsburg, A.A., & Jarrold, M.F., 1996 "An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions," Chemical Physical Letters, Vol.261 pp.86-91

Shvartsburg, A.A., & Smith, R.D., 2008 "Fundamentals of Travelling Wave Ion Mobility Spectrometry," Analytical Chemistry, Vol.24 pp.9689-9699

Smith, R., *et al.*, 1990 "Sensitivity considerations for large molecule detection by capillary electrophoresis-electrospray ionization mass spectrometry," Journal of Chromatography, Vol.516 pp.157-165

Sobott, F., & Robinson, C.V., 2004 "Characterizing Electrosprayed Biomolecules Using Tandem-MS - The Noncovalent GroEL Chaperonin Assembly," International Journal of Mass Spectrometry, Vol.236 pp.25-32

Stephens, W.E., 1946 "A Pulsed Mass Spectrometer with Time Dispersion," Physical Review, Vol.69 pp.691

Sun, J., *et al.*, 2007 "Method for identifying nonspecific protein-protein interactions in nanoelectrospray ionization mass spectrometry," Analytical Chemistry, Vol.79 pp.8301-8311

"The Nobel Prize in Chemistry 1922," Nobelprize.org. Access date; 7th August 2013 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1922/

"The Nobel Prize in Chemistry 2002," Nobelprize.org. Access Date; 7th August 2013 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/

"The Nobel Prize in Physics 1906," Nobelprize.org. Access date; 7th August 2013 http://www.nobelprize.org/nobel_prizes/physics/laureates/1906/

"The Nobel Prize in Physics 1989," Nobelprize.org. Access date; 7th August 2013 http://www.nobelprize.org/nobel_prizes/physics/laureates/1989/

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, Vol.15 pp.596-601

Tanaka, K., 1988 "Protein and Polymer Analyses up to *m/z* 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry," Rapid Communications in Mass Spectrometry, Vol.2 pp.151-153

Taylor, Sir G., 1964 "Disintegration of Water Drops in an Electric Field," Proceedings of the Royal Society of London A, Vol.280 pp.383-397

Thalassinos, K., *et al.*, 2004 "Ion Mobility Mass Spectrometry of Proteins in a Modified Commercial Mass Spectrometer," International Journal of Mass Spectrometry, Vol.236 pp.55-63

Thalassinos, K., *et al.*, 2009 "Characterization of Phosphorylated Peptides Using Travelling Wave-Based and Drift Cell Ion Mobility Mass," Peptides, Vol.81 pp.248-254

Thalassinos, K., & Scrivens, J.H., 2009 "Applications of Travelling Wave Ion Mobility-Mass Spectrometry," March, R.E., & Todd, J.F.J., (Editor), Practical Aspects of Trapped Ion Mass Spectrometry Vol. V, CRC Press, Chapter 8, pp.205-236

Thomson, J.J., 1897 "Cathode Rays," Philosophical Magazine, Series 5, Vol.44 pp.293-316

Thomson, J.J., 1907 "Rays of Positive Electricity," Philosophical Magazine, Series 6, Vol.13 pp.561-575

Thomson, J.J., 1912 "Further Experiments of Positive Rays," Philosophical Magazine, Series 6, Vol.24 pp.209-253

Thomson, J.J., 1913 "Rays of Positive Electricity and their Application to Chemical Analysis," Longmans Green & Co., London

Thomson, J.J., 1914 "Rays of Positive Electricity," Proclamations of the Royal Society London, (Series A), Vol.89 pp.1-20

Tito, M., 2000 "Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid," Journal of the American Chemical Society, Vol.122, pp.3550-3551

Uetrecht, C., *et al.* 2010 "Ion Mobility Mass Spectrometry of Proteins and Protein Assemblies," Chemical Society Reviews, Vol.39 pp.1633-1655

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

Valentine, S.J., *et al.*, 1999 "A Database of 660 Peptide Ion Cross Sections: Use of Intrinsic Size Parameters for Bona Fide Predictions of Cross Sections," Journal of the American Society for Mass Spectrometry, Vol.10 pp.1188-1211

van Berkel, G., & Kertesz, V., 2012 "Utilizing the Inherent Electrolysis in a Chip-Based Nanoelectrospray Emitter System to Facilitate Selective Ionization and Mass Spectrometric Analysis of Metallo Alkylporphyrins," Analytical & Bioanalytical Chemistry, Vol.403 pp.335-343

van den Heuvel, R. H. H., & Heck, A.J.R., 2004 "Native Protein Mass Spectrometry: From Intact Oligomers to Functional Machineries," Current Opinion in Chemical Biology, Vol.8 pp.519-526

von Helden, G., *et al.*, 1991 "Structures of Carbon Cluster Ions from 3 to 60 Atoms: Linears to Rings to Fullerenes," Journal of Chemical Physics, Vol.95 pp.3835-3837

von Helden, G., *et al.*, 1993 "Carbon Cluster Cations with up to 84 Atoms: Structures, Formation Mechanism, and Reactivity," Journal of Physical Chemistry, Vol.97 pp.8182-8192

Wang, W., Kitova, E., & Klassen, J., 2003 "Influence of Solution and Gas Phase Processes on Protein-Carbohydrate Binding Affinities Determined by Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry," Analytical Chemistry, Vol.75 pp.4945-4955

Wangler, T., 1998 "RF Linear Accelerators," Wiley Series in Beam Physics & Accelerator Technology. John Wiley, New York, USA.

Waters Corporation, 2006, "Waters Unveils New Mass Spectrometry Systems," Waters.com. Access Date; 7th August 2013. http://www.waters.com/waters/newsDetail.htm?locale=en_US&id=10001372

Williams, J.P., & Scrivens, J.H., 2008 "Coupling Desorption Electrospray Ionisation and Neutral Desorption/Extractive Electrospray Ionization with a Travelling-Wave Based Ion Mobility Mass Spectrometer for the Analysis of Drugs," Rapid Communications in Mass Spectrometry, Vol.22 pp.187-196

Wilm, M.S., & Mann, M., 1994 "Electrospray and Taylor-Cone Theory, Dole's Beam of Macromolecules at Last?" International Journal of Mass Spectrometry and Ion Processes, Vol.136 pp.167-180

Wilm, M., & Mann, M., 1996 "Analytical Properties of the Nanoelectrospray Ion Source," Analytical Chemistry, Vol.68 pp.1-8

Whitehouse, C.M., *et al.*, 1985 "Electrospray Interface for Liquid Chromatographs and Mass Spectrometers," Analytical Chemistry, Vol.57 pp.675-679

Winston, R.L., and Fitzgerald, M.C., 1997 "Mass Spectrometry as a Readout of Protein Structure and Function," Mass Spectrometry Reviews, Vol.16 pp.165-179

Wiza, J.L., 1979 "Microchannel Plate Detectors," Nuclear Instruments and Methods, Vol.162 pp.587-601

Wolynes, P. G., 1995 "Biomolecular folding in vacuo!!!(?)" Proclamations of the National Academy of Sciences, Vol.92 pp.2426-2427

Wyttenbach, T., *et al.* 1997 "Effect of the Long-Range Potential on Ion Mobility Measurements," Journal of the American Society for Mass Spectrometry, Vol.8 pp.272-282

Wyttenbach, T., Witt, M., and Bowers, M., 2000 "Effect of the Long-Range Potential on Ion Mobility Measurements," Journal of the American Chemical Society, Vol.122 pp.3458-3464

Wyttenbach, T., Bleiholder, C., & Bowers, M.T., 2013 "Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements," Analytical Chemistry, Vol.85 pp.2191-2199

Yamashita, M., & Fenn, J.B., 1984 "Electrospray Ion Source. Another Variation on the Free-Jet Theme," Journal of Physical Chemistry, Vol.88 pp.4451-4459

Zenobi, R., & Knochenmuss, R., 1998 "Ion Formation in MALDI Mass Spectrometry," Mass Spectrometry Reviews, Vol.17 pp.337-366

2. Instrumentation and Analysis

Overarching instrumentation and analysis methods are outlined here. For investigation relevant materials and methods please see the appropriate chapter.

2.1. Mass Spectrometry: Instrumentation

All data sets were acquired using a first generation Synapt HDMS (Waters, Manchester, UK), hybrid quadrupole orthogonal accelerator mass spectrometer with TriWave IM capabilities (Schematic previously depicted in Fig. 1.13). This instrument has been modified for high mass by installing a reduced RF frequency quadrupole generator enabling the study of ions up to *m/z* 32 000. For all studies presented here, instrumentation was combined with a nESI source using *in-house* prepared, gold coated 'bee-stinger' type nanospray capillaries. Instrumental condition control and data acquisition were performed using MassLynx v4.1 (SCN728) (Waters, Manchester, UK).

ToF calibration was performed using an aqueous solution of 24 mg/ml caesium iodide diluted with 30 % isopropanol, producing a suitable calibration range exceeding *m/z* 16 000 (Sterling *et al.* 2010). $N_{2(g)}$ was used for IM separation, with $Ar_{(g)}$ used within the Trap and Transfer TriWave components (see schematic in Fig. 1.13) for collision studies. When required, sample cone cleaning was performed using a solution of 45 % methanol, 45 % ddH_2O and 10 % formic acid, and sonicated at 37 kHz for no more than 15 minutes, prior to ToF calibration and analysis.

'Bee-stinger' type nanospray capillaries were produced *in-house* using, 1 mm outside / 0.5 mm inside diameter, borosilicate glass capillaries (Sutter Instruments, Novato, CA, USA) and calibrated using a P-97 flaming/brown type micropipette puller (Sutter Instruments, CA, USA). Calibrated capillaries were coated with a fine layer of gold using a SC7620 'Mini' sputter coater (Quorum Technologies, West Sussex, UK). Nanospray capillaries were loaded with 2.5 μL of solvent-diluted analyte combined with an estimated sample flow rate less than 10 nL/min enabling nESI durations exceeding 30 minutes, suitable for native mass spectrometric analysis.

2.2. Mass Spectrometry: Data Analysis

Mass spectra and t_D analysis were performed using MassLynx v4.1 (SCN712) (Waters, Manchester, UK), combined with Driftscope v2.1 (Waters, Manchester, UK) for extended IM spectra analysis. All spectra herein are shown with minimal smoothing and no baseline subtraction.

IM t_D values were converted to CCS using power fit curves in accordance with the methods outlined in Hilton *et al.* (2010), and Thalassinos *et al.* (2009). Calibration was performed using experimental $He_{(g)}$ drift cell CCS values presented by Bush *et al.* (2010), and corrected for the $N_{2(g)}$ mobility gas used in the IM separator. Calibration curves were fitted against experimental t_D values, obtained from the FWHM of the ion peak, to calculate CCS using Microsoft Office Excel 2003 (Microsoft, WA, USA). Calibrants exhibiting estimated cross sections deviating more than 5 % from the published cross sections were removed from calibration curves to optimise fitting, whilst retaining a suitable t_D spread.

Conversion of PDB data to CCS values for comparison against IM experimental values was performed using MobCal to apply the PA and EHSS models, averaged over 30 returned simulations (Mesleh *et al.* 1996, and Shvartsburg & Jarrold 1996). Due to its computationally demanding nature, the TM model was not used. This is supported by observations by Scarff and co-workers (2008) reporting comparable accuracy between the TM and EHSS methods for proteins similar in size to those studied here.

2.3. Sample Analysis: Buffer Preparation and Exchange Methods

Native-mass spectrometry analysis buffers were produced using crystalline analytical reagent grade AmAc (> 99.9 % PURITY, Fisher Scientific Ltd, Loughborough, UK), diluted to concentration using ddH_2O. Buffer pH was measured using a SevenMulti pH probe (Mettler-Toledo Ltd, Leicester, UK), and modified to the required value using glacial acetic acid (> 99.9 % purity, Fisher Scientific Ltd, Loughborough, UK) and 33 % ammonia solution (Sigma-Aldrich Ltd, Dorset, UK).

Two interchangeable methods for buffer exchanging samples were applied either individually or, as sequential complementary methods. Initial sample buffer exchange applied Micro Bio-Spin 6 columns (BioRad, Hertfordshire, UK). Columns were typically prepared using four 1000 mM AmAc (pH dependent on the protein sample) washes, followed by three at the experimental concentration and pH. Two or more Bio-Spin column washes were not used, as observations indicated a sample concentration loss of ~20 % for each Bio-Spin 6 column used (data not shown).

Amicon 1.5 ml 10 kDa MWCO centrifugal concentrator columns (Millipore, Dundee, UK) were typically used as a final buffer exchange, and sample concentration step. Amicon columns were prepared using two washes at 1000 mM AmAc (pH dependent on the protein sample), with the excess solvent removed via centrifugation. This step was followed by a single wash at the required experimental AmAc buffer concentration and pH. Sample buffer exchange was performed using a single column with two to three wash cycles. Minimum sample recovery volume was 50 μL, with no notable loss in concentration (< 10 % compared to theoretical).

Samples of native Bovine serum albumin (BSA) and *S. cerevisiae* alcohol dehydrogenase (ADH, Sigma-Aldrich Ltd, Dorset, UK) IM-MS calibrants were produced by resuspending the lyophilised protein to 20 μM in 200 mM AmAc (pH 7). These samples were then buffer exchanged into the required experimental AmAc buffer concentration and pH using a single Bio-Spin 6 column prior to analysis.

Samples of non-native Equine myoglobin (Sigma-Aldrich Ltd, Dorset, UK) IM-MS calibrant were produced by resuspending the lyophilised protein to 20 μM in denaturing buffer: 49.5 % ddH_2O, 49.5 % acetonitrile, 1 % formic acid. For denatured protein mass analysis a single wash using the above Amicon MWCO centrifugal concentrators were applied to buffer exchange samples into denaturing buffer. Denatured control analysis, of all proteins studied here in, has been used to accurately calculate experimental masses and applied to improve the accuracy of molarity calculations.

2.4. Additional Software

Extended t_D and mass analysis, including graph and table production, was performed using Microsoft Office Excel 2003, and 2010 (Microsoft, WA, USA). Images and other figures, not referenced to outside sources, were produced using Illustrator CS3 (Adobe, CA, USA).

2.5 References

Bush, M.F., *et al.* 2010 "Collision Cross Sections of Proteins and their complexes: calibration framework and Database for Gas-Phase Structural Biology," Analytical Chemistry, Vol.82, pp.9557-9565

Hilton, G.R., *et al.*, 2010 "Structural Analysis of Prion Proteins by Means of Drift Cell and Travelling Wave Ion Mobility Mass Spectrometry," Journal of the American Society for Mass Spectrometry, Vol.21 pp.845-854

Mesleh, M.F., *et al.*, 1996 "Structural Information From Ion Mobility Measurements: Effects of the Long-Range Potential," Journal of Physical Chemistry, Vol.100 pp.16082-16086

Scarff, C., *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 Spectrometry," Rapid Communications in Mass Spectrometry, Vol.22 pp.3297-3304

Shvartsburg, A.A., & Jarrold, M.F., 1996 "An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions," Chemical Physical Letters, Vol.261 pp.86-91

Sterling, H.J., *et al.*, 2010 "Effects of Buffer Loading for Electrospray Ionization Mass Spectrometry of a Noncovalent Protein Complex that Requires High Concentrations of Essential Salts," Journal of the American Society for Mass Spectrometry, Vol.21 pp.1045-1049

Thalassinos, K., *et al.*, 2009 "Characterization of Phosphorylated Peptides Using Travelling Wave-Based and Drift Cell Ion Mobility Mass," Peptides, Vol.81 pp.248-254

Chapter 3: Probing Conformational Intermediates of an α_1-Antitrypsin Variant

---

Exhibiting a native metastable fold, mutations of α_1-Antitrypsin can promote misfolding events increasing the susceptibility to aggregation events at sites of synthesis through toxic gain of function. Formation of these α_1-Antitrypsin misfolding intermediates are known to propagate under biologically relevant conditions. Using IM-MS we have characterised the structure and stability of the slow polymerisation mutant K154N. The results obtained have shown that this mutant populates an intermediate of increased structural stability upon incubation at biologically relevant temperatures.

3.1.1. The Serine Protease Inhibitor Super-Family

The serine proteinase inhibitor (serpin) superfamily, are a family of proteins originally termed for their identification in mammalian plasma (Carrell & Travis 1985). To date however, over 1500 serpin-like proteins have been characterised in all taxonomic kingdoms (Huntington 2011, Janciauskiene *et al.* 2011, and Silverman *et al.* 2001).

Ranging between 350 and 500 amino acids in length, serpins inhibit the chymotrypsin serine family of proteases. Inhibition of proteinases by serpins progresses by a suicide substrate mechanism, promoting a rearrangement of the typical serpin metastable fold that binds the serpin and proteinase irreversibly. These irreversibly bound serpin-proteinase complexes are then targeted for destruction (Silverman *et al.* 2001).

Of the 1500 identified species, the 36 of human origin are required for tightly regulated functions that include blood coagulation, fibrinolysis and inflammation (Huntington 2011). Additionally these human serpins are recognised to be required in cross-class inhibition, hormone transport, and blood pressure regulation (Huntington 2011, and Silverman *et al.* 2001).

3.1.1.1. α_1-Antitrypsin: Structure and Function

The human glycoprotein α_1-Antitrypsin (AAT) (UniProt ID; P01009, 46 kDa, 418 amino acids) is the most abundant human serpin and is predominantly secreted into the blood serum by the liver, which accounts for 70 to 80 % of all expressions (Janciauskiene *et al.* 2011, and Massi & Chiarelli 1994). Additional, low level sites of expression and secretion include the alveolar macrophages and intestinal wall epithelial cells (Massi & Chiarelli 1994).

Like all serpins, AAT can be structurally characterised by a metastable fold of nine α-helices (A-I), and three β-pleated sheets (A-C) (Fig. 3.1) (Patschull *et al.* 2011). β-sheet A is the primary scaffold that the overall metastable tertiary structure forms around, and is heavily associated with overall protein function (Dafforn *et al.* 1999).

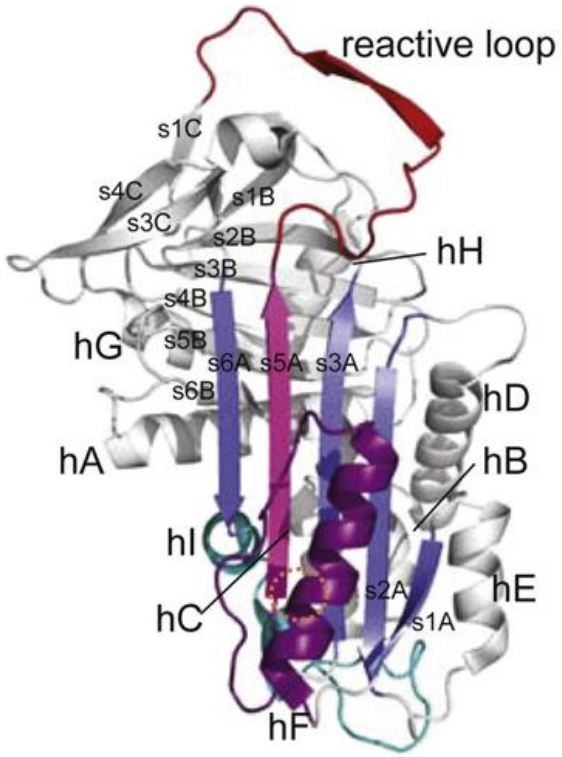

Figure 3.1. Crystal Structure of α_1-Antitrypsin

α-helices labelled 'h*x*' and β-pleated sheets labelled 's*yx*' ('*x*' represents the helix or sheet, and '*y*' the strand number). Relevant to work presented here is the K154N mutation (orange circle) which is located within α-helix F (purple). PDB ID: 3NE4

Figure reproduced and adapted from Nyon *et al.* 2012

AAT primarily functions to inhibit neutrophil elastase (section 3.1.1.2), binding with a high affinity (K_a 10^7 $M^{-1}s^{-1}$, Taggart *et al.* 2000) in a 1:1 ratio (Fig. 3.2), to prevent elastolytic destruction of tissues (Gooptu & Lomas 2009). Neutrophil elastase binding to

AAT occurs at the protruding reactive centre loop, mediated by the amino acids Met358 and Ser359 (Massi & Chiarelli 1994, and Taggart *et al.* 2000). Upon formation of the native 1:1 oligomer, the complex is quickly targeted for phagocytosis by macrophages (Massi & Chiarelli 1994).

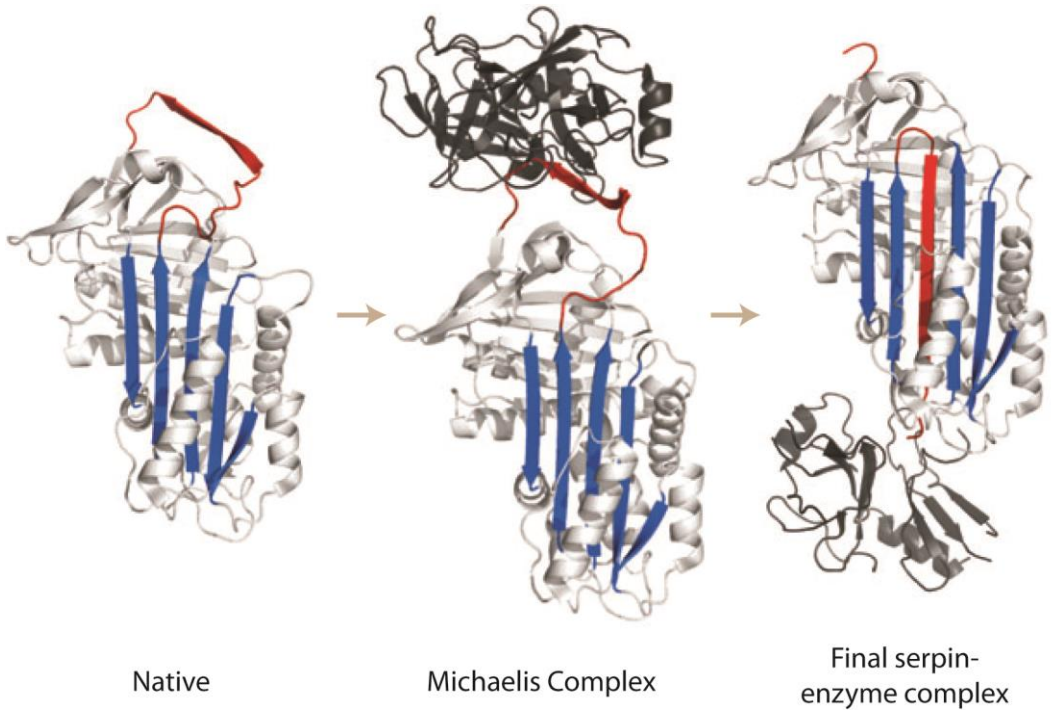

Figure 3.2. Neutrophil Elastase Inhibition by α_1-Antitrypsin

Native AAT adopts a metastable fold with a protruding reactive centre loop (red, shown here as a β-sheet) that acts as a pseudo substrate for the target neutrophil elastase (black). This forms an initial Michaelis complex that promotes the protease function of neutrophil elastase, cleaving the AAT reactive centre loop. Upon cleavage of the reactive centre loop, AAT undergoes structural rearrangement, inserting the reactive centre loop within the β-sheet A (blue), to form the fourth (of six) strand. This structural rearrangement promotes the translocation of the neutrophil elastase from pole to pole, in an irreversibly bound state that causes inactivation.

Figure reproduced and adapted from Gooptu & Lomas 2009

Additionally AAT is required in both immunological resistance and immunoregulation. As an immunoregulator, investigations into fibroblast proliferation have shown that AAT is able to act as a mitogen, promoting tissue repair in a range of tissues (Dabbagh *et al.* 2001). Studies of immunological resistance to pathogens presented by Knappstein and co-workers (2004) showed that AAT could bind, and inhibit *E. coli*

EspB. EspB induces pore formation in host cell membranes, allowing effector protein injection and correlates positively with blood cell lysis. Therefore, AAT inhibition of EspB is able to prevent haemolysis of blood cells (Knappstein *et al.* 2004).

Work performed by Griese and co-workers (2007) explored the pharmaceutical applications of ATT. Their work showed that inhalation of an aerosolised form of AAT, by cystic fibrosis patients suffering from pathogenic bronchial infection, reduced lung inflammation. In addition to these expected reduced levels of inflammation, typical of *in situ* AAT activity, inhalation reduced the observed levels of the pathogen *P. aeruginosa* (Griese *et al.* 2007) in line with its previously identified immunological function.

3.1.1.2. Neutrophil Elastase

Neutrophil elastase (UniProt ID; P08246, 28.5 kDa, 267 amino acids) is a serine protease that is recruited to the lungs at sites of inflammation and infection, acting as a front line defence against gram-negative bacteria, spirochaetes, and fungal attack (Chua & Laurent 2006). Recruitment of neutrophil elastase occurs through the innate immune systems neutrophil cells that secrete azurophilic granules containing an estimated 67 000 activated elastase proteins (Kelly, Greene & McElvaney 2008).

Upon secretion, neutrophil elastase degrades the majority of the pulmonary extracellular matrix components elastin, laminin and proteoglycan fibronectin. These degradation events are promoted through nucleophilic attack connected with the neutrophil elastases catalytic region (His^{41}-Asp^{99}-Ser^{173}) serine γ-oxygen. Additionally neutrophil elastase is able to induce epithelial cell apoptosis through surface receptor activation (Kelly, Greene & McElvaney 2008, and Lungarella *et al.* 2008). Therefore, unregulated neutrophil elastase activity, through the loss of AAT activity, can be highly detrimental to an individual's health.

3.1.2. Pathology and Mechanisms of α_1-Antitrypsin Genetic Disorders

AAT was originally identified by Jacobson (1955) as the primary inhibitor of blood serum, exhibiting the highest serum circulation level of any protease (Heimburg & Haupt 1965). Subsequent work by Laurell & Eriksson (1963) connected AAT deficiency with emphysema. This deficiency additionally promotes the development of serpinopathies, such as liver cirrhosis, which occur due to the toxic gain of function. These toxic gain of function disorders occur as a feature of the ordered polymerisation of serpins at the site of synthesis (Ekeowa *et al.* 2010).

Deficiency and aggregation disorders of AAT affect 1 in 2000 Caucasian individuals of north European ancestry (Blanco *et al.* 2001, Knaupp & Bottomley 2009). Therefore, AAT has become subject to intense study within the scientific community. These studies have subsequently identified in excess of 100 allelic variants of AAT (Belorgey *et al.* 2007). These allelic variants are classified according to an isoelectric point and electrophoretic pattern based phenotypic identification (Pi) system, measuring the phenotypic expression level, and mobility pattern on a polyacrylamide gel (McElvaney *et al.* 1997, and Massi & Chiarelli 1994).

The Pi system classifies alleles in two ways. The first of these is polyacrylamide gel mobility pattern separation, which groups alleles into one of four alphabetic classifications dependent on their conformationally extensive nature: F = *fast*, M = *medium*, S = *slow*, and Z = *very slow*. F alleles represent compact structural conformations, in contrast to the more extended, slow moving, S and Z classifications. Therefore, these classifications indicate the propensity of an allelic variant to aggregate in contrast to the wild type (wt) protein. Individuals are thereby classified by their autosomal-codominant status e.g. ZZ, SZ, etc, with heterozygous phenotypes exhibiting a combination of both allelic variants using the polyacrylamide gel analysis methods (Blanco *et al.* 2001, and Hutchinson *et al.* 1983).

The Pi system additionally classifies alleles by their expression levels, again into one of four categories, dependent on blood serum levels: *normal* (1.5-3.5 mg/ml), *deficient* (0.15-2 mg/ml), *null* (no expression due to introduction of premature stop codon),

and *dysfunctional* (*normal* serum levels, functionally ineffective) (Gooptu & Lomas 2009, Massi & Chiarelli 1994, and McElvaney *et al.* 1997).

Clinically the general population exhibits a mixture of the predominant MM Pi phenotype (86 % population), and MS Pi phenotypes (9 % population). Both the MM and MS exhibit *normal* blood serum levels (Kelly *et al.* 2010, and Sandford *et al.* 1999). Comparatively the third most common phenotype is the MZ Pi phenotype (3 % population) which exhibits a *deficient* level of AAT (57 % of normal) (Sandford *et al.* 1999).

Common allele mutations causing AAT *deficiency* levels are the S-type E264V, and Z-type E342K (Fra *et al.* 2012, and Kelly *et al.* 2010). Phenotypes with commonly associated clinical effects are the SS homozygous, and SZ and ZZ heterozygous, phenotypes. The SS phenotype results in mild *deficient AAT* levels, with an increased risk to developing emphysema (Fra *et al.* 2012, and Sandford *et al.* 1999). In contrast, the SZ and ZZ phenotypes exhibit more severe serum level reductions, 25 % and 15 % of *normal* respectively, and account for the majority of clinically diagnosed deficiency disorders (Fra *et al.* 2012).

The common Z mutation, E342K, causes a conformational change in β-sheet A that distorts its relationship with the reactive centre loop, allowing polymerisation (Ekeowa *et al.* 2010, and Gooptu & Lomas 2009). Consequently, Z phenotypes have been shown to promote the development of a range of liver diseases (neonatal hepatitis, juvenile cirrhosis, and hepatocellular carcinoma), glomerular renal damage, chronic obstructive pulmonary disease. This is in addition to a predisposition to early onset emphysema resulting from the reduced serum neutrophil elastase inhibitory activity (Dafforn *et al.* 1999, Gooptu & Lomas 2008, and Larsson 1978).

As well as a predisposition to emphysema, AAT deficiency has been connected to the appearance of other disorders such as panniculitis, which affects an estimated 1 in 1000 *deficient* patients (Rajpara, Erickson & Driscoll 2010). The clinical manifestation of AAT associated panniculitis is linked to a range of deficiency alleles and typically characterised by painful, weepy cutaneous nodules at sites of trauma, with areas of fat necrosis (diagnosed by deep skin biopsies) (Stoller & Aboussouan 2005). Pathogenesis of panniculitis has been demonstrated, in the skin of one individual, as a co-localization of neutrophils with Z allele AAT polymers. This suggests that these polymers aid the process

of panniculitis development (Gross *et al.* 2009). Due to the rarity of this disorder the current hypothesis is that panniculitis is a most likely a multi-factorial based disease, combining AAT disorders with other factors and genetic predispositions, with work still ongoing (Rajpara, Erickson & Driscoll 2010, and Stoller & Aboussouan 2005).

3.1.3. α_1-Antitrypsin Polymerisation Pathways

The native folding pathway of AAT requires that it must bypass a global free energy minimum to favour the active metastable state required to bind neutrophil elastase. Typically, folding intermediates in this pathway tend to be short-lived preventing polymerisation. AAT point mutations however, can modify folding intermediate activity, formation, and lifetime. These modifications are subsequently able to promote AAT misfolding, enabling polymerisation (Yamasaki *et al.* 2011).

Formation of AAT mutant intermediates, within the first step of the polymerisation pathway, has been shown *in vitro* to populate under several conditions. These conditions include: chaotropic agents (Kim & Yu 1996, and Tew & Bottomley 2001), low pH (Devlin *et al.* 2002), and high temperature (Egelund *et al.* 2001, and Ekeowa *et al.* 2010). Ekeowa and co-workers (2010) concluded that these conditions induce different solution based structural intermediates that support a number of polymerisation models (outlined below). Polymerisation sequentially progresses in a concentration, and temperature dependent manner (Lomas *et al.* 1992, and Lomas *et al.* 1993).

To date three models of AAT polymerisation have been proposed: the intermolecular reactive centre loop insertion (section 3.1.3.1.), β-hairpin insertion (section 3.1.3.2.), and C-terminal domain swapping models (section 3.1.3.3.). All three of these polymerisation models progress via a two-step process, which requires AAT to form a polymerogenic intermediate that promotes the formation of polymers (Nyon *et al.* 2012, and Yamasaki *et al.* 2008). These polymers exhibit a 'beads on a string' morphology that is able to permit self-termination by circularising (Huntington *et al.* 1999, Lomas *et al.* 1992, and Lomas *et al.* 1993).

3.1.3.1. Intermolecular Reactive centre loop Insertion Model

The earliest, or 'classical,' model of AAT polymerisation is the intermolecular reactive centre loop insertion model (Fig. 3.3). Destabilisation of β-sheet A, promoted by the Z-mutant E342K, enables the reactive centre loop to exhibit a more open intermediate conformation (Lomas *et al.* 1992). This destabilised open reactive centre loop extends and inserts within the β-pleated sheet A of a second monomer. Repetition of these unfolding and binding events promotes AAT polymerisation (Dafforn *et al.* 1999, Gooptu *et al.* 2000, and James & Bottomley 1998).

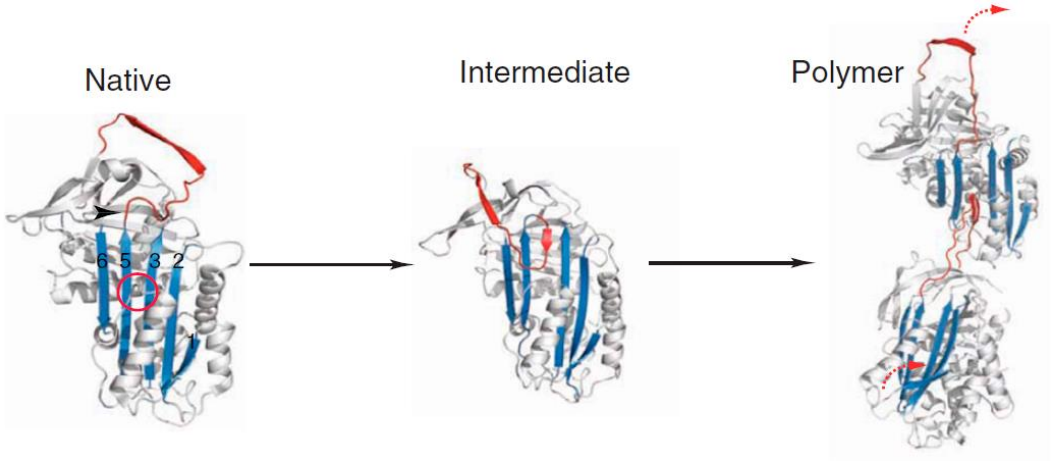

Figure 3.3. Intermolecular Reactive centre loop Insertion Model of α_1-Antitrypsin Polymerisation

The Z mutation (E342K, red circle) destabilises β-sheet A (blue) forming an activated intermediate species. Intermolecular linkage then occurs through donation of the reactive centre loop (red) from one molecule, to the open lower portion of the central β-sheet A channel of a second molecule.

Figure reproduced and adapted from Ekeowa *et al.* 2010

3.1.3.2. β-Hairpin Insertion Model

The β-hairpin insertion model (Fig. 3.4) supports a relatively extensive level of domain swapping in comparison to the other models presented. This polymerisation pathway model is strongly dependent on the classical Z mutation, E342K, located within the head of β-sheet A, strand 5 (red circle, Fig. 3.3) (Yamasaki *et al.* 2008). Formation of the intermediate progresses through an unfolding of α-helix I, which promotes the formation of a β-hairpin intermediate (consisting β-sheet A5, and the reactive centre loop). This

promotes the formation of an open region within β-sheet A that enables β-hairpin intermediate binding by a second monomer (Krishnan & Gierasch 2011, and Yamasaki *et al.* 2008).

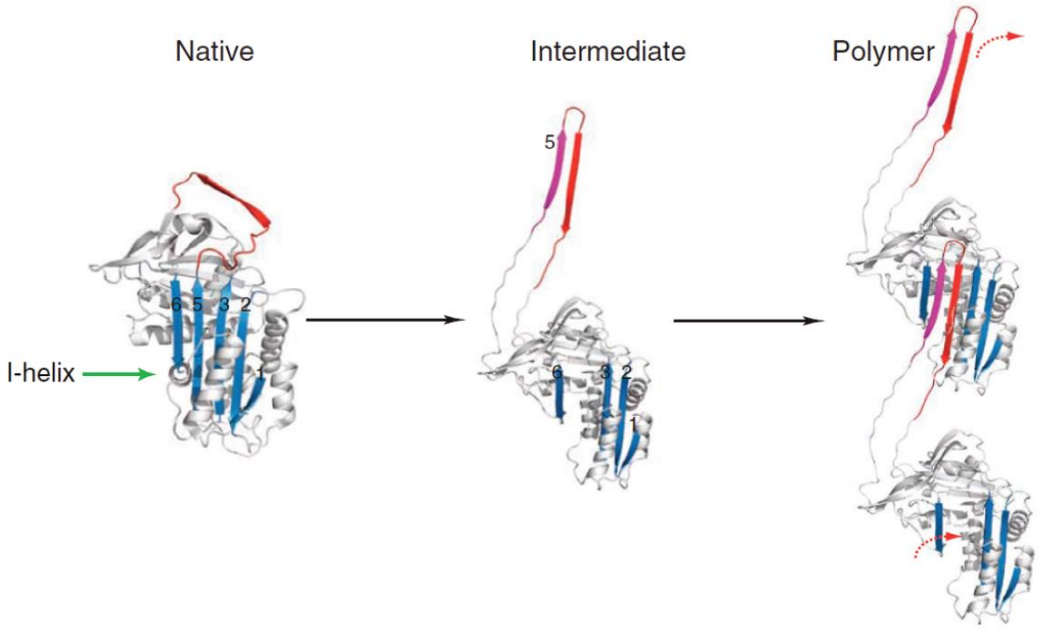

Figure 3.4. β-Hairpin Insertion Model of α_1-Antitrypsin Polymerisation

β-hairpin insertion model first requires that α-helix I is unfolded (green arrow), allowing the extension of β-sheet A5 (purple). This extension provides an intermediate state with a generated binding interface comprising the reactive centre loop (red) and β-sheet A5 (purple). This enables insertion of the hairpin into the open β-sheet A region of a second monomer.

Figure reproduced and adapted from Ekeowa *et al.* 2010

3.1.3.3. C-Terminal Domain Swapping Model

The C-terminal domain swapping model,[16] recently reported by Yamasaki and co-workers (2011), has shown that AAT polymerisation can be promoted by the formation of a β-hairpin comprising β-sheet 4B and 5B. Formation of this β-hairpin additionally promotes dissociation of β-sheet 1C (Fig. 3.5). This model has shown that the unassimilated β-sheet 4B and 5B are able to assimilate into the incomplete β-sheet of a second monomer, completing the native like fold. Data additionally supports that this unassimilated

[16] Alternative known as the 'triple strand model'

β-sheet B loop is able to re-assimilate into the native fold, of the same protomer, promoting self-termination (Yamasaki *et al.* 2011).

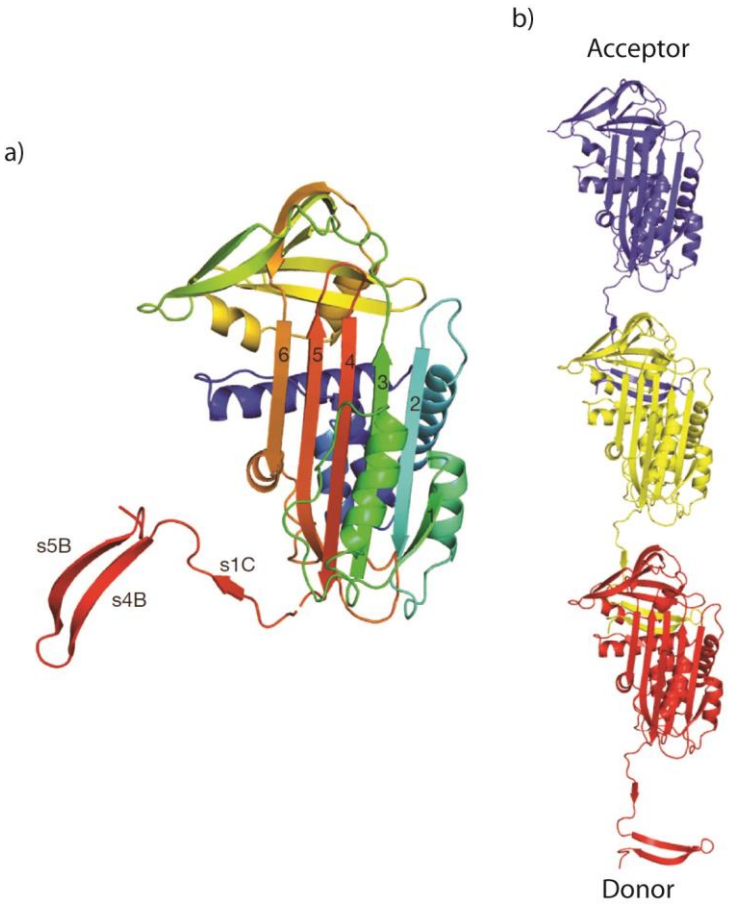

Figure 3.5. C-terminal Domain Swapping Model of α_1-Antitrypsin Polymerisation

a) Crystal structure of the polymerogenic intermediate required for polymerisation by the C-terminal domain swapping model. The fold exhibits a dissociated β-sheet 4B, 5B and 1C. The unassimilated β-sheet 4B and 5B form a β-hairpin that is able to embed within the β-sheet B of the second monomer completing the native metastable fold. b) This assimilation promotes the head to tail morphology of the polymer. The open C-terminal domain of the red monomer subunit can be subsequently self donate enabling self-termination of polymerisation.

Figure reproduced and adapted from Yamasaki *et al.* 2011

3.2. Experimental Aims and Objectives

All three models of AAT polymerisation exhibit common conclusions: dissociated β-sheets, and 'beads on a string' polymer morphology (Nyon *et al.* 2012, and Yamasaki *et al.* 2008). Despite these common conclusions, research has yet to provide a conclusive answer regarding the mode of AAT polymerisation. Consequently, recent investigations have progressed to study the polymerisation pathway using rarer mutations that stabilise the intermediate step, slowing the polymerisation pathway. Stabilisation of this intermediate

step should help to confer support for either, or all three, of the above models dependent on the conditions and methods used.

The work presented here has been published, in part, in Nyon *et al.* (2012) (Appendix I), targeting the study of the recently reported AAT mutation, K154N (orange circle, Fig. 3.1). This mutation is located within α-helix F (purple, Fig. 3.1), promoting *deficient* serum levels (0.6 mg/ml) and a phenotype between the ZZ homozygote, and the MZ heterozygote. The K154N substitution causes a 36 % loss in functional activity (apparent association rate studies), resulting from perturbed electrostatic interactions caused by the loss of H-bonding between K154 and the positively charged K174 (hF to s3A linker). Additionally this substitution reduces salt bridging interactions with E151 (hF) that are required to negatively regulate AAT polymerogenic intermediate formation (Nyon *et al.* 2012).

Using IM-MS we have studied the conformational stability of the AAT mutant K154N, upon incubation at physiological temperatures. Any observed changes in conformation, upon incubation, should support previous observations that AAT mutations can promote polymerisation under biologically relevant temperatures (Ekeowa *et al.* 2010). In addition to these incubation studies, CIU has been applied to study the stability of K154N unfolding intermediates. Increases in stability would support that the K154N substitution is able to promote stable, long lived, conformational intermediates that can promote AAT polymerisation.

This work has been used to expand upon observations made using NMR spectroscopy techniques to report the structural differences between wt AAT and K154N. Compared to all other instrumental methods applied in Nyon *et al.* (2012) for studies of the AAT K154N mutation, the major advantage of IM-MS is the ability to derive accurate changes in CCS, and separation of co-existing structural conformations.

3.3. Materials and Methods

3.3.1. Molecular Biology: Preparation and Purification of α_1-Antitrypsin

All AAT samples were purified by the Gooptu research group, prior to buffer exchange and mass spectrometric analysis protocols, using the methods outlined below.

Mutations were introduced into AAT complementary DNA (cDNA), and inserted into pQE31, and the pcDNA3.1 plasmids by polymerase chain reaction mutagenesis, using the methods outlined in Zhou, Carrell & Huntington (2001). Plasmids containing cDNA encoding His-tagged (purification tag comprising 6 sequential C-terminal histidine amino acids required for nickel column separation) recombinant AAT were then transfected into XL1 Blue *E. coli* (Stratagene, CA, USA).

Wt and K154N proteins were expressed and purified from the soluble fraction of the *E. coli* lysate using HiTrap nickel chelating column and glutathione-sepharose (GE Healthcare Life Sciences, Bucks, UK) column chromatography. Samples were dialysed into 50 mM Tris, 50 mM KCl at pH 7.5, with purity confirmed using 12 % weight to volume SDS-PAGE (Parfrey *et al.* 2003).

3.3.2. Mass Spectrometric Analysis of α_1-Antitrypsin

Prior to mass spectrometric analysis, samples were buffer exchanged using a single Micro Bio-Spin 6 column (BioRad, Hertfordshire, UK) equilibrated to 100 mM AmAc pH 7. Sample concentrations were measured using serial dilution Nanodrop 1000 (Thermo Scientific, DE, USA) analysis. Extinction coefficients were applied to correct Nanodrop 1000 mg/ml values (table 3.1). Using the corrected mg/ml values, sample molarity was calculated using the experimentally calculated mass (from denatured protein mass analysis) (table 3.1). Upon measurement of concentration, samples were diluted to 20 μM for analysis, using 100 mM AmAc pH 7.

Calibrations of experimental AAT t_D values, to calculate CCS, were performed using the method outlined in section 2.2. All data were acquired at a constant wave velocity of 300 m/s, with a wave amplitude of 10 V. Denatured Equine myoglobin (Sigma-Aldrich Ltd, Dorset, UK) was used as the CCS calibrant. CCS data are the mean of three repeats, plus/minus ($^{+}/_{-}$) mean standard deviation (MSD).

Incubations of AAT, for structural stability studies at increasing temperatures, were performed using a BTD dry block heating system (Grant, Cambridge, UK) set at 20, 34 or, 39 ˚C for 120 minutes. Control spectra were recorded at time zero. Mobility data were extracted from the FWHM of the wt and K154N lowest observed charge

+12 ion (*m/z* 3798.5 and 3797.4 respectively). These studies were performed using static Trap and Transfer (see schematic in Fig. 1.13) voltages of 6 and 4 V, respectively, to minimise unfolding prior to mobility analysis.

Quadrupole mass isolation was used to select the *m/z* corresponding to the +12 ion of both the wt and K154N (*m/z* above), prior to collisional unfolding and mobility studies. The TriWave Trap voltage was ramped from 6 to 72 V, in increments of 6 V, refined to 3 V between 36 and 54 V. The TriWave Transfer voltage was kept constant at 4 V.

Additional instrumental values not previously stated: Capillary voltage; 0.9 to 1.2 kV, Nanoflow pressure; 0.08 to 0.15 Bar, Sampling cone; 20 V, Extraction cone; 1.0 V, Source temperature; 40 ˚C, Trap DC bias; 22 V, Trap Gas; 5 ml/min, IM separator gas flow; 20 ml/min (0.5 mBar). All other Trap values: release time; 100 μs, amplitude; 10 V, and extract amplitude; 5 V. Instrumental pressures: source backing; 5.75 mBar, Trap; 0.04 mBar, mobility separator; 0.55 mBar, ToF; $2.5x10^{-6}$ mBar. Mass spectra were acquired over an *m/z* range of 500 to 8 000, for two minutes using two second scans.

3.4. Results

3.4.1. MobCal Analysis of the α_1-Antitrypsin Native Structural Conformation

To confirm the biological relevance of our IM-MS studies, MobCal analysis was used to compare CCS data between the native AAT X-ray crystallography structure 3NE4[17] (Patschull *et al.* 2011). Triplicate analysis of the quadrupole isolated lowest observed AAT charge state, +12 (*m/z* 3798.5), indicates a CCS equal to 3310.21 ($^{+}/_{-}$ 38.35) $Å^2$. MobCal calculated CCS from the PDB structure over thirty simulations provided an average of 2661.22 ($^{+}/_{-}$ 27.68) $Å^2$ for the PA, and 3441.596 ($^{+}/_{-}$ 42.98) $Å^2$ for the EHSS. PA and EHSS calculated values exhibit an excellent agreement with the lowest observed charge wt CCS conformation. The closest agreement between the gas phase and MobCal calculated CCS data, is the EHSS. Experimentally this model produces the most accurate results, when compared to the PA (Scarff *et al.* 2008). These data confirm that the conditions used here

[17] Amino acids 1-418 (full length wt), with a resolution of 2.0 $Å^2$

are able to maintain the native-like structure of AAT, and are therefore capable of producing biologically valid results.

3.4.2. Thermal Stability Studies of α_1-Antitrypsin Wild Type and K154N

Work by Ekeowa and co-workers (2010) showed that incubation of the common AAT Z-mutant, E342K, at a temperature of 37 ˚C for four hours promoted AAT polymerisation (Ekeowa *et al.* 2010). Compared to common AAT mutants such as the above, K154N represents a slow polymerisation mutant with a hypothesised stable polymerogenic structural intermediate. Therefore, K154N represent a unique opportunity to study the conformational state of an AAT structural intermediate required for polymerisation. Both wt AAT and K154N were incubated at increasing temperatures (20, 34 and 39 ˚C) to compare the stability of the protein metastable state at physiological and near-physiological temperatures using IM-MS.

Table 3.1. Theoretical and Experimentally Calculated Masses of α_1-Antitrypsin Proteins

Protein	Purification Tag	Mol. E, Coeff. (M^{-1} cm^{-1})	Theoretical Mass (Da)	Experimental Mass (Da)	Mass Error ($^{+}/_{-}$ Da)	Mass Difference (Expt. - Theo.)
wt	His-tag	25565	45550.89	45570.26	2	+19.37
K154N	His-tag	25565	45536.82	45557.26	1.1	+20.44

Prior to initiation of all incubations control mass analysis was performed on both the wt AAT and K154N variant (Fig. 3.6). These data indicate no marked differences in the oligomeric state or CSD between either of the proteins. These spectra are consistent with the oligomer and CSDs of AAT after incubation at all experimental temperatures used below. This work supports observations made by other individuals using native PAGE (polyacrylamide gel electrophoresis) showing that within the experimental time range used, no polymerisation events are observed (Fig. 3.7) (Nyon *et al.* 2012).

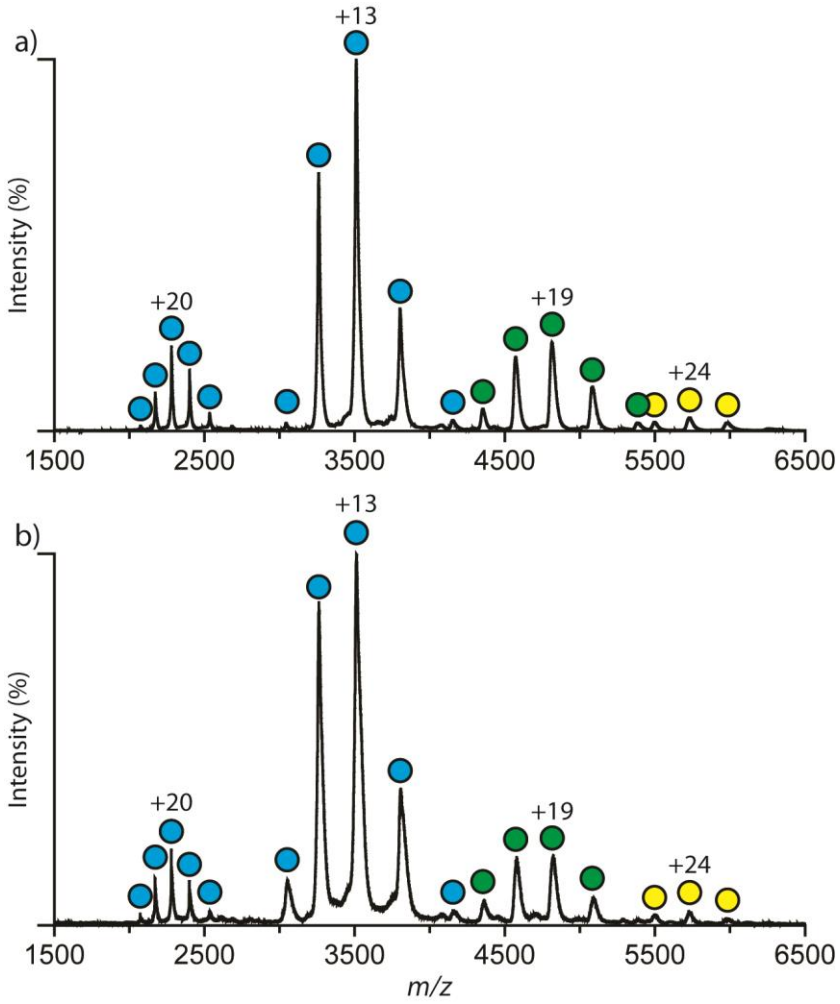

Figure 3.6. Control Mass Spectra of Wild Type α_1-Antitrypsin and K154N

Control mass spectra analysis of (a) wt and (b) K154N at time zero. The K154N substitution promotes no difference in oligomerisation behaviour or CSD (monomer = blue) with similar levels of unexpected higher-order oligomers (green = dimer, yellow = trimer). The monomer CSD, between *m/z* 1750 to 2600, represents a population of extended conformation AAT ions indicative of increased solvent accessible basic amino acids.

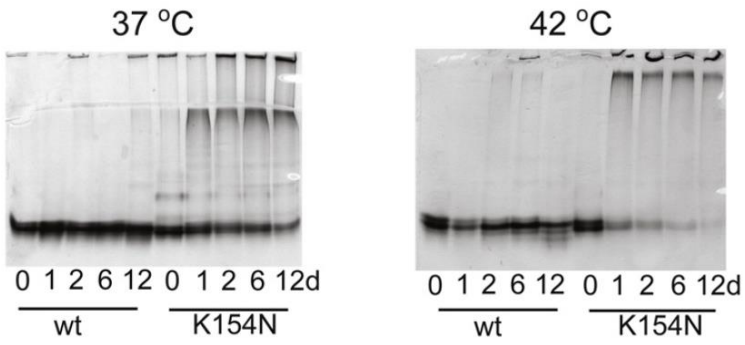

Figure 3.7. Native PAGE Gel Analysis of Wild Type α_1-Antitrypsin and K154N

7.5 % native PAGE polymerisation study of wt and K154N AAT (0.5 mg/ml) studied over twelve days at 37 ˚C and 42 ˚C. Time points indicate polymerisation levels at 0, 1, 2, 6 and 12 days. Polymerisation is indicated by the loss of the monomeric band lower down, shifting to bands higher up the gel.

Figure reproduced and adapted from Nyon *et al.* 2012

Comparisons of CCS after the 120 minute incubation period indicate no observable differences between both the wt and K154N at 20 and 34 ˚C (Fig. 3.8 and 3.9, a and b). At 39 ˚C, compared to the native wt conformation under the same conditions, the K154N mutation promotes a CCS increase from 2991.07 ($^{+}/_{-}$ 0) to 3239.71 ($^{+}/_{-}$ 49.39) Å^2 (Fig. 3.8c and 3.9c). This represents a CCS gain of 8.3 % compared to the wt, and an 8.2 % increase over the K154N prior to incubation (2994.66 $^{+}/_{-}$ 87 Å^2). Using the Student's paired t-test, the 8.3 % increase between the wt and K154N and 39 ˚C represents a significant change in collision cross section with a calculated p-value of 0.0065.

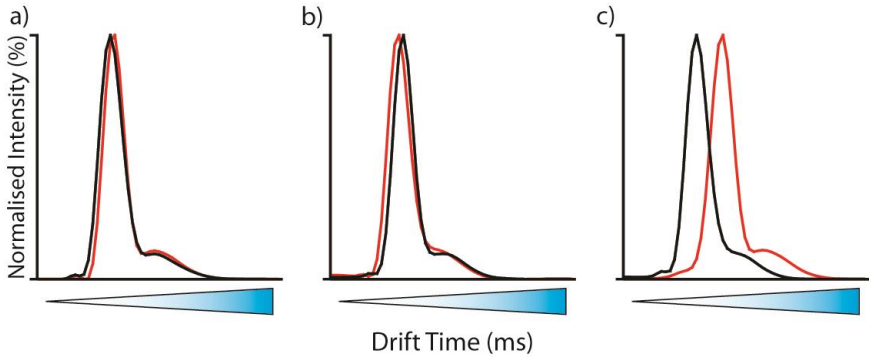

Figure 3.8. Drift Time Analysis of Incubated Wild Type α_1-Antitrypsin and K154N

t_D analysis of incubated K154N (red) and wt (black) AAT after 120 minute incubation at (a) 20 ˚C, (b) 34 ˚C, and (c) 39 ˚C. Observations indicate no difference between wt and K154N structural conformation at 20 and 34˚C. In contrast, at 39 ˚C the K154N exhibits an increased t_D when compared with the wt, under the same conditions, indicative of an extended structural conformation. These observations are expanded in figure 3.9. Data represents t_D values extracted from the FWHM of the lowest observed charge +12 ion (*m/z* 3798.5 and 3797.4, wt and K154N respectively).

These data show that the K154N substitution destabilises the AAT metastable fold, promoting the formation of an extended conformational state at biologically relevant temperatures. Additionally these observations are supported by thermal denaturation circular dichromism (CD) experiments performed by other individuals, which indicate that at increasing temperatures K154N unfolds more readily consistent with a reduced stability fold (Fig. 3.10) (Nyon *et al.* 2012).

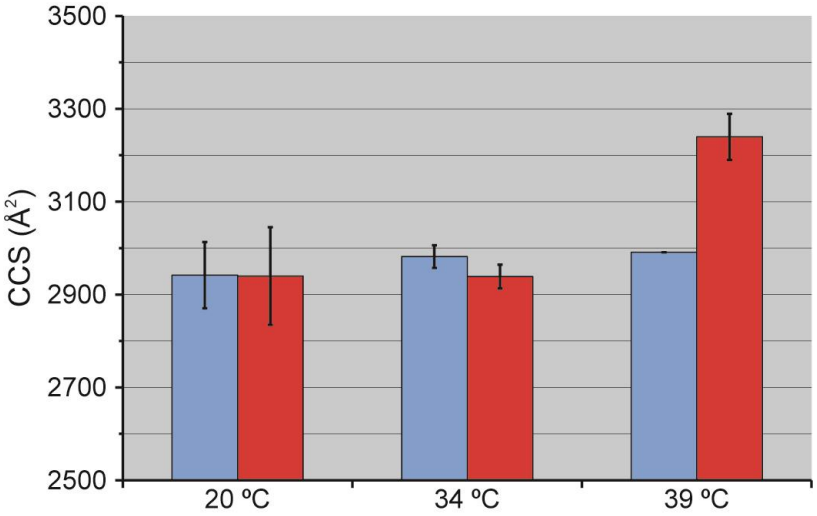

Figure 3.9. CCS Difference Analysis of Wild Type α_1-Antitrypsin and K154N

Despite the lowest observed charge AAT +12 ion (*m/z* 3798.5 and 3797.4, wt and K154N respectively) exhibiting no conformational change within MSD at 20 ˚C and 34 ˚C, at 39 ˚C a CCS increase of 8.3 % is observed for the K154N (red) against the wt (blue).

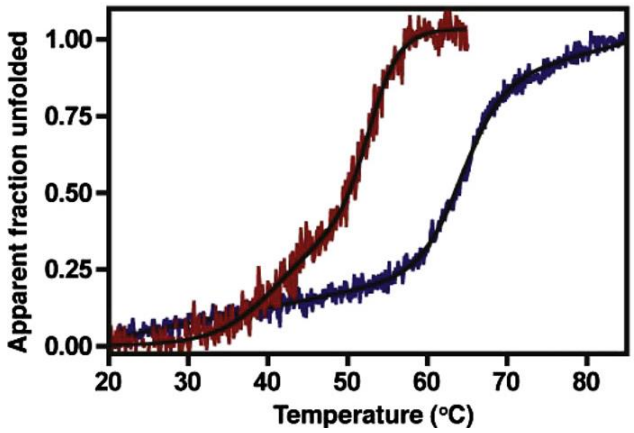

Figure 3.10. Thermal Denaturation Circular Dichromism of Wild Type α_1-Antitrypsin and K154N

Thermal dissociation CD spectra (mean ellipticity at 222 nm, n=10) for wt (blue) and K154N (red) (Nyon *et al.* 2012). Data supports IM data, indicating that the K154N substitution reduces the stability of the AAT metastable fold.

Figure reproduced and adapted from Nyon *et al.* 2012

In addition to the above, observations of K154N at 39 ˚C indicated aggregation events concluded from increased sample turbidity. These observations are supported by difficulties in sample ionisation due to increased sample viscosity. Mass spectra however

failed to provide evidence of these higher order oligomers in the gas phase, and therefore this line of questioning was taken no further.

3.4.3. Collision Induced Unfolding of α_1-Antitrypsin Wild Type and K154N

The folding pathway of the AAT metastable intermediate for neutrophil elastase binding activity requires passage of a global free-energy minimum. AAT mutations can increase the stability of folding intermediates, enabling long-lived conformations that can promote AAT polymerisation (Yamasaki *et al.* 2008, and Gooptu & Lomas 2009). CIU (previously described in section 1.4.3) has been applied to compare the conformation stability and number of structured unfolding intermediates of wt AAT and the K154N substitution. Advantageously this method provides a level of accuracy not currently reproducible using other instrumental techniques, enabling the study of stable structural unfolding intermediates and their relative contributions within the acquired arrival time distribution.

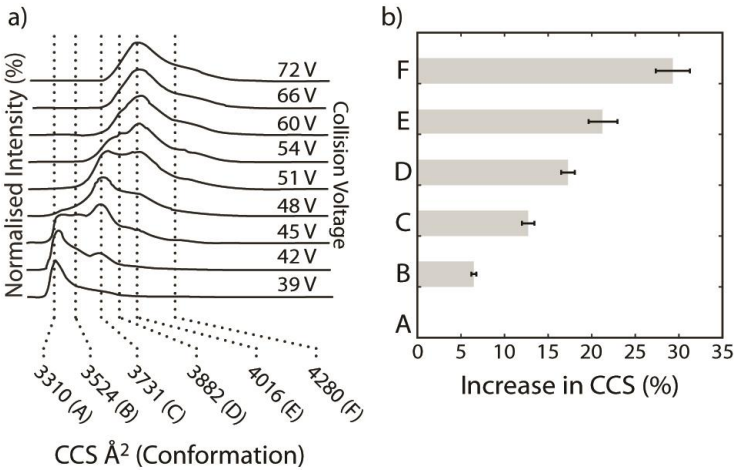

Figure 3.11. Collision Induced Unfolding Analysis of α_1-Antitrypsin Conformational Populations

a) CIU of the AAT lowest observed charge +12 ion (*m/z* 3798.5 and 3797.4, wt and K154N respectively) indicates that as the collision voltage of the TriWave Trap region is increased, AAT undergoes structural rearrangement from the native conformation (A) leading to five extended stable conformations (B to F). b) These conformations represent significant increases in CCS.

Collisional unfolding of the quadrupole isolated lowest observed charge +12 (*m/z* 3798.5 and 3797.4, wt and K154N respectively) monomer ion, was performed

using a TriWave Trap voltage ramp (6 to 72 V), prior to mobility separation. This conformational analysis (Fig. 3.11) has resolved five significantly extended conformational states (B-F) for both the K154N and wt. No unique, wt or K154N, structural conformations were observed.

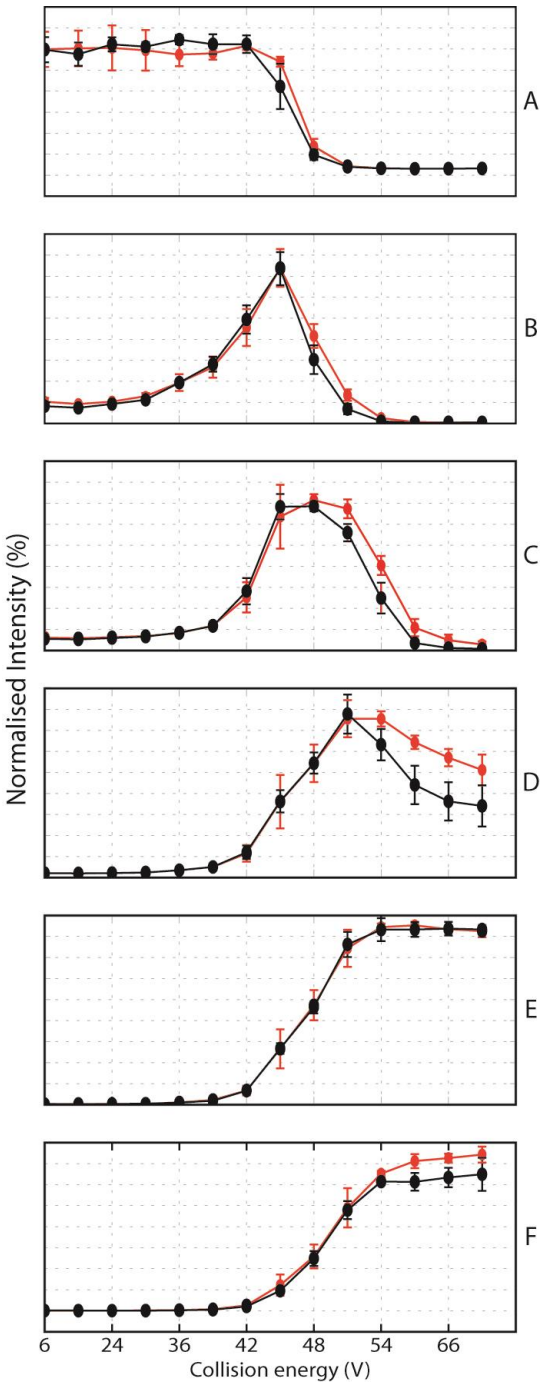

Figure 3.12. Comparative Stability Studies of Wild Type α_1-Antitrypsin and K154N Conformations

Comparison of conformational stability between K154N (red), and wt (black) ATT. Stability is measured as a function of the normalised intensity of each conformation, relative to other structural intermediates at each dissociation condition. Data indicates that relative to the wt significant increases in conformational stability of K154N are observed for conformational species C (3731 $Å^2$), and D (3882 $Å^2$).

The conformational stability of the wt and K154N AAT variants have been compared as a function of the average normalised intensity (%) of their unfolding intermediate conformations across the applied collision voltage ramp (Fig 3.12). Neither the native conformation (A, 3310 Å^2) nor the initial unfolding intermediate (B, 3524 Å^2) exhibit any differences in stability between the wt and K154N, when accounting for MSD. In contrast the second (C, 3731 Å^2) and third (D, 3882 Å^2) unfolding intermediates indicate an increased level of K154N structural stability relative to the wt between 50 to 60 V (Fig. 3.12). These results therefore support that the K154N substitution is able to promote the existence of extended non-native structures of increased structural stability, in line with observations using other AAT mutations.

3.5. Discussion

In all three models of AAT polymerisation, extended structural intermediates are required to promote the 'beads on string' polymer morphology (Dolmer & Gettins 2012, Krishnan & Gierasch 2011, Lomas *et al.* 1992, and Yamasaki *et al.* 2008). Our data has shown that under biologically permissible temperatures (39 ˚C) the K154N mutant exhibits an extended conformational state, proposed to be the polymerogenic intermediate required for AAT polymer formation (section 3.4.1). This temperature stabilised conformation exhibits a CCS 8.3 % greater than the native AAT metastable state.

Previous work has shown that AAT mutations within the reactive centre loop can increase the stability of folding intermediates, promoting long-lived conformations that enable AAT polymerisation (Yamasaki *et al.* 2008). Although the initial unfolding and native conformations, B and A respectively, exhibit no significant difference in stability between the wt and K154N (accounting for MSD), studies of the extended conformations have been used to apply support to our observations. At collision voltages of 50 to 60 V, the extended K154N conformations C and D exhibit increased structural stability, accounting for MSD, compared to the wt. These data therefore support that mutations encoded within AAT structural motifs, other than the reactive centre loop (K154 encoded within hF, Fig. 3.1), are able to promote stable structural intermediates. It should be noted however, that neither C nor D have been identified as being biologically relevant

polymerogenic intermediates, and are solely evidence of increased structural stability along the unfolding pathway of the K154N mutant.

If the CIU intermediate B represents the polymerogenic, or a polymerogenic-like intermediate, required for AAT polymerisation our data indicates that this conformation is exhibited by both the wt and K154N, at similar collision voltages (45 V, Fig. 3.12). These data therefore indicate that collisional activation may be able to promote structural conformations similar to the AAT polymerogenic intermediate previously observed (section 3.4.1) in both the wt and K154N AAT variants. Conferring support from the CIU structured unfolding intermediates C and D, the K154N substitution may promote an energetically unfavourable refolding landscape upon formation of the polymerogenic intermediate. This folding landscape would therefore promote a long-lived intermediate required for AAT polymerisation. In contrast, the wt sequence would exhibit an energy landscape that favoured refolding to the metastable state.

The use of quadrupole isolation for CIU analysis prevents comparisons with the incubation study, due to differences in the experimental methods. Therefore, we are unable to confirm that the CIU structural intermediate B is the temperature stabilised polymeric intermediate observed in the incubation studies due to these experimental differences promoting variations in the CCS.

3.6. Summary and Future Directions

Work here has shown that the polymerogenic intermediate required for AAT K154N polymerisation is propagated under biologically permissive temperatures, representing a significant increase in CCS equal to 8.3 %. This substitution additionally stabilises the unfolding intermediates, required for polymerisation activity. Other data presented within Nyon *et al.* (2012), including CD and native-PAGE, have supported our observations.

These data are unable however to confirm or refute either three of the proposed models of AAT polymerisation (section 3.1.3), supporting only common observations: extended, increased stability structural conformations, propagated under biological conditions. Studies to confirm these models therefore remain a direction for future analysis.

3.7. References

Axelsson, U., & Laurell, C., 1965 "Hereditary Variants of Serum α1-Antitrypsin," American Journal of Human Genetics, Vol.17 pp.466-472

Belorgey, D., *et al.* 2007 "Protein Misfolding and the Serpinopathies," Prion, Vol.1 pp.15-20

Blanco, I., Fernández, E., & Bustillo, E., 2001 "Alpha-1-Antitrypsin PI Phenotypes S and Z in Europe: An Analysis of the Published Surveys," Clinical Genetics, Vol.60 pp.31-41

Carrell, R., & Travis, J., 1985 "α1-Antitrypsin and the Serpins: Variation and Countervariation," Trends in Biochemical Sciences, Vol.10 pp.20-24

Chua, F., & Laurent, G., 2006 "Neutrophil Elastase: Mediator of Extracellular Matrix Destruction and Accumulation," Proceedings of the American Thoracic Society, Vol.3 pp.424-427

Dabbagh, K., *et al.*, 2001 "Alpha-1-Antitrypsin Stimulates Fibroblast Proliferation and Procollagen Production and Activates Classical MAP Kinase Signalling Pathways," Journal of Cellular Biology, Vol.186 pp.73-81

Dafforn, T., 1999 "A Kinetic Mechanism for the Polymerisation of α_1-Antitrypsin," Journal of Biological Chemistry, Vol.14 pp.9548-9555

Devlin, G.L., 2002 "Acid Denaturation of α_1-Antitrypsin: Characterization of a Novel Mechanism of Serpin Polymerisation," Journal of Molecular Biology, Vol.324 pp.859-870

Dolmer, K., & Gettins, P., 2012 "How the Serpin α_1-Proteinase Inhibitor Folds," Journal of Biological Chemistry, Vol.287 pp.12425-12432

Dunstone, M.A., *et al.*, 2000 "Cleaved Antitrypsin Polymers at Atomic Resolution" Protein Science, Vol.9 pp.417-420

Egelund, R., *et al.*, 2001 "A Regulatory Hydrophobic Area in the Flexible Joint Region of Plasminogen Activator Inhibitor-1, Defined with Fluorescent Activity-neutralizing ligands," Journal of Biological Chemistry, Vol.276 pp.13077-13086

Ekeowa, U., *et al.* 2010 "Defining the Mechanism of Polymerisation in the Serpinopathies," Proceedings of the National Academy of Sciences, Vol.107 pp.17146-17151

Fra, A., *et al.*, 2012 "Three New Alpha1-Antitrypsin Deficiency Variants Help to Define a C-Terminal Region Regulating Conformational Change and Polymerisation," PloS One, Vol.7:e38405

Gooptu, *et al.*, 2000 "Inactive Conformation of the Serpin α1-Antichymotrypsin Indicates Two Stage Insertion of the Reactive centre loop Implications for Inhibitory Function," Publications of the National Academy of Sciences, Vol.97 pp.67-72

Gooptu, B., & Lomas, D.A., 2008 "Polymers and Inflammation: Disease Mechanisms of the Serpinopathies," The Journal of Experimental Medicine," Vol.205 pp.1529-1534

Gooptu, B., & Lomas, D.A., 2009 "Conformational Pathology of the Serpins: Themes, Variation and Therapeutic Strategies," Annual Review of Biochemistry, Vol.78 pp.147-176

Gooptu, B., *et al.*, 2009 "Crystallographic and Cellular Characterization of Two Mechanisms Stabilizing the Native Fold of Alpha1-Antitrypsin: Implications for Disease and Drug Design," Journal of Molecular Biology, Vol.387 pp.857-868

Griese, M., *et al.*, 2007 "Alpha1-Antitrypsin Inhalation Reduces Airway Inflammation in Cystic Fibrosis Patients," The European Respiratory Journal, Vol.29 pp.240-250

Heimburg, H., & Haupt, H., 2011 "Charakterisier Von α1 X-glykoprotein Als Chymotrypsin Inhibitor Des Humanplasmas, "Clinica Chimica Acta, Vol.12 pp.116-118

Hutchinson, D.C.S., Tobin, M.J., & Cook, P.J.L., 1983 "Alpha$_1$ Antitrypsin Deficiency: Clinical and Physiological Features in Heterozygotes of Pi Type SZ," British Journal of Diseases of the Chest, Vol.77 pp.28-34

Huntington, J., *et al.* 1999 "A 2.6 Å Structure of a Serpin Polymer and Implication for Conformational Disease," Journal of Molecular Biology, Vol.293 pp.449-455

Huntington, J.A., 2011 "Serpin Structure, Function and Dysfunction," Journal of Thrombosis and Haemostasis, Vol.9 pp.26-34

Jacobsson, K., 1955 "Studies on Fibrinogen II. Studies on the Trypsin and Plasmin Inhibitors in Human Blood Serum," Scandinavian Journal of Clinical & Laboratory Investigation, Suppl. 14:55-102

James, E.L., & Bottomley, S.P., 1998 "The Mechanism of α_1-Antitrypsin Polymerisation Probed by Fluorescence Spectroscopy," Archives of Biochemistry and Biophysics, Vol.356 pp.296-300

Janciauskiene, S.M., *et al.* 2011 "The Discovery of α1-Antitrypsin and its Role in Health and Disease," Respiratory Medicine, Vol.105 pp.1129-1139

Kelly, E., Greene, C., & McElvaney, N.G., 2008 "Targeting Neutrophil Elastase in Cystic Fibrosis," Expert Opinion on Therapeutic Targets, Vol.12 pp.145-157

Kelly, E., *et al.* 2010 "Alpha-1 Antitrypsin Deficiency," Respiratory Medicine, Vol.104 pp.763-772

Kim, D., & Yu, M.H., 1996 "Folding Pathway of Human α_1-Antitrypsin: Characterization of an Intermediate That is Active but Prone to Aggregation," Biochemical and Biophysical Research Communications, Vol.226 pp.378-384

Knaupp, A.S., & Bottomley, S.P., 2009 "Serpin Polymerisation and its Role in Disease - the Molecular Basis of α1-Antitrypsin Deficiency," IUBMB Life, Vol.61 pp.1-5

Knappstein, S., 2004 "α 1 -Antitrypsin Binds to and Interferes with Functionality of EspB from Atypical and Typical Enteropathogenic Escherichia coli Strains," Infection and Immunity, Vol.72 pp.4344-4350

Larsson, C., 1978 "Natural History and Life Expectancy in Severe Alpha$_1$-Antitrypsin Deficiency, Pi Z," Acta Medica Scandinavica, Vol.204 pp.345-351

Lomas, D.A., *et al.*, 1992 "The Mechanism of Z α_1-Antitrypsin Accumulation in the Liver," Nature, Vol.357 pp.605-607

Lomas, D.A., *et al.*, 1993 "Effect of the Z Mutation on the Physical and Inhibitory Properties of α_1-Antitrypsin," Biochemistry, Vol.32 pp.500-508

Laurell, C.B., & Eriksson, S., 1963 "The Electrophoretic α_1-Globulin Pattern of Serum in α_1-Antitrypsin Deficiency," Scandinavian Journal of Clinical & Laboratory Investigation, Vol.15 pp.132-140

Lungarella, G., *et al.*, 2008 "The Dual Role of Neutrophil Elastase in Lung Destruction and Repair," International Journal of Biochemistry and Cell Biology, Vol.40 pp.1287-1296

Massi, G., & Chiarelli, C., 1994 "Alpha1-Antitrypsin: Molecular Structure and the Pi System," Acta Paediatrica, Supplementary Vol.394 pp.1-4

McElvaney, N. G., *et al.*, 1997 "Baseline Characteristics of Enrolees in the National Heart, Lung and Blood Institute Registry of α_1-Antitrypsin Deficiency," CHEST, Vol.111 pp.393-403

Nyon, M.P., *et al.*, 2012 "Structural Dynamics Associated with Intermediate Formation in an Archetypal Conformation Disease," Structure, Vol.3 pp.504-512

Patschull, A.O.M., *et al.*, 2011 "Therapeutic Target-Site Variability in α_1-Antitrypsin Characterized at High Resolution," Acta Crystallographica, Vol.F67 pp.1492-1497

Parfrey, H., *et al.*, 2003 "Targeting a Surface Cavity of alpha 1-Antitrypsin to Prevent Conformational Disease," Journal of Biological Chemistry, Vol.278 pp.33060-33066

Rajpara, A., Erickson, C., & Driscoll, M., 2010 "Review of Alpha-1-Antitrypsin Deficiency Associated Panniculitis," The Open Dermatology Lab, Vol.4 pp.97-100

Sandford, A., *et al.*, 1999, "Z and S Mutations of the α_1-Antitrypsin Gene and the Risk of Chronic Obstructive Pulmonary Disease," American Journal of Respiratory Cell and Molecular Biology, Vol.20 pp.287-291

Scarff, C., *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 Spectrometry," Rapid Communications in Mass Spectrometry, Vol.22 pp.3297-3304

Silverman, G.A., *et al.*, 2001 "The Serpins are an Expanding Superfamily of Structurally Similar but Functionally Diverse Proteins. Evolution, Mechanism of Inhibition, Novel Functions, and a Revised Nomenclature," Journal of Biological Chemistry, Vol.276 pp.33293-33296

Taggart, C., 2000 "Oxidation of Either Methionine 351 or Methionine 358 in Alpha 1-Antitrypsin Causes Loss of Anti-Neutrophil Elastase Activity," Journal of Biological Chemistry, Vol. 275, pp.27258-27265

Tew, D.J., & Bottomley, S.P., 2001 "Probing the Equilibrium Denaturation of the Serpin α_1-Antitrypsin with Single Tryptophan Mutants; Evidence for Structure in the Urea Unfolded State," Journal of Molecular Biology, Vol.313 pp.1161-1169

Yamasaki, M., *et al.*, 2008 "Crystal Structure of a Stable Dimer Reveals the Molecular Basis of Serpin Polymerisation," Nature, Vol.455 pp.1255-1258

Yamasaki, M., *et al.* 2011 "Molecular Basis of α_1-Antitrypsin Deficiency Revealed by the Structure of a Domain-Swapped Trimer," EMBO Reports, Vol.12 pp.1011-1017

Zhou, A., Carrell, R.W., & Huntington, J.A., 2001 "The Serpin Inhibitory Mechanism is Critically Dependent on the Length of the Reactive Centre Loop," Journal of Biological Chemistry, Vol.276 pp.27541-27547

4. Analysis of the Sgt1 Dimerisation Domain, Required for Yeast Kinetochore Assembly

---

Saccharomyces cerevisiae Sgt1 dimers mediate binding between Hsp90 and Skp1, to initiate chromosome separation. Mutations of the Sgt1 dimerization domain are known to inhibit Sgt1: Skp1 binding, arresting the cell cycle at the G2/M interface. Work here has shown that this dimer is stabilised by an Ascomycota specific structural loop within the dimerisation domain, with potential contributions from other domains. Using MS/MS, we have shown that this dimer does not represent the structural minimum for Skp1 binding.

4.1.1. The Kinetochore

The *Saccharomyces cerevisiae* kinetochore (Fig. 4.1) is a complex that comprises more than 70 different proteins, organised into a variety of sub-complexes.[18] Additionally these sub-complexes can be divided into three different layers: the inner DNA anchoring layer, a dynamic central bridging/linker layer, and the outer layer that functions as the microtubule binding interface (Westermann, Drubin & Barnes 2007). Overall, this complex mediates the interactions between chromosomes and spindle microtubules required for cell division (Bansal *et al.* 2009a, Cho *et al.* 2011, and Westermann, Drubin & Barnes 2007).

4.1.2. The Inner Kinetochore Sub-Complex CBF3

The inner kinetochore contains two multimeric sub-complexes, the centromeric nucleosome and the CBF3 (centromere binding factor 3) (Westermann, Drubin and Barnes 2007). Of interest here is the CBF3, an essential component of kinetochore initiation and formation, required for passage of the G2/M spindle checkpoint assembly interface (Bansal *et al.* 2009a, and Makhnevych & Houry 2012).

[18] For a larger overarching review on the topic of kinetochore sub-complexes the author directs the reader to the review by Westermann, Drubin & Barnes (2007)

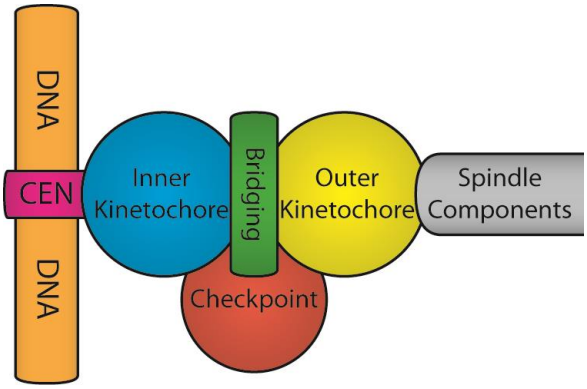

Figure 4.1. Three Domain Architecture of the Kinetochore

The *S. cerevisiae* kinetochore exhibits a three-domain architecture, which anchors spindle fibres to DNA. The inner domain binds to the centromere DNA (CEN) and is required to establish kinetochore formation and activity. The kinetochore attaches to the spindle components required for chromosome separation at the S phase through the outer kinetochore components. The inner and outer kinetochore domains are connected through a central bridging/linker layer. Together, the overall structure combined with required checkpoint proteins allows passage of the G2/M interface and the sequential separation of sister chromatids (Westermann, Drubin & Barnes 2007).

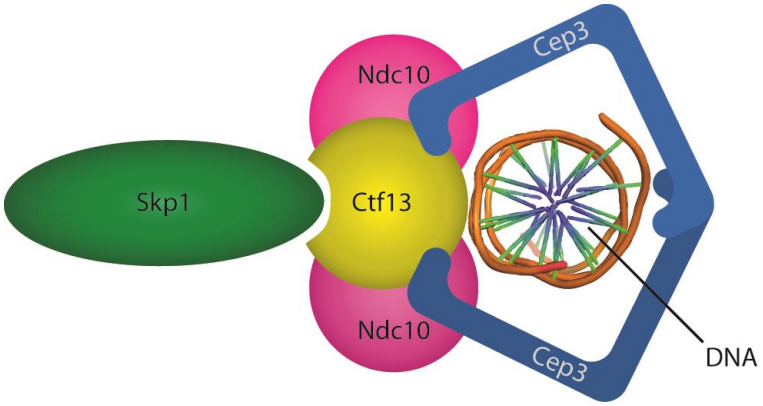

Figure 4.2. Structure of the Inner Kinetochore: The CBF3 Complex at the DNA Binding Interface

Skp1 and Ctf13 form the CBF3 initiation complex, allowing binding of the Cep3 dimer (Kaplan, Hyman & Sorger 1997). Dimerisation of Cep3 creates a cleft in the quaternary structure required for Cep3-DNA binding at CDEIII and CDEII (Rodrigo-Brenni *et al.* 2004). The model proposed by Purvis & Singleton (2008), suggests that Ctf13 binds within the Cep3 cleft between the two CDE components, opposite the homodimerisation domains, stabilising DNA binding. Ndc10 associates with the CBF3 by binding to Ctf13. Complex stability promotes further Ndc10 interactions outside of the CBF3 complex (Cho *et al.* 2011). Skp1 feedback interactions with Bub1 additionally act as a checkpoint of kinetochore assembly as a measure of complex stability (Kitagawa *et al.* 2003). Image not to scale, DNA PDB ID; 3BSE.

The CBF3 is a heteromeric complex that comprises four essential proteins (Fig. 4.2): Skp1 (section 4.1.2.4), Ctf13 (section 4.1.2.1), Cep3 (section 4.1.2.2), and Ndc10 (section 4.1.2.3). This complex is anchored to the DNA at the 25 bp long CDEIII (centromere determining element 3) motif of the centromere (Ortiz *et al.* 1999, Kitagawa & Hieter 2001, and Makhnevych & Houry 2012).

4.1.2.1. Ctf13 Within the CBF3

Ctf13 (Uniprot ID; P52286, 22.3 kDa, 194 amino acids) and Skp1 (section 4.1.2.4) form the initiation complex of the CBF3 complex. Structurally, Ctf13 mediates the interactions between Skp1 and the other components of the CBF3 complex. The Ctf13 internal binding domain (amino acids 139-336) binds Cep3 towards its C-terminal end, with Skp1 and Ndc10 binding to the N-terminus of this sequence (Russell, Grancell & Sorger 1999). Regulation of Ctf13 and Ndc10 binding events are currently proposed to occur through one of two methods, post-translational phosphorylation of Ctf13 by Skp1 (Kaplan, Hyman & Sorger 1997), or the involvement of Hsp90 chaperones (Stemmann *et al.* 2002).

As Ctf13 is structurally unstable within the cytosol (Russell, Grancell & Sorger 1999), it currently represents the *in vitro* limiting step for characterising both the individual protein and CBF3 complex.

4.1.2.2. Cep3 Within the CBF3

Cep3 (Uniprot ID; P40969, 71.4 kDa, 608 amino acids) dimers regulate the DNA binding activity of the CBF3. Cep3 binds DNA at the CDEIII domain of the centromere (Fig. 4.3). Although Cep3 is able to interact with the CDEIII regardless of further interactions with Skp1 or Ctf13, stable binding requires formation of the Skp1: Ctf13: $Cep3_2$ complex (Russell, Grancell & Sorger 1999).

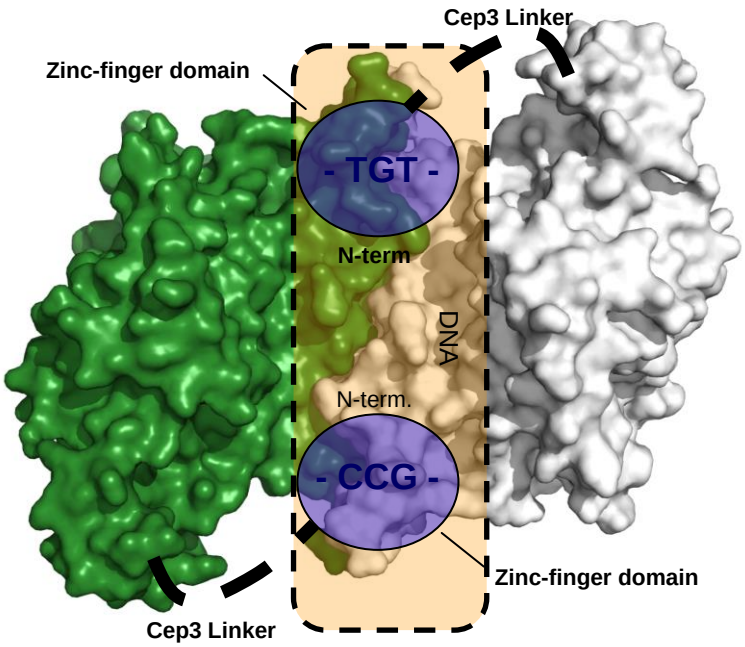

Figure 4.3. DNA Binding Model of Cep3

DNA binds within a groove formed on the underside of the Cep3 dimerisation interface. DNA binding is promoted through the Cep3 N-terminal Zn_2Cys_6 class zinc finger domain, to a conserved CCG site within the CDEIII domain (Espelin, Kaplan & Sorger 1997) with additional interactions with a TGT motif 3' of the CDEII motif (Espelin, Kaplan & Sorger 1997, and Jehn *et al.* 1991). The linker between the zinc-finger and C-terminal domains confers the necessary flexibility of the N-terminus to bind DNA in this manner. These binding events sequentially stabilise interactions on the opposite side of the DNA binding groove (Purvis & Singleton 2008). PDB ID: 2VEQ.

Figure reproduced and adapted, with permission, from Willhöft 2013

4.1.2.3. Ndc10 Within the CBF3

Ndc10 (Uniprot ID; P32504, 112 kDa, 956 amino acid) within the CBF3 is dimeric and has been identified as the final component to associate. Structurally Ndc10 is a tri-domain protein, that contains, an N-terminal DNA binding domain, central dimerisation domain (Cho *et al.* 2011), and C-terminal regulatory domain (Peng *et al.* 2011). Ndc10 binds to DNA non-specifically, orientating at the CDEIII when complexed with the CBF3 (Cho *et al.* 2011, and Perriches & Singleton 2012).

Ndc10 performs a scaffolding role within the CBF3, connecting this complex to other components of the kinetochore (Cho *et al.* 2011, and Perriches & Singleton 2012). Ndc10 recruits the histone chaperone Scm3 to form bridging interactions with the Cse4

component of the centromeric nucleosome. Although it may be suggested that the Ndc10 functions to initiate further layers of the kinetochore to form, it cannot be concluded to proceed within this role as a checkpoint protein (Camahort *et al.* 2007, and Mizuguchi *et al.* 2007).

The activity and localisation of Ndc10 is regulated by Ipl1 and CK2 mediated phosphorylation of the C-terminus. Ndc10 phosphorylation by Ipl1 protein inhibits Ndc10 interactions with Bir1, preventing targeting of the protein to the anaphase spindles. In contrast, phosphorylation by the highly conserved phosphorylation protein CK2 prevents the localisation and incorporation of Ndc10 within the CBF3 kinetochore complex (Peng *et al.* 2011).

4.1.2.4 Skp1 Within the CBF3

The *S. cerevisiae* protein Skp1 (Suppressor of kinetochore protein 1, Uniprot ID; P52286, 22.3 kDa, 194 amino acids) (Fig. 4.4) is required for initiation of the CBF3 complex through Hsp90 mediated interactions with Ctf13. This Hs90: Skp1 interaction is promoted by interactions mediated by Sgt1 (Lingelbach & Kaplan 2004), and is essential for passage of the G2/M cell cycle transitions (Zhang *et al.* 2008). Initiation of the CBF3 complex by Hsp90 will be discussed in increased detail in section 4.1.4.

Further to its requirement at the G2/M as part of the CBF3 complex, Skp1 is additionally required at the G1/S transition as a component of the SCF functional E3 ubiquitinase (Yoshida, Murakami & Tanaka 2011).

4.1.3. Hsp90 Structure and Function

The *S. cerevisiae* protein, Hsp90 (Heat shock protein 90 kDa, Uniprot ID; P02829, 82.4 kDa, 709 amino acids) [19] is a dimeric, ATP dependent molecular chaperone that comprises 1-2 % of the *S. cerevisiae* soluble cell fraction (Borkovich *et al.* 1989, and Neckers & Workman 2012). The Hsp90 class of proteins is highly conserved in nature,

[19] Synonym; Hsp82. The Sgt1 *S. cerevisiae* homologue provides an exception to the rule of Hsp# proteins classified in terms of their molecular weight. Hsp90 is typically used for naming due to the functional cross specie homologues.

present in eukaryotes and prokaryotes (Makhnevych & Houry 2012),[20] and therefore represents an essential component of organism survival (Johnson 2012). In addition to the inducible form of Hsp90 outlined here, the *S. cerevisiae* genome additionally encodes the constitutively expressed, Hsc82 (Uniprot ID; P15108).

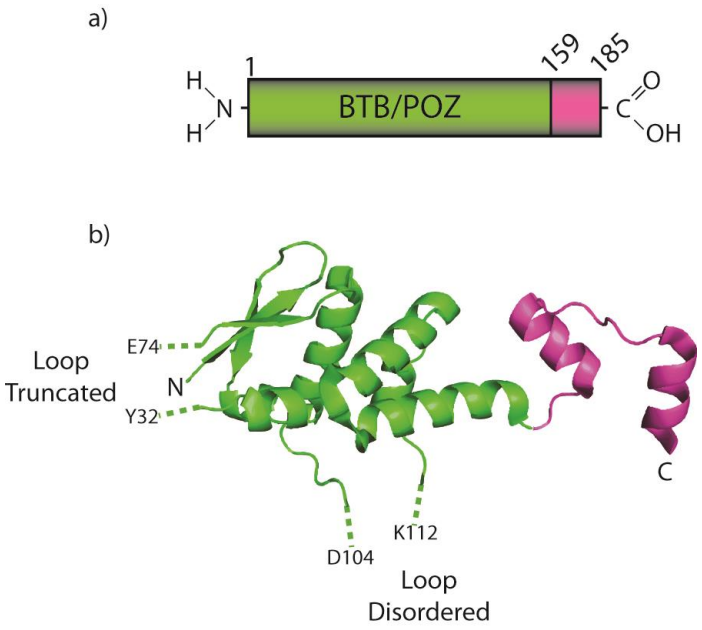

Figure 4.4. Domains and Crystal Structure of *S. cerevisiae* Skp1

a) Skp1 contains a single structured domain, comprising the BTB/POZ (amino acids 1-159), and a C-terminal F-box interaction domain (magenta, amino acids 162-185). The BTB/POZ domain is the binding site for Sgt1. The F-box domain allows binding of Ctf13 required promoting formation of the CBF3, along with chaperone activity in the SCF complex (Willhöft 2013, and Zhang *et al.* 2008). b) The intact crystal structure of this domain has been resolved to 2.6 Å, highlighting both domains (PDB ID; 3MKS).

Figure 4.5. Domains of *S. cerevisiae* Hsp90

Hsp90 comprises, an N-terminal ATPase domain (red, amino acids 1-220) which further acts as a drug binding domain, a central (blue, amino acids 255-540) client binding domain, and a C-terminal (green, amino acids 541-709) homo-dimerisation domain (Pearl & Prodromou 2006). There is currently no complete PDB crystal structure of the *S. cerevisiae* Hsp90, with the largest available structure comprising amino acids 1-677 (PDB ID; 2CG9, resolved to 3.1Å).

S. cerevisiae isoform of Hsp90 comprises a three domain architecture: an N-terminal ATPase domain, central co-chaperone binding domain, and a C-terminal

[20] There are no known members in archaea (Pearl & Prodromou 2006).

dimerisation domain (Fig. 4.5). Hsp90 dimerisation at the C-terminus occurs at a K_d of 60 nM, and is therefore observed as a constitutive dimer (Richter *et al.* 2001).

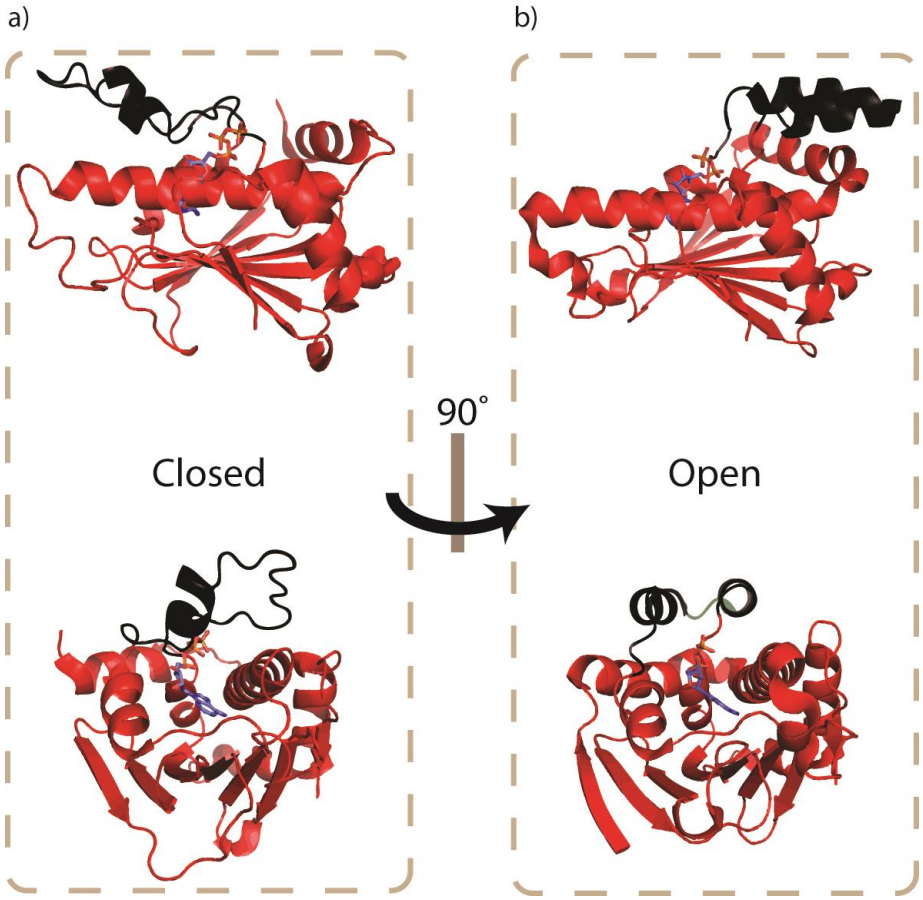

Figure 4.6. N-terminal ATPase Domain of Hsp90

a) The Hsp90 ATPase domain in its ATP bound state exhibits closed lid segment conformation (PDB ID; 2CG9, *S. cerevisiae*, resolved to; 3.1 Å). b) ATP hydrolysis, producing ADP, promotes conformational rearrangement of the lid segment to an open conformation (PDB ID; 2IOR, *E. coli* - HtpG, resolved to; 1.65 Å). The 90˚ structural rotation highlights the nucleotide-binding pocket. Despite the species difference, the highly conserved nature of the protein, and its N-terminal ATPase domain, allows direct comparison between homologues (Makhnevych & Houry 2012).

The conformational state of the Hsp90 N-terminal ATPase domain regulates its co-chaperone binding activity. Binding of ATP, includes conformational changes in the N-terminus and middle segment catalytic loop (Fig. 4.6). These changes aid positioning of the ATP hydrolysis apparatus (Ali *et al.* 2006), which represents the rate limiting step of the hydrolysis mechanism (Cunningham, Krukenberg & Agard 2008). Dimerisation of the

N-terminus has been additionally shown as a requirement for ATP hydrolysis (Vaughan *et al.* 2009). Through changes in the monomeric/dimeric state of the Hsp90 N-terminal domain and ATP/ADP conformational states, co-chaperone and client binding activity can be regulated. The current model of the ATP dependent conformational cycle is outlined in figure 4.7.

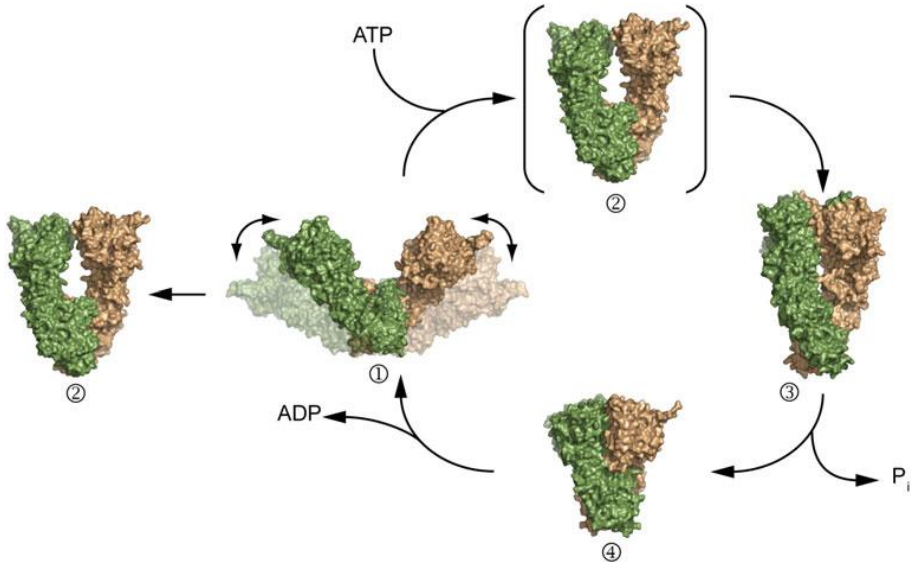

Figure 4.7. ATP Dependent Conformational Hydrolysis Cycle of Hsp90

Structural rearrangement of the Hsp90 dimer occurs in an ATP dependent manner. (1) From an open dimeric conformation, (3) binding of ATP induces conformational rearrangement to a closed state (2) through a proposed intermediate step identified using the crystal structure of Grp94 (Humans, endoplasmic reticulum Hsp90 homologue, Dollins *et al.* 2007). This structure (2) however, may further represent a non-catalytic ATP/ADP unbound, conformation (Krukenburg *et al.* 2011). (4) Hydrolysis of bound ATP causes further compaction of the protein, releasing ADP, and allowing the cycle to restart. Despite the cross species nature of these conformational states, a high level of sequence similarity across Hsp90 homologues supports the relevance of the model to *S. cerevisiae* (Vaughan *et al.* 2009).

Figure reproduced from Krukenberg *et al.* 2011

4.1.3.1. Hsp90 Co-chaperone Activity

Hsp90 interacts with a range of co-chaperones, including kinases and transcription factors, which are typically termed 'clients' (Hartl, Bracher & Hayer-Hartl 2011, and Verghese *et al.* 2012). Many of these clients are at the epicentre of genetic cascades, and therefore Hsp90 is closely connected with a number of oncogenes (genes with the potential

to cause cancer). Subsequently, Hsp90 and its binding partners are an area actively studied in terms of drug design to inhibit cancer-related pathways (Neckers *et al.* 2012).

Currently, no specific sequence or structure has been identified to promote targeting of Hsp90 to client proteins for folding support. Co-chaperones that mediate the activity between Hsp90 and these clients have been shown to contribute to this recognition process (Taipale *et al.* 2012, Vaughan *et al.* 2006, and Willhöft *et al. In Preparation*). These include the protein Sgt1 (described below) which acts as an Hsp90 client adaptor protein to promote Skp1 binding (Catlett & Kaplan 2006).

4.1.4. Sgt1

The *S. cerevisiae* protein, Sgt1 (Uniprot ID; Q08446, 44.9 kDa, 395 amino acids), was originally identified as a dosage suppressor of G two allele of skp1 mutants (Kitagawa *et al.* 1999). Sgt1 homologues have been identified in other species, including humans where two splice variants are known to exist (Niikura & Kitagawa 2003).

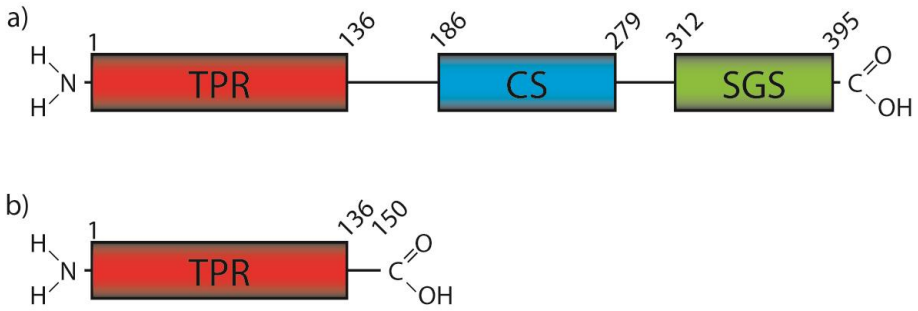

Figure 4.8. Domains of Wild Type Sgt1 and the TPRκ Mutant

a) Sgt1 contains three domains (TPR in red, CS in blue and SGS in green), connected by flexible linker regions (Bansal *et al.* 2009a and Bansal *et al.* 2009b). b) Relevant to work performed here, the Sgt1 TPRκ truncate has been used to study the influence of the CS and SGS domains on the oligomeric state. This protein is defined as the TPR domain plus capping helix (κ).

Structurally, Sgt1 (Fig 4.8a) comprises a three structured domain architecture: an N-terminal TPR (Tetratricopeptide repeat) domain (section 4.1.1.1), central CS (CHORD protein- and Sgt1- specific) domain, and a C-terminal SGS (Sgt1 specific sequence) domain

(section 4.1.1.2). Flexible linkers connect all three domains (Bansal, Abdulle, & Kitagawa 2004, and Bansal *et al.* 2009a).

The tertiary structure is highly conserved between Sgt1 homologues. In contrast, the number of Sgt1 protomers bound at the quaternary level varies between species. In *S. cerevisiae* Sgt1 exists *in vivo* as a dimer (Bansal *et al.* 2009a) whilst the human homologue is monomeric. Oligomers of the *A. thaliana* and *H. vulgare* Sgt1 homologues are concentration dependent (Nyarko *et al.* 2007).

4.1.4.1. The Sgt1 TPR Domain

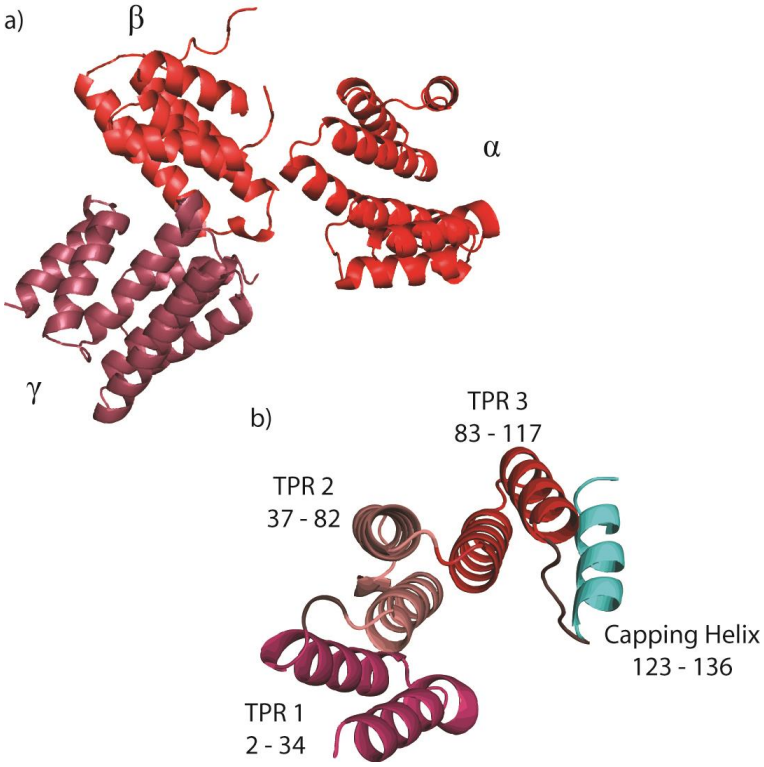

Figure 4.9. Crystal Structure of the TPR Domain

a) Crystal structure of the Sgt1 dimerisation domain resolved to 4.3 Å. The structure has been solved using the dimerisation domain only TPRκ mutant (amino acids 137-150 remain unresolved in the final structure), in an unexpected trimer (α, β and γ). Additions of sequential Sgt1 monomers occur with a 90˚ rotation about the binding interface. b) The curved binding interface of the dimerisation domain is promoted by the organised stacking of TPR repeats against the capping helix (Willhöft 2013, and Willhöft *et al. In Preparation*). Structure currently unpublished in the PDB.

TPR motifs, jointly identified in 1990 (Hirano *et al.* 1990, and Sikorski *et al.* 1990), are a widespread structural motif in nature (> 5 000 identified containing proteins) (Zeytuni & Zarivach 2012). These motifs are defined by their 34 amino acid, anti-parallel

helix-turn-helix topology (Main, Jackson & Regan 2003, and Main *et al.* 2003), and are involved in a range of protein-protein interactions.

TPR mediated interactions are dependent on peptide sequence and variations in the number of repeats contributing to the larger super secondary and tertiary structure (Zeytuni & Zarivach 2012). In *S. cerevisiae* TPR containing proteins are involved in a broad range of functions and interactions. Ssn6 contains ten consecutive N-terminal TPR motifs which enable Opi1 and Sin3 binding. This interaction is subsequently able to inhibit phospholipid biosynthesis (Jäschke *et al.* 2011). In contrast, Sti1 contains nine TPR repeats spread across three domains (TPR1, TPR2A and TPR2B). Of these three TPR domains, TPR1 and TPR2A function as binding domains required to modulate the activity of Hsp70 and Hsp90 respectively (Flom *et al.* 2006).

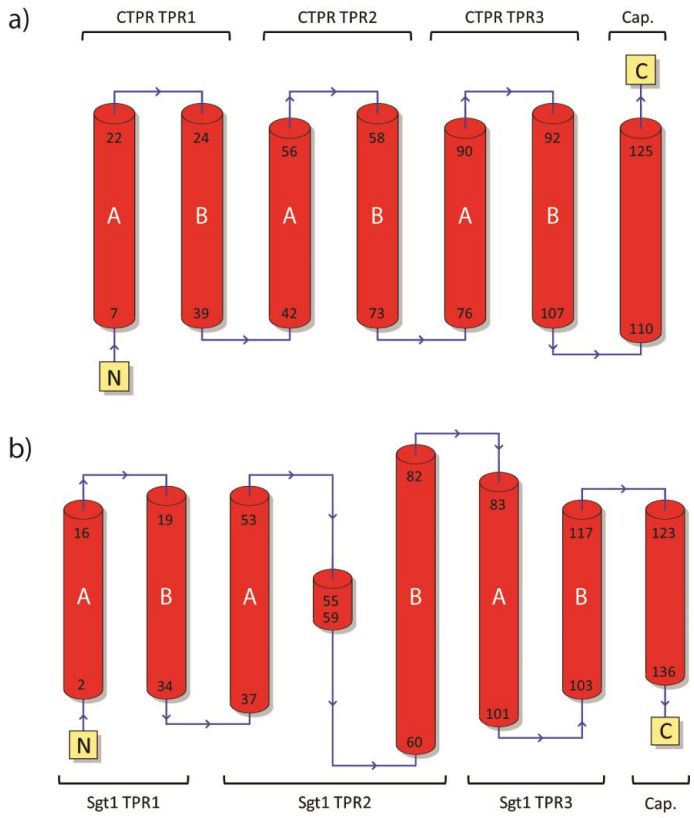

Figure 4.10. Topology Comparison of the Consensus and Sgt1 Sequence TPR Domains

TPR topology of the (a) consensus sequence (CTPR, Main, Jackson & Regan 2003) and (b) *S. cerevisiae* Sgt1 dimerisation domain. Despite amino acid sequence differences, the core alpha helical fold is conserved. Variations in the primary structure of the two proteins, resulting from sequence substitution and insertions, promote differences in helix length, along with an inserted sequence between TPR2 helix A, and B. Please note A and B highlight the N- and C-terminal helices within each TPR repeat respectively.

Figure reproduced and adapted, with permission, from Willhöft 2013

The Sgt1 TPR domain (*S. cerevisiae* crystal structure Fig. 4.9b, and topology model Fig. 4.10b) contains three sequentially encoded TPR repeats (1, 2 and 3), and promotes protein-protein interactions. In *S. cerevisiae* this domain acts as the Sgt1: Sgt1 dimerisation

domain (Bansal *et al.* 2009a). Sgt1 dimerisation occurs through a perpendicular arrangement of monomers (Fig. 4.9a), modulated by a concave face arrangement of one monomer stacking against the convex face of a second. Crystal structure analysis has shown that this ordered arrangement of stacked Sgt1 monomers is stabilised by an Ascomycota[21] specific structured insert between TPR2A and B (amino acids 54-59) (Fig. 4.10b), critical for *S. cerevisiae* dimerisation. This interaction is stabilised by a pi stacking interaction between the W58 of one monomer, to the H59 of a second (Fig. 4.11a). These amino acids are orientated within the convex and concave faces respectively (Willhöft *et al. In Preparation*).

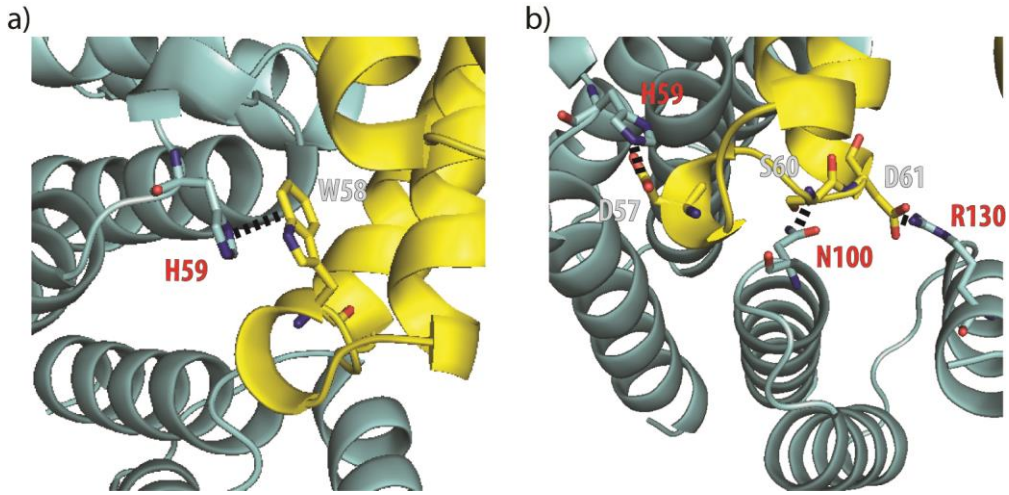

Figure 4.11. Sgt1 Stacking Interactions

a) *S. cerevisiae* Sgt1 dimerisation is stabilised by a pi stacking interaction between W58 and H59, of the convex (yellow, with grey text) and concave (blue, with red text) faces of adjacent TPR domains respectively. Both amino acids are encoded within the Ascomycota specific structured insert between TPR2A and B. b) This interaction is additionally stabilised by interactions between the convex face D57, S60, D61 amino acids, to their respective partners H59, N100, R130 in the concave face. Additionally, the D57-H59 interaction is proposed to stabilise the orientation of the latter within the binding face (Willhöft 2013). Structure currently unpublished in the PDB.

Stabilisation of *S. cerevisiae* Sgt1 dimerisation is additionally promoted by hydrogen bonding interactions between neighbouring monomers. These are formed between the convex face amino acids: D57, S60 and D61 (all located within the extended

[21] Taxonomic phylum within the kingdom Fungi.

loop region), and their binding partners within the concave face: H59, N100 and R130 (Fig. 4.11b) (Willhöft *et al. In Preparation*).[22]

The *H. vulgare* Sgt1 homologue contains three cysteine amino acids within the homo-oligomerisation domain. Two of these cysteines, C84 and C117, promote dimerisation through disulphide linkages. Formation of this bond has been suggested to function as an oxidative stress sensor during pathogenic attack (Nyarko *et al.* 2007). Concerning the *S. cerevisiae* homologue studied here, sequence analysis by Oliver Willhöft indicated the presence of a single cysteine residue (C47) orientated within the dimerisation domain (TPR2A). SV-AUC (sedimentation velocity, analytical ultracentrifugation), with and without DTT, has shown that this residue is not involved in dimerisation (Willhöft 2013).

4.1.4.2. The Sgt1 SGS and CS Domains

Sgt1 dimerisation is a regulated event, and inhibition of this interaction arrests the *S. cerevisiae* cell cycle at the G2/M interface by preventing kinetochore assembly. Observations by Bansal and co-workers (2009b) showed that the *S. cerevisiae* Sgt1 S361, a residue within the SGS domain, acts as a phosphorylation site for CK2. Phosphorylation of this residue has been proposed to modify the structure of the SGS domain, and the CS to SGS linker (amino acids 270-312). This phosphorylation subsequently inhibits Sgt1 dimerisation, arresting the cell cycle by abolishing the Skp1 interactions mediated by this oligomeric state (Bansal *et al.* 2009b). Therefore, this represents a dual role of CK2 regulated kinetochore formation, in addition to the previously outlined inhibition of Ndc10 function (section 4.1.1.2) (Peng *et al.* 2011). Furthermore this domain has been shown to interact with leucine rich repeat domains, such as that of the Hsp90 client protein Ctf13 (section 4.1.1.1), an early component of the CBF3 complex mediated by Sgt1: Skp1 interactions (Austin *et al.* 2002).

The CS domain is a highly conserved structural cross-species homologue, required to promote Sgt1: Hsp90 binding in both *S. cerevisiae* and Humans (Bansal, Abdulle & Kitagawa 2004, and Zhang *et al.* 2008).

[22] N100 and R130 are located within TPR3B and capping helices respectively.

4.1.5. Initiation of the CBF3 Complex

Formation of the CBF3 initiation complex occurs in a step wise manner. This step-wise interaction allows Hsp90 to promote Skp1: Ctf13 binding, through additional interactions with Sgt1 (Fig. 4.12) (Catlett & Kaplan 2006, Stemmann *et al.* 2002). This interaction is proposed to induce Skp1 mediated Ctf13 phosphorylation (Kaplan, Hyman & Sorger 1997) that subsequently promotes stable CBF3 formation (Fig. 4.2) (Russell, Grancell & Sorger 1999, and Yoshida, Murakami & Tanaka 2011).

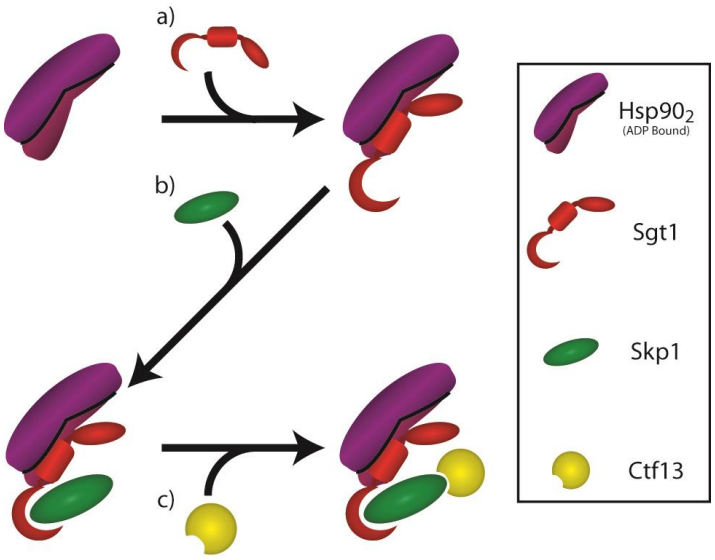

Figure 4.12. Sequential Interactions Mediating the Activation of Ctf13 by Hsp90

a) Sgt1 binds to Hsp90 in its ADP bound compact dimeric state (Zhang *et al.* 2008). b) This interaction promotes Sgt1: Skp1 binding (Bansal, Abdulle & Kitagawa 2004). c) Interactions between Sgt1 and Skp1 subsequently allows stable Skp1: Ctf13 binding through the latter's weakly conserved F-box domain (Russell, Grancell & Sorger 1999). This interaction allows Skp1 mediated Ctf13 phosphorylation, required for CBF3 complex formation (Kaplan, Hyman & Sorger 1997).

Hsp90 in its ADP bound form binds, to the CS domain of Sgt1, through its ATPase domain supporting a 1:1 interaction (Fig. 4.13) (Zhang *et al.* 2008, and Zhang *et al.* 2010). Formation of the Hsp90: Sgt1 complex promotes Skp1 binding supporting a binding promoter role for Sgt1 (Bansal, Abdulle & Kitagawa 2004, and Lingelbach & Kaplan 2004). Observations have shown however that Sgt1: Skp1 binding can occur in the absence of Hsp90, with a K_d of 6 μM (Willhöft *et al. In Preparation*).

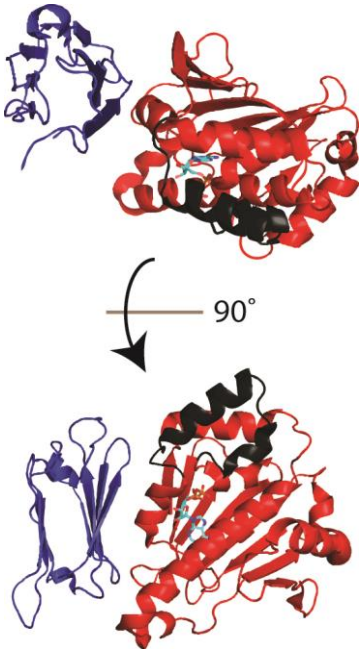

Figure 4.13. Crystal structure of the Hsp90 ATPase domain complexed with the Sgt1 CS domain

Crystal structure of the Hsp90 ATPase domain (red, lid-segment highlighted in black), ADP bound form, interacting with the CS domain of Sgt1 (blue) (PDB ID; 2JKI, resolved to 3.3 Å) (Zhang *et al.* 2008).

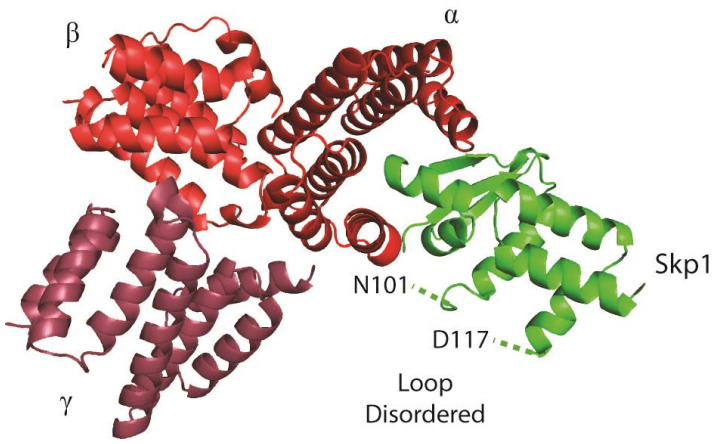

Figure 4.14. Crystal Structure of Sgt1 TPRκ$_3$Skp1 ΔBTB/POZ

The crystal structure of the Sgt1: Skp1 binding interface has been recently solved using Skp1ΔC and Sgt1 TPRκ (α, β and γ). Despite a Sgt1: Skp1 incubation ratio of 2:1, the resolved crystal produced an unexpected 3:1 oligomer (resolution 2.8 Å). Skp1 binds within the concave face of Sgt1 TPRκ α. β and γ stack sequential on top of α (Willhöft 2013, and Willhöft *et al. In Preparation*). The structure of Skp1 remains unresolved between amino acids 101-177, consistent with other crystal structures including 3KMS (Fig. 4.4b) (Willhöft *et al. In Preparation*). Structure currently unpublished in the PDB.

In *S. cerevisiae* the Hsp90: Skp1 interactions are Sgt1 dimer dependent. In contrast, human homologues require Sgt1 monomers to promote these interactions (Bansal *et al.* 2009a, and Zhang *et al.* 2008). The structure of the *S. cerevisiae* Sgt1: Skp1 binding

interface has recently been solved, using an Sgt1 N-terminal TPR domain only mutant (TPRκ, amino acids 1-150) and Skp1ΔC (Fig. 4.14). Sgt1: Skp1 binding progresses through the burial of the Skp1 N-terminus within the concave face of the Sgt1 TPR domain (Willhöft 2013, and Willhöft *et al. In Preparation*). This interaction is stabilised by the strongly conserved region of acidic amino acids within the Skp1 sequence (amino acids 65-73), and a region of basic amino acids orientated within the concave face of the Sgt1 TPR domain (Willhöft *et al. In Preparation*).

Formation of the Hsp90: Sgt1: Skp1 interaction promotes a 3.2x increase in Skp1: Ctf13 binding (Bansal, Abdulle & Kitagawa 2004). This interaction is stabilised by a weakly conserved F-box domain within the Ctf13 (N-terminal 58 amino acids), and the Skp1 C-terminus F-box interacting domain. This interaction confers an E3 ubiquitin ligase like interaction similar to Skp1's activity within the SCF complex (Russell, Grancell & Sorger 1999).

4.2. Experimental Aims and Objectives

The X-ray crystal structure of the Ascomycota specific loop (amino acids 54-59) within the *S. cerevisiae* Sgt1 dimerisation domain supports sequential H59-W58 pi-stacking interactions between adjacent Sgt1 monomers (Fig. 4.11a). In the absence of the C-terminus a deletion mutant encoding only the dimerisation domain, TPRκ (amino acids 1-150, TPR plus capping helix; κ), has been shown to enable Sgt1 trimerization (Fig. 4.9). This is in contrast to the native Sgt1 dimer required for *in vivo* activity. Modifying the protein concentration we have studied the ability of the Ascomycota specific loop to support higher order oligomers, both in the presence (wt) and absence (TPRκ) of the C-terminal domains.

Analysis of the X-ray crystallography structure has shown that dimerisation is stabilised by H-bonding interactions outside of the Ascomycota specific loop (Fig. 4.11b). These interactions are thought to be salt sensitive, modifying the oligomeric state at higher ionic strengths. Therefore, we have studied the influence of increasing ionic strengths to modify the oligomeric state of wt Sgt1 and the deletion mutant TPRκ.

Amino acid substitutions have been applied to further study the requirement of Ascomycota specific loop, and additional H-bonding interactions, to promote stable dimerisation behaviour. These studies have been performed using H59A, and D61R substitutions. The H59A substitution has been designed to study the role of the ascomycete specific loop in dimerisation, inhibiting its interaction with W58 (Fig. 4.11a). The D61R substitution has been designed to inhibit its interactions with R130 (Fig. 4.11b), allowing studies of the extended H-bonding interactions that are proposed to stabilise the dimerisation interface.

In *S. cerevisiae* Sgt1 dimers mediate Hsp90: Skp1 binding. Analysis of the crystal structure of the TPRκ_3Skp1 (Fig. 4.14) however, has shown that the interaction between the Sgt1 dimer and Skp1 is stabilised by a 1:1 interaction. MS/MS has been applied to study the organisation and stability of the Sgt1: Skp1 binding interface under dissociative conditions.

4.3. Materials and Methods

4.3.1. Molecular Biology, Preparation, and Purification of Proteins Studied

All Sgt1 and Skp1 samples were purified by the Vaughan research group. Sample preparation was performed using transformed *E. coli* (BL21) for all proteins, except TPRκ that utilised T7 Express LysY/IQ. Vector systems for transfection of genetic material encoding the required proteins are outlined in table 4.1.

Protein samples were purified using three to four chromatography steps (table 4.1). Samples were supplied in the following, final elution buffer: 25 mM HEPES, 150 mM NaCl, 2 mM DTT, and 1 mM EDTA at pH 7.0.

4.3.2. Mass Spectrometric Analysis of Sgt1

Prior to mass spectrometric analysis, samples were buffer exchanged using a single Micro Bio-Spin column (BioRad, Hertfordshire, UK), and three sequential washes using a single Amicon 1.5 ml 10 kDa MWCO centrifugal concentrator column (Millipore, Dundee, UK). All columns were prepared using the required experimental concentration of AmAc (pH 7).

Table 4.1. Biological Purification Methods for Sgt1 and Skp1 Proteins

Protein	Protein Name	Purification Tag	Expression System (Vector)	Purification Steps			
				1	2	3	4
Sgt1	wt	His-tag	pCOLADuet-1 (MCS 2)	His6-Affinity Pulldown (1 mL HisTrap FF)	IEX (1 mL Resource S)	SEC (HiLoad 16/60 S200)	-
	H59A		pET28a				
	D61R		pET28a				
	TPRκ	None	pRSETa	IEX (10 mL HiTrap SP FF)	SEC (HiLoad 16/60 S75)	IEX (5 mL HiTrap SP FF)	SEC (HiLoad 16/60 S200)
Skp1	wt	None	pET22b	IEX (10 mL HiTrap SP FF)	SEC (HiLoad 16/60 S75)	IEX (5 mL HiTrap SP FF)	SEC (HiLoad 16/60 S200)

IEX; Ion Exchange Chromatography. All columns purchased from GE Healthcare Life Sciences, Buckinghamshire, UK.

Adapted, with permission, from Willhöft 2013

Two methods were applied to measure sample concentrations. The low mass proteins (native oligomer less than 69 kDa) Skp1 and TPRκ, were measured using a

Qubit 2.0 fluorometer (Invitrogen, Paisley, UK), and a 1 in 100 dilution of sample: running buffer. The high mass proteins (native oligomer greater than 69 kDa) wt Sgt1, H59A, and D61R (including incubated wt: Skp1 samples post purification), were measured using serial dilution Nanodrop 1000 (Thermo Scientific, DE USA) analysis. Extinction coefficients were applied to correct Nanodrop 1000 mg/ml values (table 4.2). Using the corrected mg/ml values, sample molarity was calculated using the experimentally calculated mass (from denatured protein mass analysis) (table 4.2).

These differences in concentration analysis occur due to identified accuracy limitations of the Qubit method to measure proteins with masses greater than the bovine serum albumin (BSA) standard used for calibration (Jones, Haugland & Singer 2003).[23] In all cases the recovered sample was diluted to the experimental concentration using the required ionic strength of AmAc buffer (pH 7).

4.3.2.1. Mass Spectrometric Analysis of Sgt1: Area Under Curve Analysis

The following conditions are relevant to the work performed in sections 4.4.2 and 4.4.3. Analysis of oligomer contributions to each spectrum were performed using the recently developed Amphitrite software (Sivalingam *et al.* 2013). This software deconvolutes individual mass spectra using Gaussian curve fitting to calculate component contributions to mass spectra as a normalised percentage. This software is therefore able to calculate the area under curve (AUC) contributions of oligomeric species normalised to the observed dominant ion peak.

Amphitrite deconvolution of experimental spectra was performed on individually-imported files with a grain value of 500,[24] reducing the computational demands of the analysis. Additional light smoothing of mass spectra was applied within Amphitrite to improve the automatic peak detection. Manual selection of CSD peaks was performed to minimise calculated mass errors. Individual spectra were sequentially simulated with a variety of 'single peak FWHM' values to maximise deconvolution of all charge peaks

[23] These observations confirmed using Nanodrop 1000 vs. Qubit sample concentration measurements, on the TPRκ and FL proteins. Data not shown.
[24] Grain represents spacing between imported values, 1 = 0.0116 *m/z*. Grain of 500 therefore samples spectra points every 5.8 *m/z*.

within the distribution. After successful simulation of oligomer CSD's, the respective AUC values were copied from the graphical user interface for analysis using Microsoft Office Excel 2003 (Microsoft, WA, USA).

Prior to mass spectrometric analysis, all homogenous samples of wt Sgt1, TPRκ, H59A and D61R were incubated at 4 ˚C for 60 minutes. Prior to incubation, all samples were diluted to a 20 µL final volume at the required experimental protein concentration and ionic strength, acquiring control spectra at time zero. All spectra and results presented here represent data acquired post incubation.

Wt and TPRκ samples were studied over a range of ionic strength and protein concentrations. Samples were incubated in the following AmAc concentrations, at pH 7: 150, 500, and 1000 mM. At each ionic strength the following protein concentrations were analysed: 4.5, 11.2, and 22.3 µM. At 150 mM AmAc, wt Sgt1 was additionally analysed at the following protein concentrations: 7.9, 14.9, and 18.6 µM. H59A and D61R were analysed at a concentration of 11.2 µM, incubated in 150 mM AmAc (pH 7). All data sets are the mean of five repeats, $^{+}/_{-}$ MSD.

Instrumental values for mass spectrometric analysis were: capillary voltage; 0.9 to 1.5 kV, Nanoflow pressure; 0.00 to 0.15 Bar, Sampling cone; 80 V, Extraction cone; 1 V, Source temperature; 40 ˚C, Trap collision; 5 V, Transfer collision; 10 V, Trap DC bias; 8 V, Trap Gas; 4 ml/min. Spectra were typically acquired over two minutes using two second scans, unless longer was required to improve the signal to noise value (between 5 to 10 minutes). Instrumental pressures: Trap; 0.02 mBar, mobility separator; 0.002 mBar, ToF; 1.5×10^{-6} mBar. Wt, H59A and D61R mass spectra were acquired over an *m/z* range of 1 000 to 16 000 with a source backing pressure of 5 mBar. TPRκ mass spectra were acquired over an *m/z* range of 500 to 8 000 with a source backing pressure of 4.5 mBar.

4.3.2.2. Mass Spectrometric Analysis of Sgt1: Ion Mobility

The following conditions are relevant to the work performed in section 4.4.1.1. Ion mobility data have been acquired for the lowest observed charge of the TPRκ trimer ion

(+13, *m/z* 4000.65) at a concentration of 11.2 μM. Analysis was performed using an AmAc buffer concentration of 150 mM, pH 7.

Calibration of the experimental t_D values, to produce relevant CCS values, was performed using the method outlined in section 2.2. Denatured Equine myoglobin (Sigma-Aldrich Ltd, Dorset, UK) was used as the CCS calibrant. All data sets are the mean of five repeats, $^{+}/_{-}$ MSD, acquired at a range of mobility separation conditions: 250 m/s 7 V, 275 m/s 7.5 V, and 300 m/s 8 V.

Due to the requirement of increased operating pressures for IM separation, the following values were used: Trap DC bias; 20 V, Trap Gas; 5 ml/min, IM separator gas flow; 20 ml/min. Instrumental pressures: Trap; 0.04 mBar, mobility separator; 0.52 mBar, ToF; $2x10^{-6}$ mBar. All other instrumental parameters are consistent with those detailed in section 4.3.2.1.

4.3.2.3. Mass Spectrometric Analysis of Sgt1: Tandem Mass Spectrometry Analysis of Skp1 Binding

The following conditions are relevant to the work performed in section 4.4.4. Unless stated samples of wt Sgt1 or TPRκ were mixed with Skp1 in 2:1. Samples were mixed to produce a theoretical final $Sgt1_2Skp1$, or $TPR\kappa_2Skp1$, concentration of 30 μM, calculated using the minimum Amicon 1.5 ml 10 kDa MWCO centrifugal concentrator column (Millipore, Dundee, UK) recovery volume of 50 μL. Sample concentration measurements were performed according to the methods previously outlined above (section 4.3.2).

Samples of Sgt1: Skp1 were suspended in 400 μL incubation buffer (25 mM HEPES, 150 mM NaCl, 2 mM DTT, 1 mM EDTA, pH 7.0) and sequentially inverted using a SRT6 roller mixer (Stuart, Staffordshire, UK), at a machine defined 33 rpm, for 60 minutes at 4 ˚C. Prior to buffer exchange samples were concentrated to the minimum recovery volume of the Amicon centrifugal concentrator columns, removing the excess incubation buffer.[25] Prior to recovery, samples were buffer exchanged was

[25] Micro Bio-Spin 6 columns (BioRad, Hertfordshire. UK) were not used for buffer exchange protocols due to sample volumes exceeding the recommended load volume greater than six fold.

performed by three sequential washes into 150 mM AmAc (pH 7) using a single Amicon 1.5 ml 10 kDa MWCO centrifugal concentrator columns (Millipore, Dundee, UK).

Quadrupole mass isolation was used to select the *m/z* corresponding to the $Sgt1_2Skp1$ +23 ion (*m/z* 4856) prior to CID. The TriWave Trap (see schematic in Fig. 1.13) voltage was ramped from 10 to 120 V, in increments of 10 V. Analysis of the $TPR\kappa_2Skp1$ complex was performed by isolating the *m/z* value corresponding to the +15 ion (*m/z* 3791). The TriWave Trap voltage was ramped from 10 to 80 V, in increments of 5 V. Analysis of both complexes was performed with a static TriWave Transfer voltage of 10 V.

Protein independent instrumental settings for mass spectrometric analysis inconsistent with the previously stated values in section 4.3.2.1 were: Trap DC bias; 20 V, Trap Gas; 2 ml/min. Spectra were typically acquired over two minutes, unless longer was required to improve signal to noise, using two second scans. Instrumental pressures: Trap; 0.015 mBar, mobility separator; 0.002 mBar, ToF; $1.5x10^{-6}$ mBar. The *m/z* range for analysis was 1 000 to 25 000.

Protein variable values, excluding collision voltages, are as follows, wt_2Skp1: LM and HM values; 4 and 15 respectively, source backing; 5 mBar. $TPR\kappa_2Skp1$: LM and HM values; 4 and 19 respectively, source backing; 4.5 mBar.

4.4. Results

4.4.1. Control Mass Analysis of the Sgt1 and Skp1 Proteins Studied

As a prerequisite for mass analysis of all Sgt1 and Skp1 proteins studied here, denatured protein mass analysis was performed to calculate accurate experimental mass values using the buffer exchange method outlined in section 2.3. Experimentally calculated masses, along with mass difference comparisons with the theoretical are described in table 4.2.

The observed Sgt1 and Skp1 mass differences are proposed to occur because of N-terminal methionine clipping (M1), explaining an average loss of ~130 Da. This is supported by denatured protein mass analysis of selenomethionine labelled TPRκ, required for crystallography phase determination (Fig. 4.15b).[26] Analysis of selenomethionine

[26] Crystallography work performed by Oliver Willhöft.

labelled TPRκ showed two successful methionine to selenomethionine substitutions, in lieu of the expected three encoded by the amino acid sequence. In contrast, if C-terminal clipping were occurring we would expect a mass difference of 101 Da (Threonine) for TPRκ, 156 Da (Arginine) for Skp1, and 137 Da (Histidine from His-tag) for the wt, H59A, and D61R proteins.

Table 4.2. Theoretical and Experimentally Calculated Masses of Sgt1 and Skp1 Proteins

Protein	Protein Name	Tag	Mol. E, Coeff. (M^{-1} cm^{-1})	Theoretical Mass (Da)	Experimental Mass (Da)	Mass Error ($^{+}/_{-}$ Da)	Mass Difference (Expt. - Theo.)
Sgt1	wt	His-tag	64 860	47 023	46 895.88	6.23	-127.12
	H59A		59 360	46 957	46 832.41	4.92	-124.59
	D61R		59 360	47 064	46 931.63	6.51	-132.37
	TPRκ	None	22 920	17 462	17 331.83	3.23	-130.17
Skp1	wt	None	20 970	22 330	22 200.25	4.04	-129.75

Denatured mass analysis of H59A and D61R indicated a large number of additional masses. Analysis of the H59A (Fig. 4.16b) denatured spectra indicates species with and without N-terminus methionine clipping. In contrast, D61R (Fig. 4.16c) denatured spectra indicates a variety of mass values, including those with and without methionine clipping and masses corresponding to increases over theoretical values. The causes of these additional higher mass distributions currently remain unidentified, and have been observed in a number of different sample expression and purification runs obtained from our collaborators.

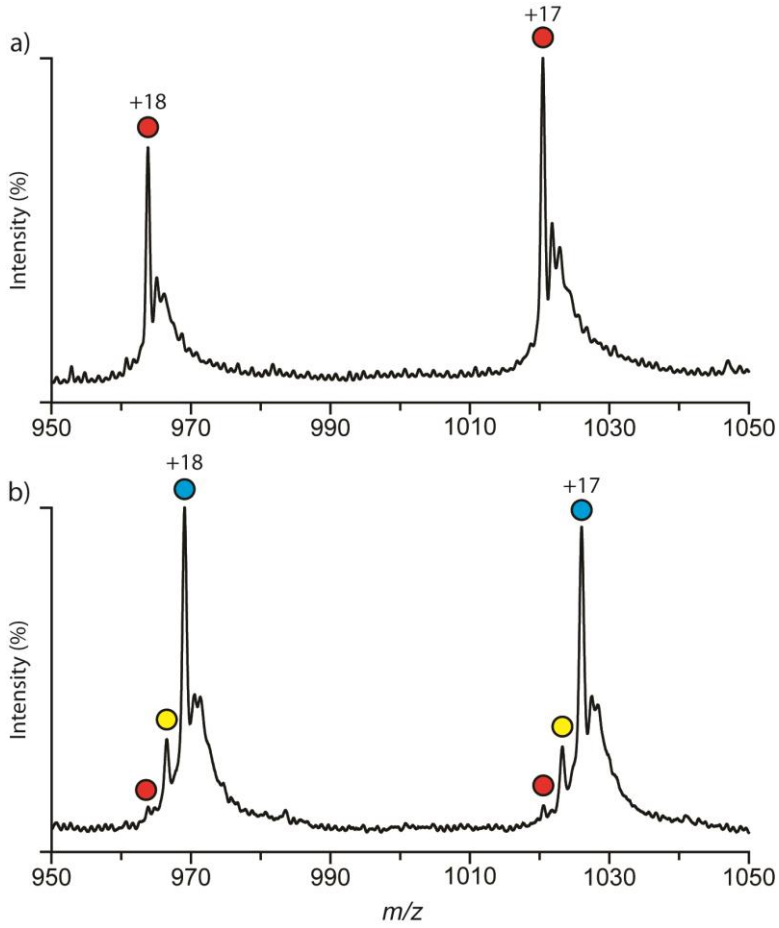

Figure 4.15. Denatured Mass Spectra of TPRκ and Methionine to Selenomethionine Substituted TPRκ

a) Denatured protein mass analysis of TPRκ prior to Methionine to Selenomethionine substitutions, indicated a mass difference of 130.17 ($^{+}/_{-}$ 3.23) Da compared to the theoretical (red). This mass loss was proposed to occur due to N-terminal methionine clipping.

b) Denatured protein mass analysis to calculate the number of Methionine to Selenomethionine substitutions. Analysis was performed to aid the calculation of the crystallography phase for the TPRκ oligomer structure solved by Oliver Willhöft (Fig. 4.9a). Of the three methionines encoded by the TPRκ (M1, M64 and M76), mass analysis shows up to two observed substitutions (red = 0, yellow =1, blue =2) are successful, with no evidence of a third. This therefore supports that the observed average mass loss of 132 Da, occurs due to non-specific cleavage of M1 during the expression and purification. Mass difference between methionine and selenomethionine is 46.9 Da.

Accounting for N-terminal methionine clipping occurring in samples studied the theoretical mass was updated to reflect these observations. Comparison between the theoretical masses lacking methionine and the experimental are presented in table 4.3, and represent good agreement between the two values.

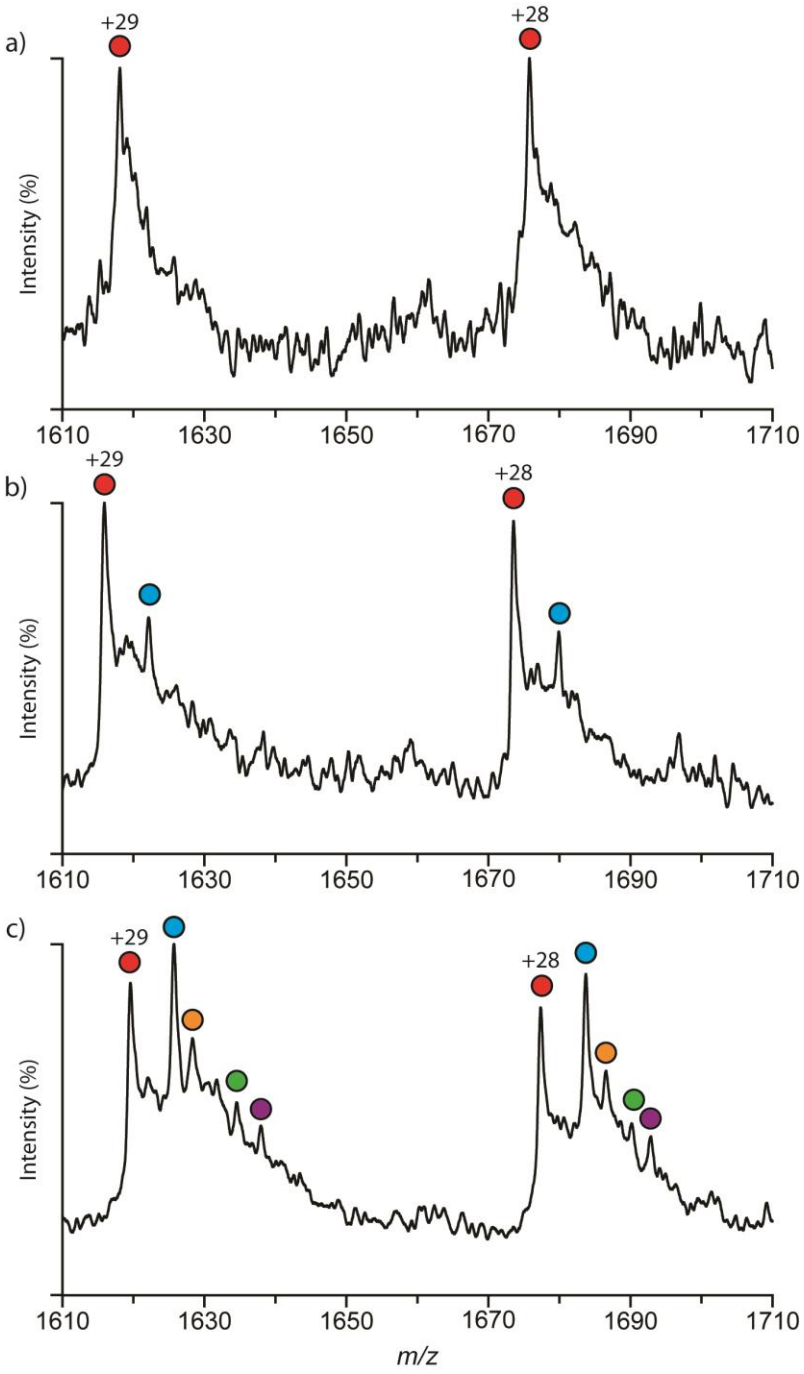

Figure 4.16. Denatured Mass Spectra of the Sgt1 Wild Type, H59A & D61R Proteins

a) Denatured protein mass analysis of wt Sgt1 indicates methionine (M1) clipping (Red). This clipping is additionally observed in (b) H59A and (c) D61R, with masses corresponding to the theoretical native (blue). Of concern are the observed masses in the D61R spectra that correspond to gains in mass. These represent masses of 130 (orange), 1929 (green) and 3686 (purple) Da greater than the D61R theoretical.

Table 4.3. Comparison of Theoretical and Experimental Mass Differences of Sgt1 and Skp1 Proteins Accounting for Methionine Clipping

Protein	Protein Name	Theoretical Mass (Da) Accounting for Methionine Clipping	Experimental Mass (Da)	Mass Error ($^{+}/_{-}$ Da)	Mass Difference (Expt. - Theo.)
Sgt1	wt	46 891	46 895.88	6.23	+4.88
	H59A	46 825	46 832.41	4.92	+7.41
	D61R	46 932	46 931.63	6.51	-0.37
	TPRκ	17 330	17 331.83	3.23	+1.83
Skp1	wt	22 198	22 200.25	4.04	+2.25

Removal of N-terminal Methionine represents a subtraction of 132 Da from the theoretical mass.

4.4.1.1. MobCal Analysis of the TPRκ Trimer Native Structural Conformation

This investigation has drawn on a key advantage provided by the Synapt HDMS (Waters, Manchester, UK) over other commercial ToF mass spectrometers. Using the inbuilt T-wave mobility separator, we are able to confirm or refute the biological relevance of the mass spectrometry data sets by comparing IM calibrated CCS, with MobCal analysis of available crystal structure data.

IM-MS was used to calculate the CCS of TPRκ lowest observed charge +13 ion (*m/z* 4000.65). The t_D extracted from the FWHM of this ion exhibits two structural conformations (1 and 2, Fig. 4.17d). The calibrated CCS of this ion equates to 3635.6 ($^{+}/_{-}$ 76.91) Å^2 for the primary compact, and 3886.36 ($^{+}/_{-}$ 84.81) Å^2 for the low intensity secondary conformation. This secondary conformation represents an increase of 6.9 % over the primary compact.

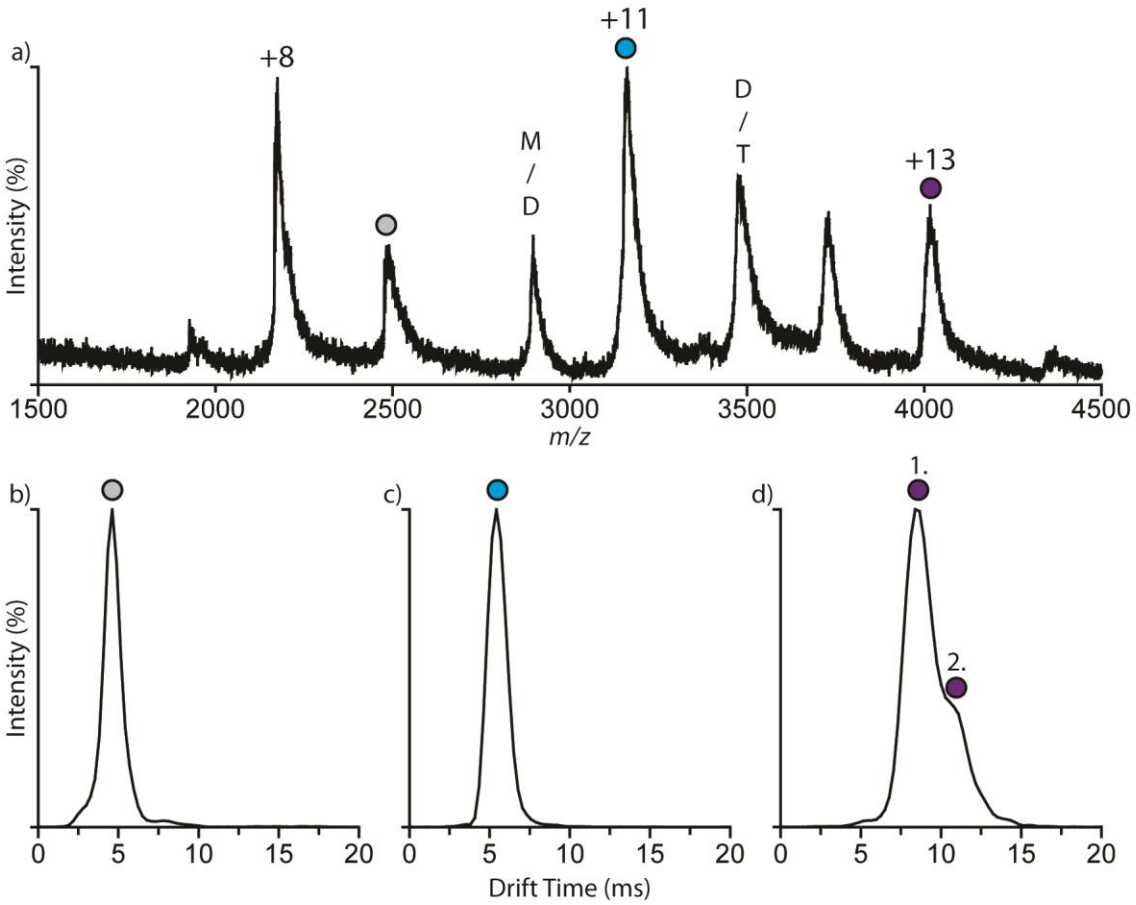

Figure 4.17. Ion Mobility Extracted Drift Time Data of the TPR Oligomers

a) Mass spectra recreated from the sum of the TPRκ mobility data acquired at a wave velocity of 300 m/s and an amplitude of 7 V. Sample concentration 11.2 μM, in a 150 mM AmAc buffer (pH 7). b) t_D data recreated from the FWHM of the monomer +7 ion (*m/z* 2476.9, grey) shows a single compact structural conformation (1580.0 $^{+}/_{-}$ 37.57 Å^2). c) Like the monomer, the dimer +11 ion (*m/z* 3152.2, blue) shows a single compact conformation (2619.2 $^{+}/_{-}$ 79.98 Å^2). d) t_D data recreated from the FWHM of the trimer +13 ion (*m/z* 4000.65, purple) indicates two conformations, an abundant compact (1) conformation (3635.6 $^{+}/_{-}$ 76.91 Å^2), and less abundant extended (2) conformation (3886.36 $^{+}/_{-}$ 84.81 Å^2). Due to the strong overlap of TPRκ oligomer (monomer, dimer & trimer) CSDs studies of the lowest charged monomer (M/D; = +6 monomer, overlapping with +12 dimer) and dimer (D/T; +10 dimer, overlapping with +15 trimer) ions could not be performed. These ions therefore represent the lowest charged *m/z* value not shared with either other oligomeric species, ensuring no additional t_D contributions used to calculate the CCS of the respective ions studied.

MobCal was used to apply the PA and EHSS to calculate the CCS of for the TPRκ$_3$ crystal (Fig. 4.9a) across thirty simulations. The PA calculated CCS equalled 3057.45 ($^{+}/_{-}$ 30.21) Å^2, with the EHSS CCS equalling 3924.68 ($^{+}/_{-}$ 44.48) Å^2. Comparison

of these MobCal data with the primary dominant conformation (3635.6 $^{+}/_{-}$ 76.91 Å^2), suggests good agreement between both the PA and EHSS. The closest agreement between the gas phase and MobCal calculated CCS data, is the EHSS. Experimentally this model produces the most accurate results, when compared to the PA (Scarff *et al.* 2008). These data confirm that the conditions used are able to maintain the native-like structure of the TPRκ trimer, and are therefore capable of producing biologically valid results.

The existence of a second conformation for the trimer is contrary to the monomer and dimer, which both exhibit only a single compact conformation (Fig. 4.17b and c, respectively). The less abundant secondary, more extended, conformation may however represent a further biologically relevant structure, the significance of which will be discussed later (section 4.5.1).

4.4.2.1. Comparison of Sgt1 Wild Type and TPRκ Oligomerisation as a Function of Concentration

The recently solved crystal structure of TPRκ (Fig. 4.9a) has shown that Sgt1 dimerisation is promoted by a W58-H59 stacking interaction between adjacent monomers (Fig. 4.4a). In the absence of the C-terminal steric clashes, that inhibit higher order-order oligomerisation (e.g. trimer, tetramer), these amino acids are orientated to enable sequential stacking of Sgt1 monomers. Therefore, we have applied mass spectrometry to study the effect of concentration on the oligomerisation behaviour of TPRκ to promote higher order oligomers, comparing these with wt analysis.

Samples of wt and TPRκ were incubated in 150 mM AmAc (pH 7), to compare oligomerisation as a function of concentration at 4.5, 11.2 and 22.3 μM (Fig. 4.18). Wt data was expanded to study oligomerisation at 7.9, 14.9, and 18.6 μM. These conditions were incorporated after initial analysis, using the sum of oligomer normalised peak intensities, produced an inconclusive relationship between the concentration and oligomeric state. This peak intensity analysis predated the Amphitrite software. Comparison of all concentration points using Amphitrite calculated AUC analysis is shown in figure 4.19.

In the absence of the Sgt1 C-terminal 151-395 amino acids, analysis indicates that the TPRκ is able to undergo concentration dependent oligomerisation. This is further

supported by the appearance of higher order Sgt1 oligomers (tetramer and pentamer), not previously observed. These data therefore indicate that the orientations of W58 and H59 allow higher order oligomerisation in the absence of C-terminal domain steric clashes.

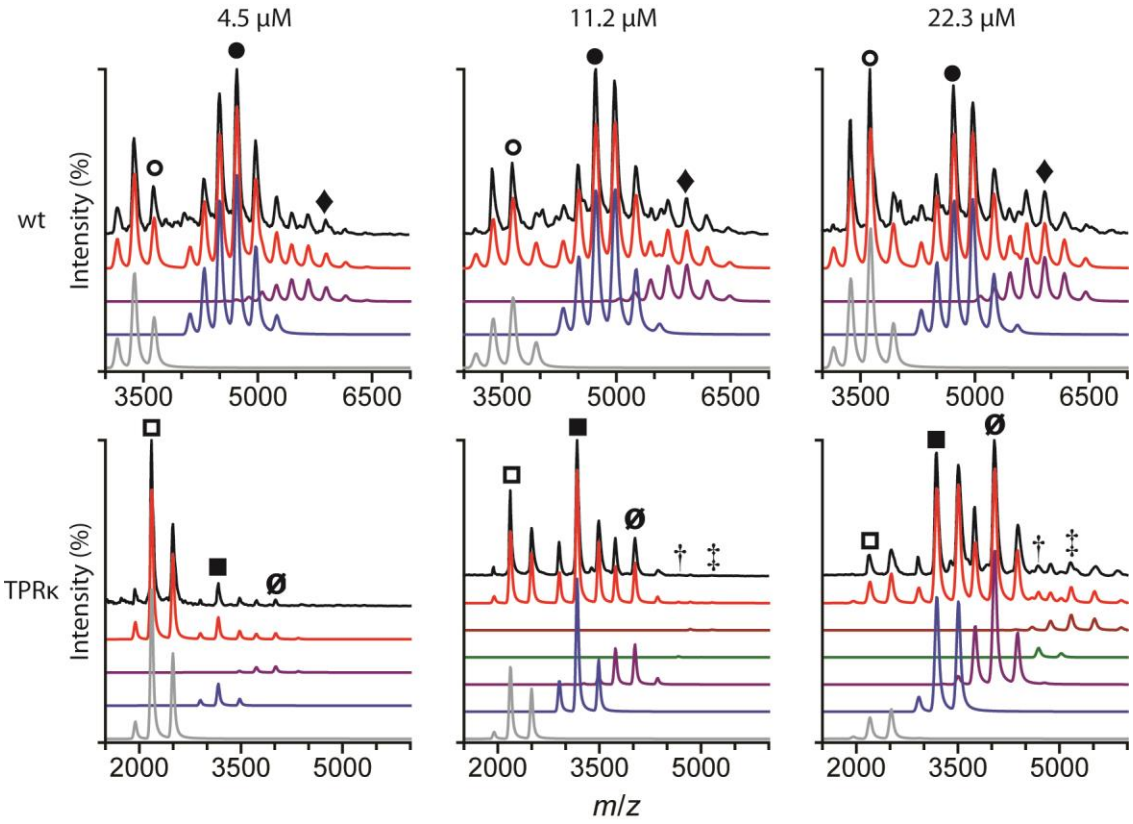

Figure 4.18. Comparison of Sgt1 Wild Type and TPRκ Oligomerisation as a Function of Concentration

Amphitrite analysis of the simulated oligomer species to calculate AUC values for the wt (top row) and TPRκ (bottom row) oligomers as a function of the concentration. Data is expanded in quantitative detail in figure 4.19. Spectra indicate the oligomerisation state of each protein at the represented concentrations, and highlight a positive correlation between concentration and higher order oligomers in the absence of the C-terminal amino acids 151-395.

Experimentally acquired spectra; black, simulated spectra; red, monomer; grey, dimer; blue, trimer; purple, tetramer; green, pentamer; brown. Charge states: wt +13; ○, wt_2 +20; ●, wt_3 +24; ♦, TPRκ +8; □, $TPR\kappa_2$ +11; ■, $TPR\kappa_3$ +13; ø, $TPR\kappa_4$ +15; †, $TPR\kappa_5$ +17; ‡

Wt AUC decreases only fractionally outside of MSD across the low to high concentration range (Fig. 4.19b). These decreases in wt dimer AUC, from low to high concentrations, suggests that the CS and SGS domain clashes may not be able to completely prevent concentration dependent oligomerisation. This observation is supported

by an increase in trimer AUC between 4.5 and 18.6 μM. At 22.3 μM, however, we observe a sudden unexplained change in oligomer contributions; an increase in monomer is observed with decreased levels of trimer inconsistent with trends in oligomerisation.

Despite a small reduction in AUC, in both figures (Fig. 4.18 and 4.19), the wt Sgt1 protein exhibits a consistent and relatively stable preference for the dimer, with an average AUC of 59.6 ($^{+}$/- 7.5) % across all protein concentrations. These results may therefore indicate that the CS and SGS domains act to promote oligomer stability. Excluding the SGS domain acting as a phosphorylation site for CK2, that negatively regulates dimerisation, contributions of the CS and SGS domains to promote oligomer stability have not been previously identified.

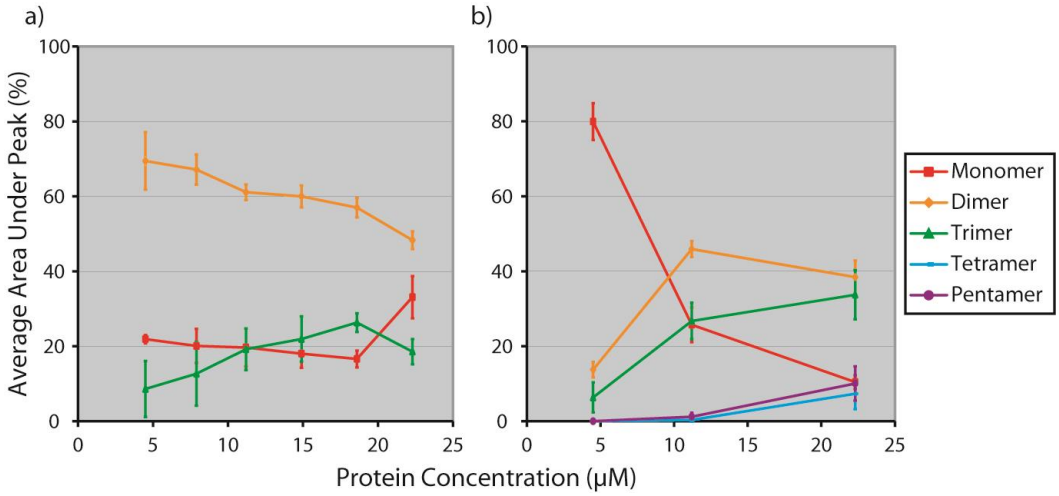

Figure 4.19. Percentage Area Under Peak Comparison of Sgt1 Wild Type and TPRκ as a Function of Protein Concentration

Comparison of average AUC Amphitrite calculated data, for the (a) wt and (b) TPRκ oligomers, as a function of protein concentration.

4.4.2.2. Comparison of Sgt1 Wild Type and TPRκ Oligomerisation as a Function of Ionic Strength

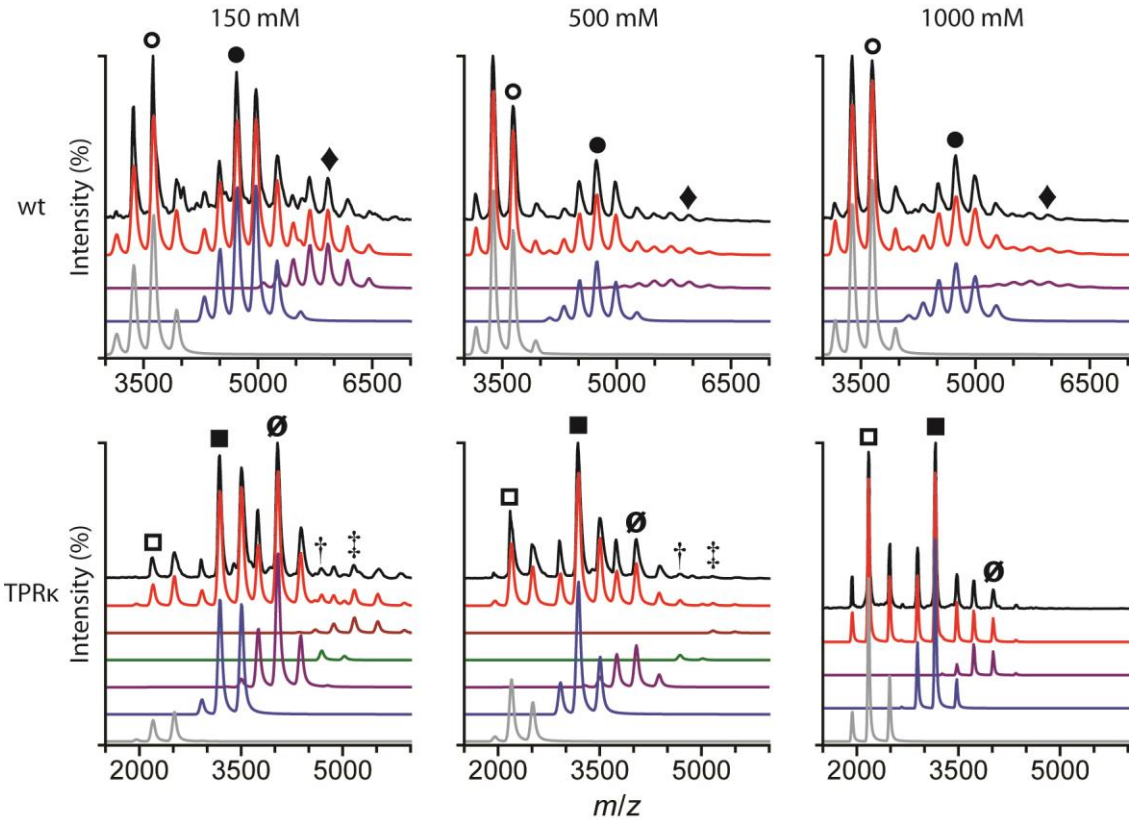

Figure 4.20. Comparison of Sgt1 Wild Type and TPRκ Oligomerisation as a Function of Ionic Strength
Amphitrite analysis of the simulated oligomer species, to derive AUC data for wt (top row) and TPRκ (bottom row) oligomers as a function of the ionic strength. Data is expanded in quantitative detail in figure 4.21. Protein concentration remains static in all spectra: 22.3 μM.
Experimentally acquired spectra; black, simulated spectra; red, monomer; grey, dimer; blue, trimer; purple, tetramer; green, pentamer; brown. Charge states: wt +13; ○, wt_2 +20; ●, wt_3 +24; ♦, TPRκ +8; □, $TPR\kappa_2$ +11; ■, $TPR\kappa_3$ +13; ø, $TPR\kappa_4$ +15; †, $TPR\kappa_5$ +17; ‡

Analysis by Oliver Willhöft, using SV-AUC had shown that the extended electrostatic interactions between D57 and D61 (convex face) and their respective binding partners H59 and R130 (concave face) required to stabilise the dimerisation interface, are perturbed at increasing ionic strengths. The former of these, the D57-H59 interaction, is required to orientate the H59 within the concave face to bind the W58 of the neighbouring Sgt1 monomer (Willhöft 2013). Mass spectrometry has therefore been applied to accurately

resolve the effect of increasing ionic strengths on Sgt1 oligomerisation behaviour. Studies were performed using AmAc, in lieu of the NaCl buffer used in previous SV-AUC experiments (Willhöft 2013), due to the non-volatile nature of the latter buffer.[27]

Samples of wt and TPRκ were incubated at a protein concentration of 22.3 μM, in 150, 500, and 1000 mM AmAc. Comparison of wt and TPRκ spectra across the ionic strength range studied is shown in figure 4.20. These data are expanded using the Amphitrite calculated AUC analysis in figure 4.21.

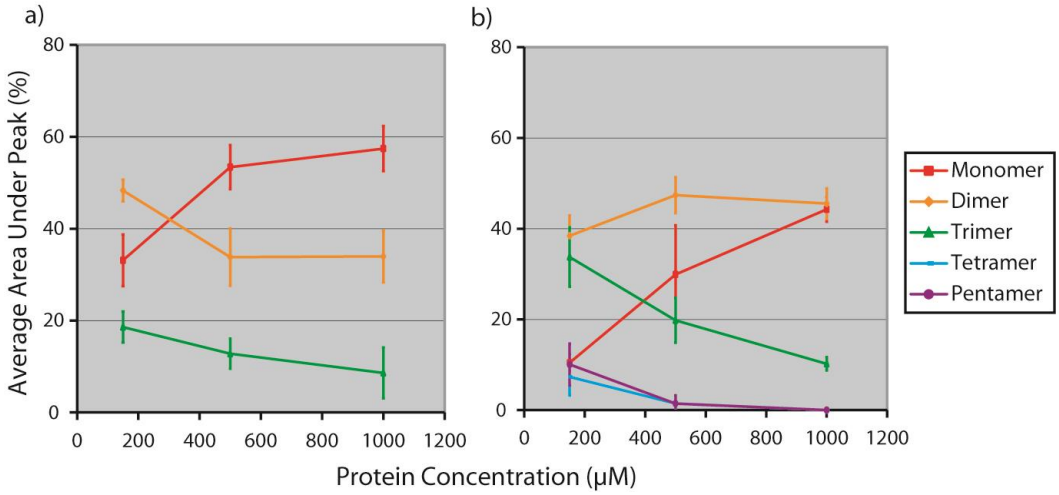

Figure 4.21. Percentage Area Under Peak Comparison of Sgt1 Wild Type and TPRκ as a Function of Ionic Strength

Comparison of Amphitrite calculated average AUC data, for (a) wt and (b) TPRκ oligomers, as a function of ionic strength.

The analysis presented here indicates that ionic strength is able to destabilise the binding interface of Sgt1. Reduced levels of higher order oligomers are observed for both wt and TPRκ as ionic strength increases, and is supported by a consistent increase in monomeric species. The largest decrease in higher order oligomers, in both the wt and TPRκ, occurs between 150 and 500 mM. Above these ionic strengths, marginal decreases in the higher order wt oligomers are observed. In contrast, at the highest ionic strength studied (1000 mM AmAc) not all TPRκ higher order oligomers have been destabilised.

[27] PubChem (http://pubchem.ncbi.nlm.nih.gov/, access date 7th August 2013) Property information; NaCl; CID 5234, 58.44277g/mol, H-Bond Donors 0, H-Bond Acceptors 1 (Net; +1 Acceptor) Ammonium Acetate; CID 517165, 77.08248g/mol, H-Bond Donors 1, H-Bond Acceptors 2 (Net; +1 Acceptor)

This suggests that higher ionic strengths may be required to promote similar TPRκ oligomer levels when compared with the wt data.

4.4.2.3. Comparison of Sgt1 Wild Type and TPRκ Oligomerisation as a Function of Concentration and Ionic Strength

In order to expand on the studies above, oligomerisation as both a function of ionic strength, and protein concentration were compared (sections 4.4.2.1 and 4.4.2.2). Wt and TPRκ were incubated in both 500 and 1000 mM AmAc, to compare oligomerisation at 4.5, 11.2 and 22.3 μM (for 150 mM data please see Fig. 4.19).

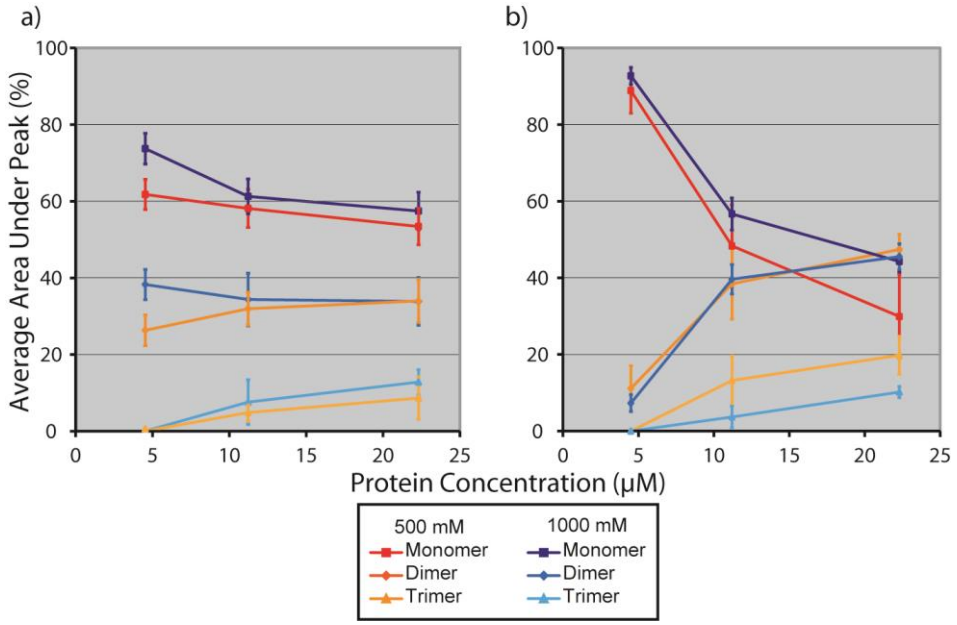

Figure 4.22. Extended Comparison: Percentage Area Under Peak Comparison of Sgt1 Wild Type and TPRκ

Comparison of average AUC calculated using Amphitrite for the oligomerisation state of the (a) wt and (b) TPRκ, as a function of ionic strength and protein concentration. For simplicity, only common oligomeric species present at all ionic strengths are shown (monomer, dimer, and trimer). Contributions from other TPRκ oligomers species represent a contribution of 2.9 % (Tetramer + Pentamer) at 22.3 μM, 500 mM AmAc. Data here represents an extension of the data presented in figure 4.19.

Compared with previous data (Fig. 4.21), results suggest that destabilisation of the salt sensitive binding interface interactions occur prior to the 500 mM AmAc concentration

(Fig. 4.22). This is supported by a minimal changes in oligomer AUC within MSD between 500 and 1000 mM AmAc, observed using both the wt and TPRκ at all protein concentrations. The remaining dimerisation interactions therefore appear to stabilise the oligomeric state, albeit at reduced levels, between both ionic strengths presented.

4.4.3. Disrupting the Binding Interface Using H59A and D61R Substitutions

Histidine 59 (H59) is a positively charged amino acid located within the extended loop region located between TPR2A and TPR2B (amino acids 54-59). This forms a stacking interaction with Tryptophan 58 (W58) in the extended TPR loop region of a second Sgt1 monomer (Fig. 4.11a) (Willhöft *et al. In Preparation*). This interaction is thought to central to the homodimerisation interaction, and it is hypothesised that a histidine to alanine substitution (H59A) should significantly reduce oligomer stability.

In addition to the W58-H59 stacking interface, further polar interactions aid stabilisation of the Sgt1 homodimerisation interface. Of interest here is the negatively charged amino acid Aspartic acid 61 (D61) located within the TPR2B helix, which forms an electrostatic interaction with the positively charged Arginine 130 (R130) of its neighbouring monomer (Fig. 4.11b). As this binding interface is supported by additional electrostatic interactions: D57 to S60, and H59 to N100 (convex to concave respectively), a D61R substitution will not completely disrupt oligomerisation, but hypothetically reduce its stability.

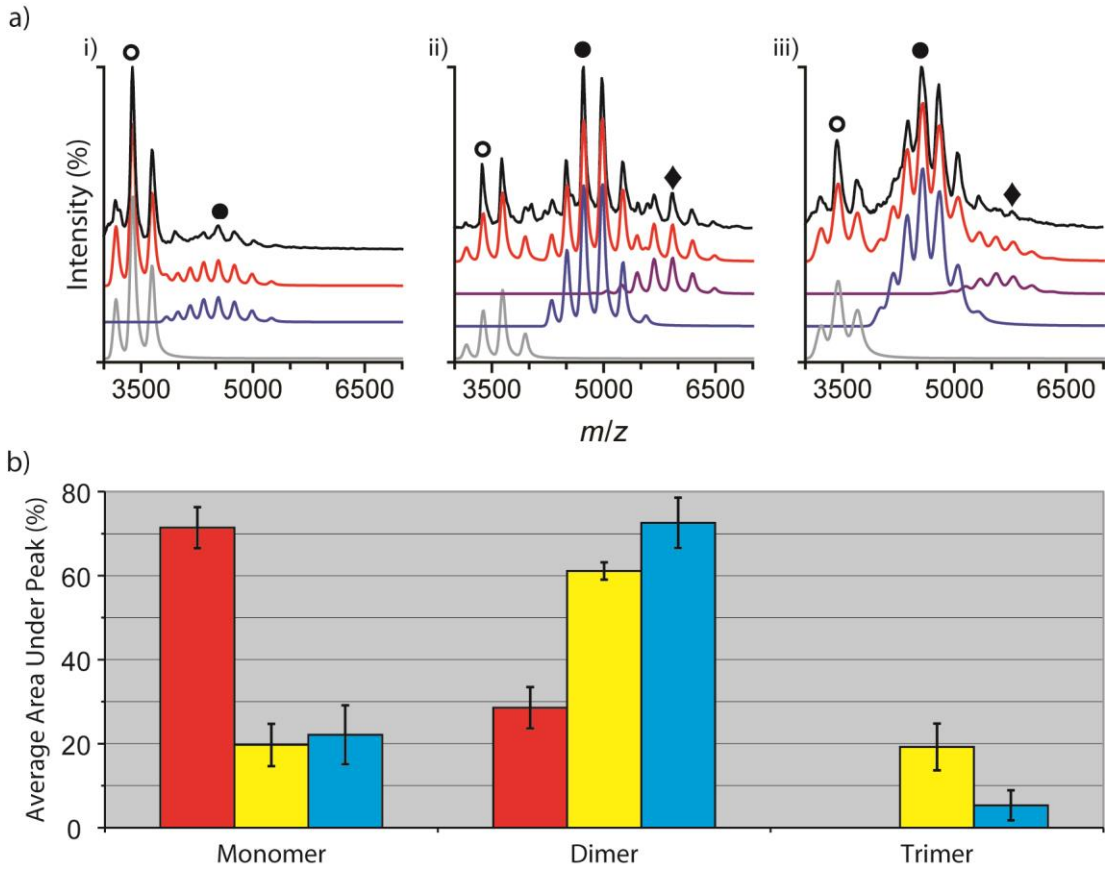

Figure 4.23. Analysis and Comparison of the Sgt1 Binding Interface Substitutions

a) Deconvolution of (i) H59A, (ii) wt, and (ii) D61R oligomeric species using Amphitrite. Experimental acquired spectra; black, simulated spectra; red, monomer; grey, dimer; blue, trimer; purple. Charge states: monomer +14; ○, dimer +20; ●, trimer +24; ♦. b) Comparison of H59A (red), and D61R (blue), oligomer AUC values (wt in yellow).

The mass heterogeneity observed using denatured mass analysis (previously described section 4.4.1 and Fig. 4.16) is proposed to account for the poor D61R dimer peak separation within acquired mass spectra.

Work presented here compares the two previously highlighted substitutions, against the wt protein, using the same solution and gas phase conditions (11.2 μM, 150 mM AmAc, pH 7). Comparisons of protein oligomeric species are shown in figure 4.23.

The H59A substitution inhibits higher order oligomerisation from 80 % in the wt (dimer + trimer) to almost a third of this value (28 %). These results represent a good agreement with the above hypothesis, supporting a central role of the H59-W58 stacking interaction to stabilise Sgt1 dimerisation.

In contrast, compared to the wt the D61R substitution appears, within MSD, to increase dimerisation events (61 to 72 %) with both proteins exhibiting comparable levels of monomer. Comparisons of trimerization oligomerisation events, however, indicate that this substitution destabilises further oligomerisation events, nearly negating the trimeric species within MSD (5.3 $^{+}$/_ 3.57 %). This is in direct contrast to the wt trimer species, which comprises nearly a fifth of all oligomers. This comparison therefore supports the above hypothesis that this D61R substitution should reduce the stability of the binding interface by inhibiting its interactions with R130.

4.4.4.1. Analysis of the $Sgt1_2Skp1$ Binding Interface Using Tandem Mass Spectrometry

The crystal structure of $TPR\kappa_3Skp1$ has shown that Sgt1: Skp1 binding is mediated by a 1:1 interaction (Fig. 4.14). MS/MS is a common method used to probe the organisation of multimeric complexes, and has been applied here to study Sgt1: Skp1 binding interface. Dissociation of a single Sgt1 subunit from the wt $Sgt1_2Skp1$ complex would support the 1:1 binding interface observed in the crystal, and indicate that the Sgt1 dimer is not required for stable Skp1 binding. In contrast, if no $Sgt1_1Skp1$ oligomers were observed, this would suggest that the second Sgt1 monomer is required to stabilise the binding interface.

Incubation of wt Sgt1: Skp1 using the method previously described (section 4.3.2.3), produced a heterogeneous mixture of Sgt1 and Skp1 complexes (Fig. 4.24c). MS/MS analysis of the isolated $Sgt1_2Skp1$ +24 (*m/z* 4856) ion at a TriWave Trap collision voltage of 70 V produced a fragmentation pattern consistent with the 1:1 binding interface, with dissociated $Sgt1_1$ and $Sgt1_1Skp1$ complexes (Fig. 4.24a). Further increases in collision voltage (90 V) improved the relative intensity of Sgt1Skp1 (Fig. 4.24b). At higher collision voltages (> 90 V) the signal to noise ratio inhibited observations of all oligomeric species.

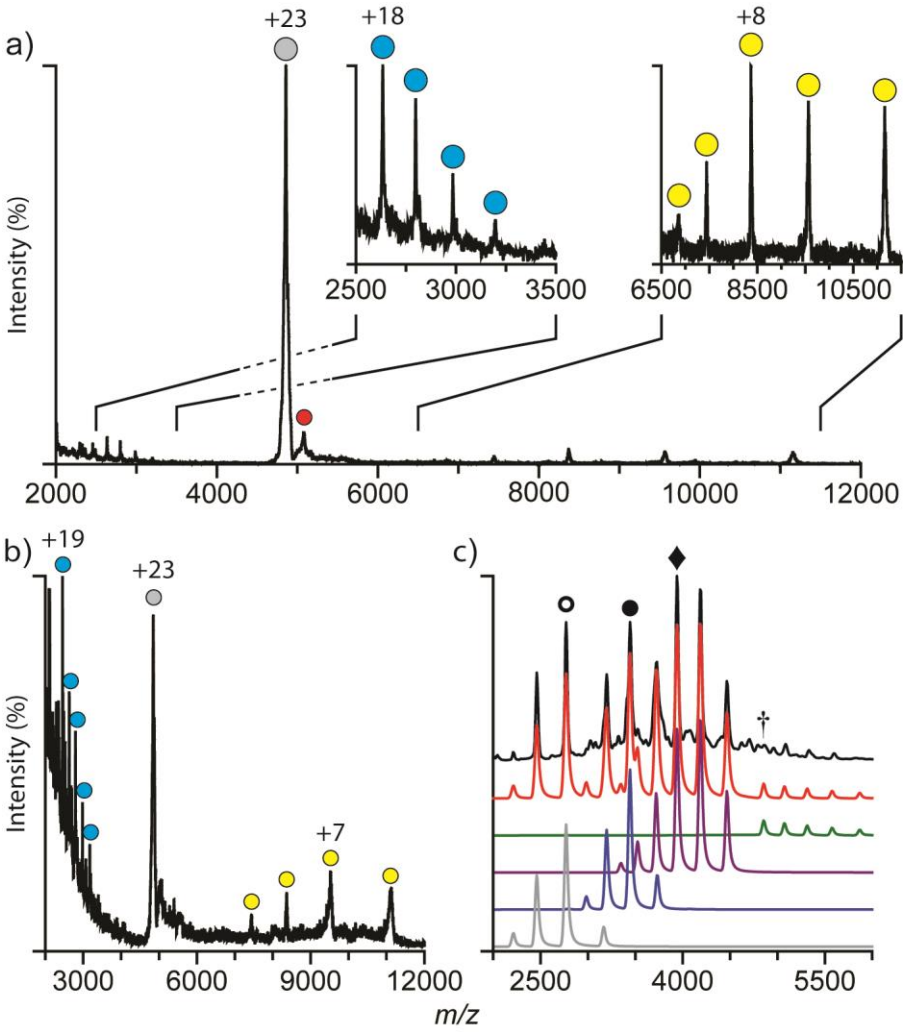

Figure 4.24. Tandem Mass Spectrometry Analysis of the Wild Type $Sgt1_2Skp1$ Complex

a) MS/MS analysis of the wt $Sgt1_2Skp1$, using the isolated +24 (*m/z* 4856, grey sphere) ion at 70 V. Data shows that a single subunit of wt Sgt1 can be dissociated (blue sphere), whilst retaining an intact $Sgt1_1Skp1$ complex (yellow sphere) (red sphere, represents the co-isolated 2:1 +22 ion). b) Increasing the collision voltage to 90 V, allows the 1:1 heterodimer to be more clearly resolved. c) Control spectra Amphitrite analysis shows the presence of unbound Skp1 (grey, +8; ○), $Skp1_2$ (blue, +13; ●) and Sgt1Skp1 (purple, +17; ♦), along with the required $Sgt1_2Skp1$ protein (green, +24; †). Experimentally acquired spectra in black, simulated spectra in red.

4.4.4.2. Analysis of the TPRκ$_2$Skp1 Binding Interface Using Tandem Mass Spectrometry

Following from studies of the Sgt1$_2$Skp1 complex, analysis was performed on the TPRκ$_2$Skp1 oligomer to support the wt data above. This would allow progression of the study to analyse the TPRκ$_3$Skp1 oligomer observed in the crystal structure (Fig. 4.14).

Using the same incubation method as for the wt protein, mass analysis of incubated TPRκ: Skp1 produced a heterogeneous mixture of increased complexity oligomers (Fig. 4.25). To identify the CSD of the TPRκ$_2$Skp1 oligomer spectra deconvolution was required. As this analysis predated the development of the Amphitrite software, deconvolution of these spectra was performed by using the quadrupole to isolate each peak across the *m/z* range. Ion decomposition was then collisionally induced to study the resultant fragmentation patterns.

The CID fragmentation patterns were used to identify the CSD's of TPRκ$_x$Skp1$_y$ oligomers (x and y = number of respective monomeric subunits). These identifications were confirmed using at least one neighbouring charge state ($z = {}^{+}/_{-} 1$) within the theoretical CSD. These data were corroborated by summing the average charge totals of the fragmentation pattern to ensure close agreement with the theoretical value expected of the oligomer at the observed *m/z* value.

Using the identified CSD of the required 2:1 heteromer, a suitable TPRκ$_2$Skp1 ion peak (+15, *m/z* 3791) was isolated using the quadrupole (highlighted as 'Q,' in Fig. 4.25), and analysed using CID (Fig. 4.26). Suitability of ions for quadrupole isolation was defined as an ion peak that did not overlap, within an appropriate *m/z* range ($^{+}/_{-}$ *m/z* 50), any other heteromeric protein that may produce a contradictory fragmentation pattern. This is in addition to retaining a suitable relative intensity for ease of analysis. MS/MS analysis has shown a fragmentation pattern consistent with a TPRκ monomer dissociated from TPRκ$_2$Skp1, retaining an intact TPRκ$_1$Skp1 heteromer. This is in excellent agreement with the wt work above (section 4.4.4.1).

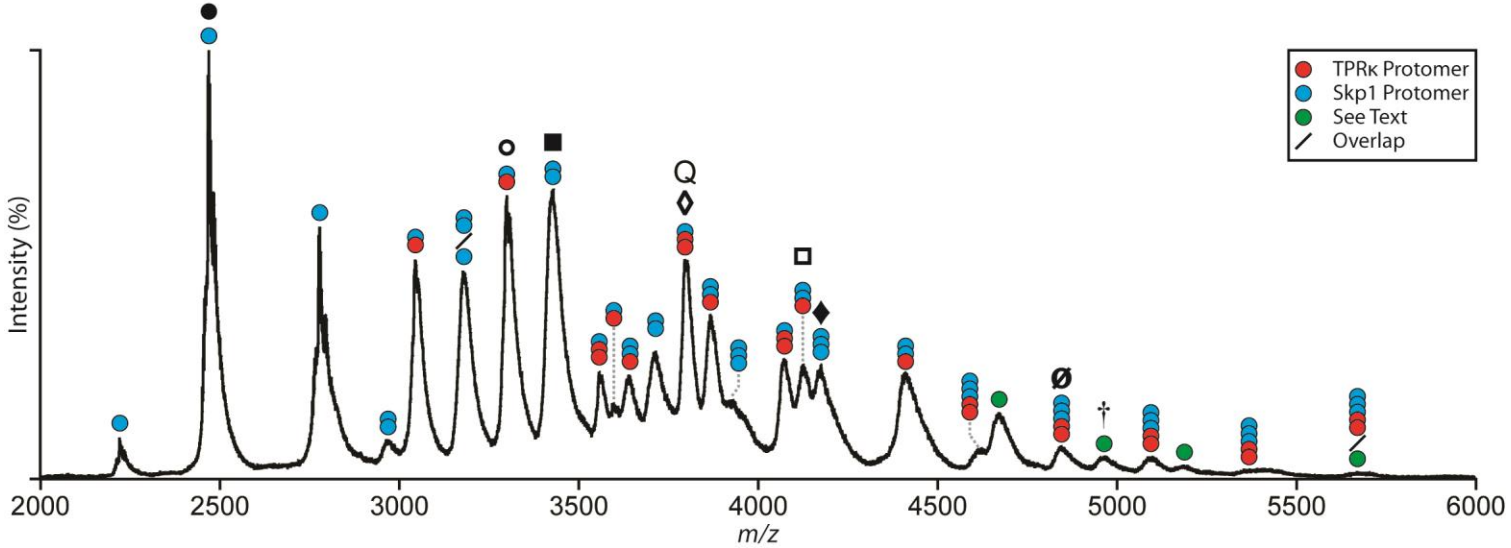

Figure 4.25. Tandem Mass Spectrometry Identified Species Within TPRκ Skp1 Mass Spectra

Oligomer identifications are indicated by the number of respectively coloured overlapping spheres (e.g. 2 blue and 1 red = TPRκ$_2$Skp1). Ion peaks highlighted in green represent an overlap of two potential proteins. Evidence supports either 1:3 (79 kDa) or 2:2 (84 kDa) Sgt1: Skp1 oligomers, due to similar fragmentation patterns and theoretical *m/z* values (within *m/z* 50 between the +14 to +17 charge states).

'Q' denotes the isolated TPRκ$_2$Skp1 +15 (3791 *m/z*) ion used for fragmentation studies (Fig. 4.26). Despite the lack of observed homogenous Sgt1 oligomers, Skp1 monomers are still present in large abundance, relative to other species identified.

Sgt1: Skp1 oligomer charge states: 0:1 +9; ●, 0:2 +13; ■, 0:3 +16; ♦, 1:1 +12; ○, 1:2 +15; □, 2:1 +15; ◊, 2:3 +21; ø, 1:3 +17 or 2:2 +16; †.

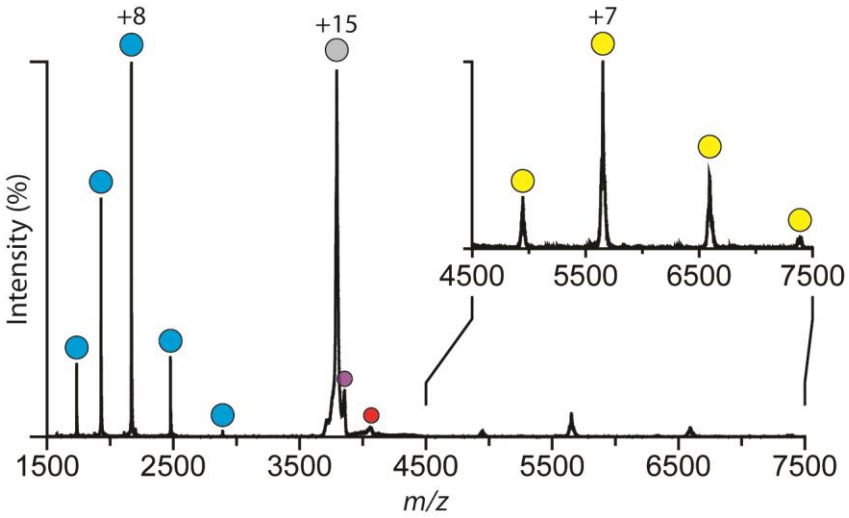

Figure 4.26. Tandem Mass Spectrometry Analysis of the TPRκ Sgt1$_2$Skp1 Complex

The isolated TPRκ$_2$Skp1 (grey) +15 (*m/z* 3791) ion was fragmented using a TriWave Trap collision voltage ramp (10-80 V). Fragmentation pattern at 45 V indicates a TPRκ monomer (blue) can be dissociated to leave an intact TPRκSkp1 heterodimer (yellow) supporting the 1:1, Sgt1: Skp1, binding interface. Also shown are the trace co-isolated +14 TPRκ$_2$Skp1 (*m/z* 4045) ion (red), and +16 TPRκSkp1$_2$ (*m/z* 3851) ion (purple).

4.4.4.3. Progressing to Study the Crystal Structure Relevant TPRκ$_3$Skp1 Binding Interface

Despite the biologically relevant oligomer that binds Hsp90 *in vivo* being Sgt1$_2$Skp1, the recently solved crystal structures of TPRκ, and TPRκ: Skp1 (Fig. 4.9 Willhöft 2013, and Fig. 4.14 Willhöft *et al. In Preparation*) have shown trimeric Sgt1 species incorporated within these oligomers. Therefore, it was proposed that MS/MS could be applied to probe the organisation of TPRκ$_3$Skp1 oligomer to allow a greater understanding of the binding interface.

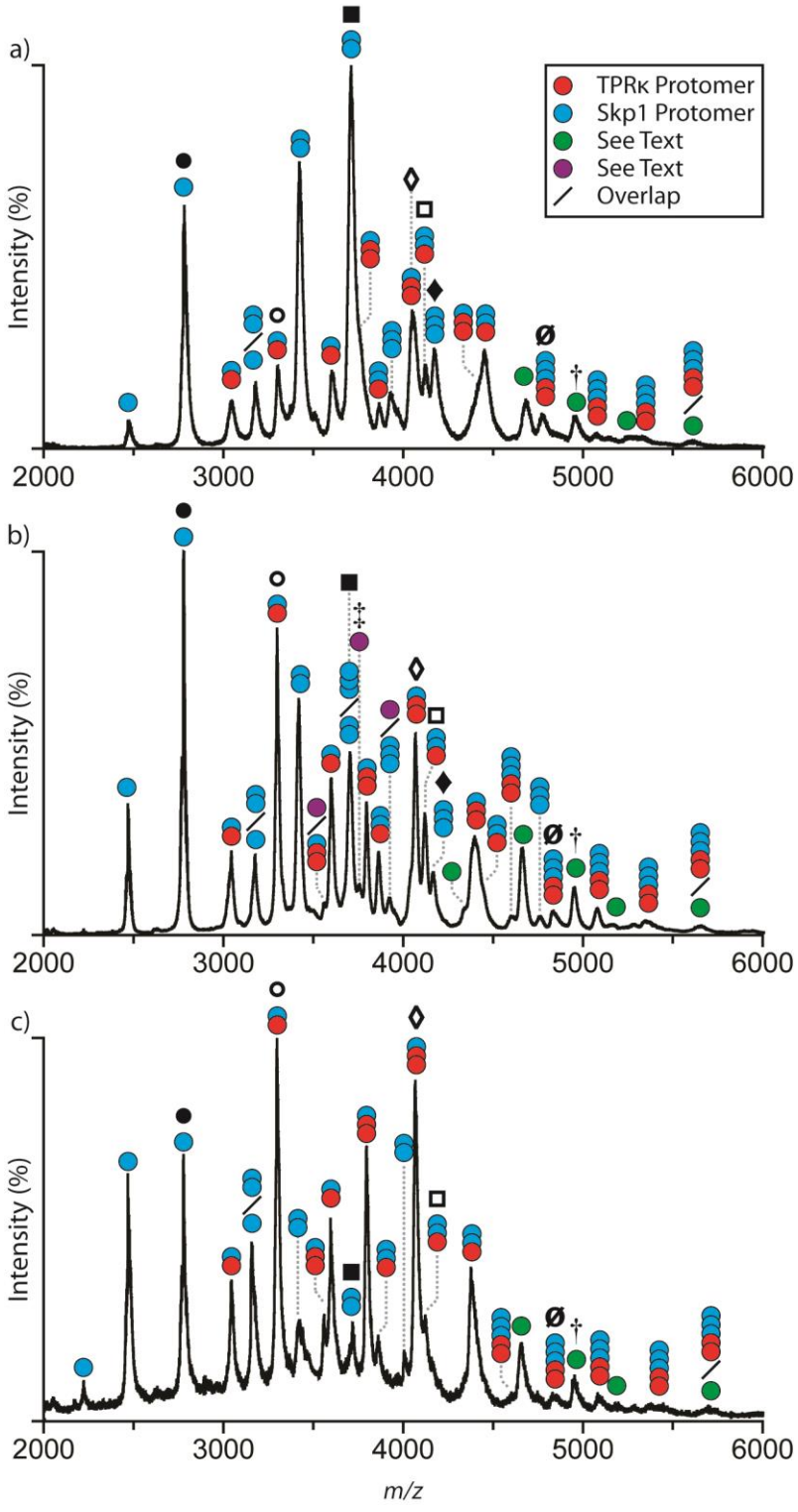

Figure 4.27. Promoting Gas Phase TPRκ$_3$Skp1: Sample Ratio Effects

Data shows TPRκ: Skp1 incubated at ratios of (a) 1:1, (b) 3:1 and (c) 5:1. Method of peak identification is outlined in figure 4.25. As above (Fig. 4.25) the green spheres represent an overlap of two potential proteins, supporting evidence for either a 1:3 (79 kDa) or 2:2 (84 kDa) TPRκ: Skp1 oligomer.

Data shows that as the concentration of TPRκ increases a reduction in higher order Skp1 oligomers occur. Spectra were unable to resolve the required TPRκ$_3$Skp1.

The purple sphere, observed in b, highlights a potential TPRκ$_3$Skp1 +20 (‡, *m/z* 3710.8) ion, albeit at a reduced relative intensity to the other ions observed. Within the theoretical CSD +19 and +21 charge states overlap too closely with other oligomeric species, to actively conclude the presence of the TPRκ$_3$Skp1 oligomer at these conditions. Mass spectrum in part b, represents the only datum that may indicate the presence of TPRκ$_3$Skp1. Sgt1: Skp1 oligomer charge states: 0:1 +8; ●, 0:2 +12; ■, 0:3 +16; ♦, 1:1 +12; ○, 1:2 +15; □, 2:1 +14; ◊, 2:3 +21; ø, 1:3 +17 or 2:2 +16; †, 3:1 +20; ‡.

Accounting for the solvent accessible area of the species, using a projected CSD comparable to the charge increase between 1:1 and 2:1 and globular protein charge prediction equations,[28] the TPRκ$_3$Skp1 oligomer was expected to appear between +17 to +21 (*m/z* 3500 to 4400). [29] Unfortunately, the incubation method used to produce observed TPRκ$_2$Skp1 heteromer studied above, failed to resolve a CSD corresponding to TPRκ$_3$Skp1.

To resolve the TPRκ$_3$Skp1 CSD, incubations varying the solution conditions were performed. Initial attempts modified the ratio of TPRκ to Skp1, studying incubation ratios of 3:1 and 5:1. These ratios were studied due to the observation of TPRκ concentration dependent oligomerisation. At 30 μM, the 3:1 ratio (Fig. 4.27b) would contain a Sgt1 concentration equal to 21 μM, whilst the 5:1 (Fig. 4.27c) would contain 23.9 μM. Therefore, we expected comparable, or improved, levels of TPRκ trimer to those observed during the oligomerisation as a function of concentration work above (section 4.4.2.1). For control methods, we additionally studied an incubation ratio of 1:1, TPRκ: Skp1 (Fig. 4.27a).

Spectra acquired at each of the incubation conditions are presented in figure 4.27, applying the previously described MS/MS fragmentation pattern method to identify TPRκ$_x$Skp1$_y$ oligomers. Incubations of TPRκ: Skp1 at ratios of 3:1 and 5:1 promoted observable improvements in the intensities of the TPRκSkp1 and TPRκ$_2$Skp1 heteromers relative to other oligomers. Excluding a single non-overlapping ion peak (purple sphere Fig. 4.27) that exhibited, with good agreement, a *m/z* value in line with the theoretical +20 (*m/z* 3710.8) ion, the TPRκ$_3$Skp1 CSD remained unobserved.[30] Due to the low intensity of this peak, relative to the neighbour ions, MS/MS analysis could not be performed to confirm its oligomeric identity. Replication of the conditions corresponding to the datum in figure 4.27b failed to reproduce observations of this ion peak.

Subsequently, it was proposed that the key limiting factor to clearly resolve the TPRκ$_3$Skp1 oligomer, was the over abundance of higher order Skp1 oligomers. This homomerisation occurs as a property of the Skp1 surface cysteine amino acids to form

[28] $z = 0.078 \times \sqrt{m}$ (Kebarle & Verkerk 2009)
[29] Theoretical *m/z* series produced using the denatured spectra derived experimental masses of both the TPRκ and Skp1 proteins (section 4.4.1.).
[30] TPRκ$_3$Skp1 +20 theoretical *m/z* = 3710.77, experimental observed peak *m/z* = 3742.8

disulphide bonds between monomers. Therefore, DTT was suspended to a final concentration of 3 mM within the AmAc (150 mM, pH 7) electrospray buffer, to reduce Skp1 disulphide-bridges.

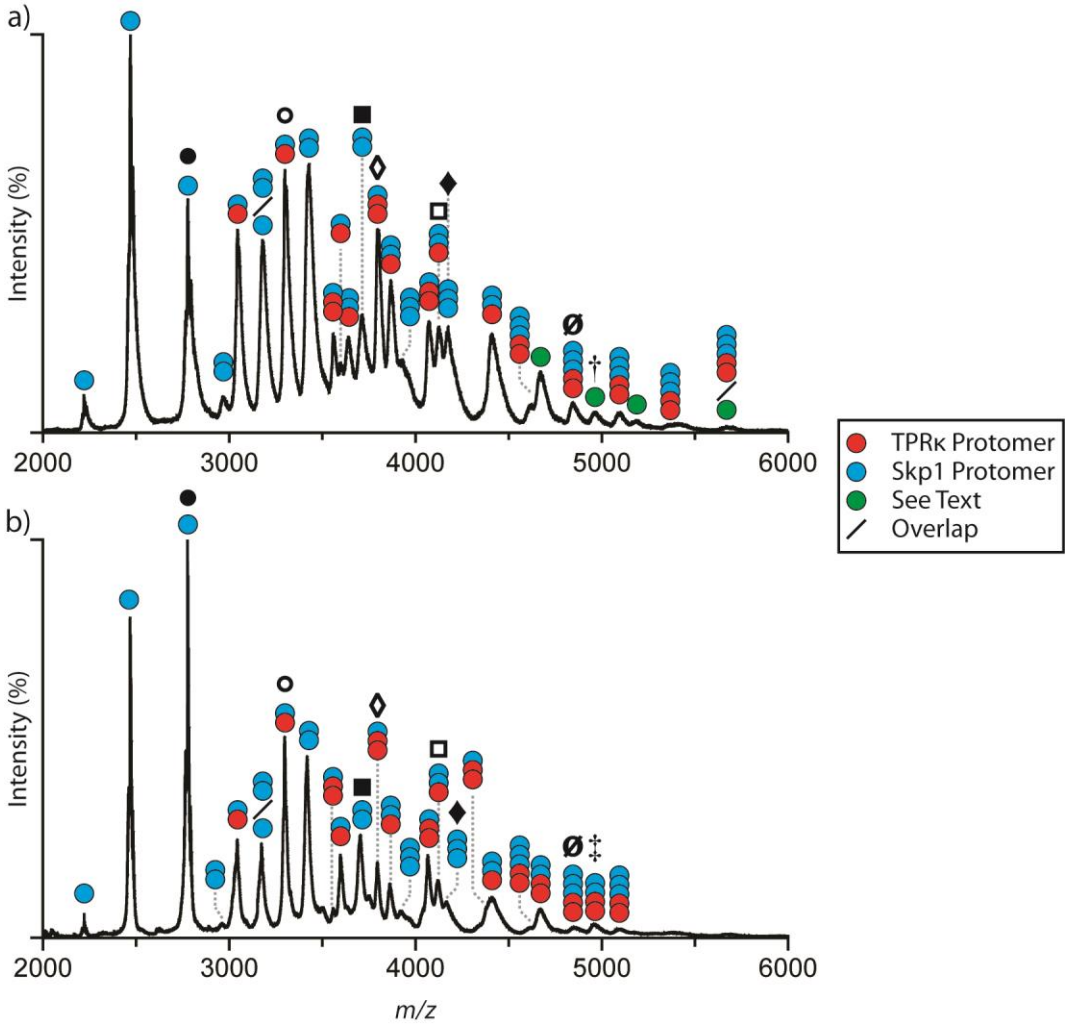

Figure 4.28. Promoting Gas Phase TPRκ$_3$Skp1: Effect of DTT (Pt. I)

The 2:1 ratio incubation of TPRκ: Skp1 readily shows the difference between electrospray buffer both, (b) with, and (a) without DTT (3 mM). In the presence of DTT the previously unknown species suspected as being either the 2:2 or 1:3 oligomer (green sphere) are concluded to be the former (2:2). This does not exclude the conclusion that the 1:3 oligomer may exist in the absence of DTT, and is supported by the presence of Skp1$_3$ oligomers. Sgt1: Skp1 oligomer charge states: 0:1 +8; ●, 0:2 +12; ■, 0:3 +16; ♦, 1:1 +12; ○, 1:2 +15; □, 2:1 +15; ◊, 2:2 +16; ‡, 2:3 +21; ø, 1:3 +17 or 2:2 +16; †.

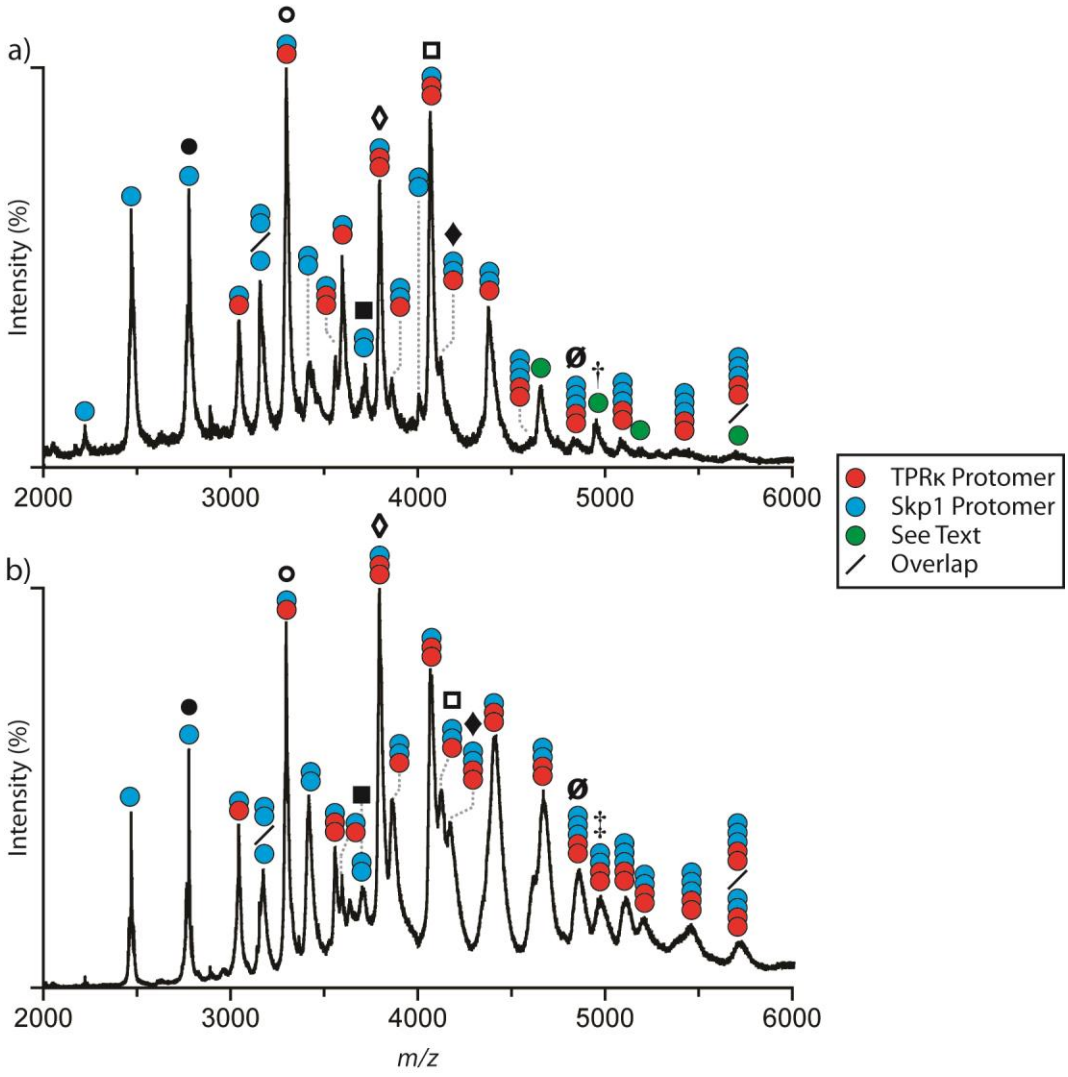

Figure 4.29. Promoting Gas Phase TPRκ$_3$Skp1: Effect of DTT (Pt. II)

5:1 ratio incubation of TPRκ: Skp1 showing the difference between electrospray buffer both, (b) with, and (a) without DTT (3 mM). In the presence of DTT the previously unknown species suspected as being either the 2:2 or 1:3 oligomer (green sphere) are concluded to be the former (2:2). This figure is a continuation of the 2:1 incubation ratio data presented in figure 4.28. Sgt1: Skp1 oligomer charge states: 0:1 +8; ●, 0:2 +12; ■, 0:3 +16; ♦, 1:1 +12; ○, 1:2 +15; □, 2:1 +15; ◊, 2:2 +16; ‡, 2:3 +21; ø, 1:3 +17 or 2:2 +16; †.

Analysis using this modified method was performed using 2:1 and 5:1 concentration ratios of TPRκ: Skp1. The 2:1 incubation ratio replicated the conditions used to produce the crystal required for X-ray crystallography (Fig. 4.14). The 5:1 incubation targeted solution conditions promoting the concentration dependent oligomerisation of TPRκ.

Figure 4.28 compares the observed oligomeric species for the 2:1 incubation of TPRκ: Skp1 in both the presence (b), and absence (a) of 3 mM DTT. Incubations of TPRκ: Skp1 at 5:1 are compared in figure 4.29. Extended analysis using the MS/MS fragmentation pattern method, previously described, again failed to indicate the presence of observable $TPR\kappa_3Skp1$ oligomers despite an apparent reduction in higher order Skp1 oligomers, complexed with and without TPRκ, relative to other species.

Concluding that the 3:1 observed in figure 4.14 was likely an artefact of the crystallisation method, resulting from the high protein concentrations required, further attempts to study this complex were not made. Additionally the MS/MS analysis of the biologically relevant $Sgt1_2Skp1$ oligomer, originally targeted by the investigation, had already been performed.

4.5. Discussion

Comparison of the mass spectrometric data presented here is produced in combination with other data presented by Willhöft (2013), and Willhöft *et al.* (*In Preparation*).

4.5.1. Regulation of Sgt1 Dimerisation

Analysis of wt Sgt1 protein, excluding 22.3 μM at 150 mM AmAc, shows stable oligomer AUC values across the protein concentration range studied (Fig. 4.20 and Fig. 4.23). This is in comparison to the TPRκ protein, which exhibits a concentration dependent mode of oligomerisation. These data suggest that the absence of the C-terminus removes the steric clashes that prevent further oligomerisation interactions, and allow additional W58-H59 stacking interactions between adjacent monomers.

SV-AUC analysis performed by Willhöft (2013) (Fig. 4.30) has shown that the TPRκ is able to form trimers in solution, and is supported by the gas phase analysis presented here. In contrast the TPRκλ (TPRκ plus linker; λ) protein (amino acids 1-178) only forms dimers. Therefore, the TPRκ (amino acids 1-150) likely represents what is close to the structural maximum that supports Sgt1 trimerization, with further additions to the protein sequence inhibiting these oligomerisation events occurring due to steric clashes.

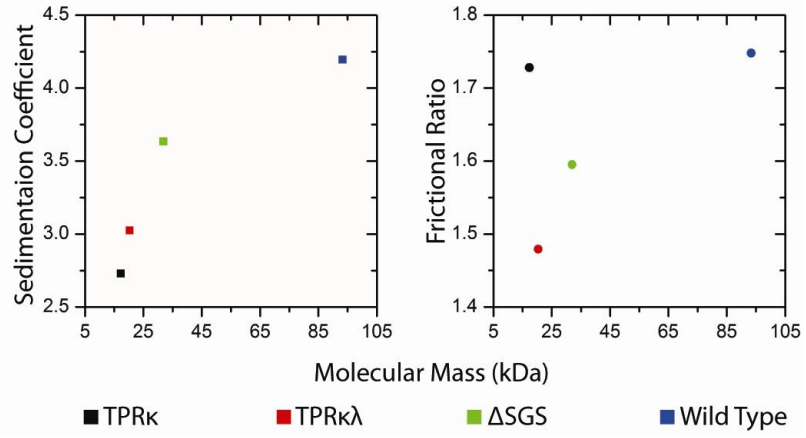

Figure 4.30. Comparison of Sgt1 Proteins Using SV-AUC

SV-AUC analysis of four Sgt1 proteins (molecular weight order), TPRκ (amino acids 1-150), TPRκλ (amino acids 1-178), ΔSGS (amino acids 1-279) and wt (full length). A positive correlation is observed in sedimentation coefficient and increasing monomeric mass (left). In terms of the observed weight-average frictional ratio, (right) proteins encompassing regions of the 151 to 395 amino acid stretch show a similar increasing trend in the frictional ratio, due to the linear addition of structural features that restrict higher order oligomerisation. The TPRκ, which does not encompass these additional amino acids, is able support additional oligomerisation events (i.e. trimerization) accounting for a weight-average frictional ratio increase inconsistent with the higher mass proteins. Data collected and analysed by Oliver Willhöft (Willhöft 2013).

Figure reproduced, with permission, from Willhöft 2013

These data are supported by analysis of the crystal structure analysis by Oliver Willhöft (2013). This analysis has shown that the TPRκ is unlikely to support further oligomerisation events due to the increasing steric clashes between neighbouring Sgt1 monomers. These observations contradict the mass spectrometric data presented here which has shown that the TPRκ protein is able to support increasing levels of tetramer and pentamer oligomers at concentrations exceeding 11.2 μM (Fig. 4.18 and Fig. 4.19). These oligomerisation events are therefore likely to support a model of non-specific gas phase interactions that are known to occur as a function of increased concentration under the charged residue model of analyte desolvation (see Fig. 1.5a) (Sun *et al.* 2007). In a contrasting explanation, these higher order oligomers may form due to a second mode of dimerisation.

Consensus TPR (CTPR) analysis performed by Krachler & co-workers (2010) using CTPR3Y3, a triple mutant variation of CTPR3, indicated a potential second mode of

dimerisation. Mutations within the capping helix destabilised hydrogen bonding that promoted displacement of this secondary structure element to enable a third CTPR3Y3 to bind to the existing dimer. In contrast to the classical stacking interface, this displacement promoted an end-on mode of oligomerisation producing a quasi-continuous oligomer (Fig. 4.31) (Krachler, Sharma & Kleanthous 2010).

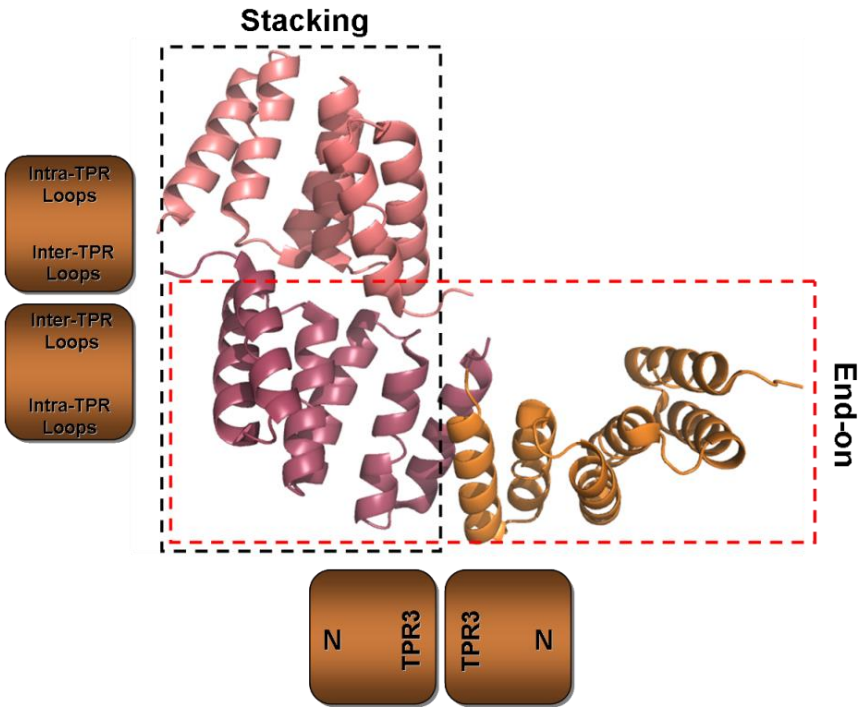

Figure 4.31. Different Modes of TPR Self-Association

The CTPR3Y3 presents two potential modes of self-association of TPR-containing proteins. Dimerisation can occur through the loop-mediated interactions (stacking) or by a hydrophobic end-on association (Krachler Sharma & Kleanthous 2010, PDB ID: 2WQH).

Figure reproduced, with permission, from Willhöft 2013

Within the studies performed here, the open C-terminal nature of the TPRκ may allow the capping helix to dissociate from the dimerisation domain forming a secondary, or end-on, binding interface. Due to the reduced number of amino acids contributing to binding between monomers compared to the stacking interface, the stability of the end-on mode of oligomerisation will be weaker in contrast. Subsequently due to these reduced binding stabilities, we would expect lower levels of these oligomers consistent with the observed tetrameric and pentameric species.

It may be argued that the use of the buffers (AmAc) required for native mass spectrometric analysis, when compared to those used for SV-AUC and SEC-MALS (Size Exclusion Chromatography-Coupled Multi-Angle Light Scattering) (NaCl), may contribute to these extended oligomerisation events. Considering that anions and cations of AmAc and NaCl share similar positions within the Hofmeister series however, suggests that this is unlikely (Curtis & Lue 2006). It is more likely that the SV-AUC and SEC-MALS analysis methods presented by Willhöft (2013), lack the dynamic range, sensitivity, and accuracy required to resolve these higher order oligomers, when compared to the mass spectrometric methods used here.

Analysis of the TPRκ IM data acquired indicates single conformational species for both the monomer and dimer (Fig. 4.17, b and c). In contrast the trimer exhibits a second, less abundant, conformation of 3886.36 ($^{+}/_{-}$ 84.81) $Å^2$ (Fig. 4.17d). This represents a conformational increase of 6.8 % over the primary, more compact, conformation (3635.6 $^{+}/_{-}$ 76.91 $Å^2$). Due to reduced structural compaction resulting from the displaced capping helix of the end-on mode of dimerisation (Krachler, Sharma & Kleanthous 2010), increases in CCS are to be expected and may support that the TPRκ is able to oligomerize using this alternative mode of oligomerisation. Unfortunately, due to the relative intensities of the tetramer and pentamer oligomers in the acquired IM mass spectra, mobility analysis cannot be readily performed to confirm these suggestions.

Although the W58 and H59 amino acids that mediate dimerisation are orientated in a way to support trimerization, analysis performed by Willhöft (2013) has indicated that the C-terminal domains of the wt protein would cause steric clashes between the α and γ monomers preventing wt $Sgt1_3$ formation (Fig. 4.9a). Data presented here however has shown that the wt is able to promote trimerization, in an interaction that is perturbed by the D61R substitution. It may be concluded that these steric clashes reduce higher order binding stability, but are unable to completely inhibit these interactions explaining why we see only low levels of trimer across the 4.5 to 22.3 μM concentration range (Fig. 4.19a). In addition to increasing steric clash events, the extended C-terminus should prevent binding by the end-on stacking interface, further explaining why we see no evidence of tetrameric or pentameric oligomers using the wt.

Comparing oligomerisation as a function of protein concentration (Fig 4.18 and 4.19), the TPRκ supports concentration dependent oligomerisation: predominantly monomeric at 4.5 μM, dimeric at 11.2 μM, with a trend indicating preference for the trimer above 22.3 μM. In contrast, the appearance of wt oligomers appear to be concentration independent, and exhibit a preference for the dimer across the range studied. These observations therefore support the requirement for the C-terminal domains to stabilise the native dimeric oligomer.

Despite indicating a requirement for C-terminal stabilised oligomers, what role these domains play to regulate these dimerisation events is non-resolvable, nor accurately suggests which domain is responsible. The previously outlined SV-AUC and SEC-MALS data indicate that increasing the length of the TPRκ by 18 amino acids (TPRκλ) inhibits trimerization. This is understood to occur because of C-terminal steric clashes, in addition to with the reduced sensitivity of the methods used for analysis compared to mass spectrometry. Comparison with the literature however suggests that the SGS domain, which acts as a site for CK2 mediated phosphorylation inhibiting dimerisation (Bansal *et al.* 2009b), is therefore the most likely site for stabilization. How this SGS domain contributes structurally to the dimer however remains currently unidentified.[31]

Comparison of ionic strength against protein concentration (Fig. 4.22) indicates that increases in ionic strength between 150 mM and 500 mM may be sufficient to destabilise the electrostatic interactions that are required for stable Sgt1 dimerisation. This is supported by similar AUC of each oligomer, within MSD, between 500 mM and 1000 mM AmAc across the protein concentration range. Above 500 mM AmAc, it is expected that only the H59-W58 stacking interaction will support Sgt1 dimerisation. This observation is in line with the previously outlined requirement of this interaction using the H59A substitution (section 4.4.3).

4.5.2. Sgt1, Skp1 Interactions

MS/MS data of both the wt and TPRκ supports that the binding interface for Sgt1: Skp1 binding is a 1:1 interaction. The requirement for the previously identified Sgt1 dimer that

[31] The structure of the SGS domain currently remains unresolved.

mediates *in vivo* interactions between Hsp90 and Skp1 in *S. cerevisiae* is yet to be concluded. These observations are supported by pull down experiments using the monomeric H59A Sgt1 mutant, which is able to bind to Skp1 without prior homo-dimerisation (Willhöft 2013). Combined these observations support the suggestion that Sgt1 dimerisation is not required to promote stable binding of Skp1.

Data presented here expands on previous observations by supporting the requirement of increased collisional voltages to dissociate a single wt subunit from the $Sgt1_2Skp1$ when compared to the $TPR\kappa_2Skp1$ oligomer. These data further support that the requirement of the C-terminal domains that have been proposed here to stabilise Sgt1 dimerisation. These observations are more likely however, to occur due to the increased mass of the wt allowing additional energy transfers with the collision gas, prior to unfolding and dissociation.

4.6. Future Directions

4.6.1. Exploring the Dimerisation Interactions Further

Of interest for future Sgt1 mass spectrometry based studies are the deletion mutants at sequence lengths between the wt and TPRκ, and their influence on oligomer stability. These include the deletion mutants TPRκλ (amino acids 1-178), and ΔSGS (amino acids 1-279, lacks CS-SGS linker domain) previously studied using SV-AUC and SEC-MALS (Fig. 4.30). Using these Sgt1 deletion mutants to study oligomerisation as a function of protein concentration (see section 4.4.2.1) would allow us to support the requirement of either the CS, SGS or TPR to CS domain linker, to promote oligomer stability.

It is hypothesised here that the shortest protein, in amino acid length, exhibiting concentration independent oligomerisation behaviour would indicate that the last domain or linker encoded at the C-terminus is required to regulate stable Sgt1 dimerisation. Therefore, if the TPRκλ exhibits concentration independent oligomerisation, this would indicate that the 18 amino acids directly C-terminus to the TPRκ are sufficient to regulate the Sgt1 dimerisation behaviour. These observations would subsequently support a role for this linker region. In contrast, if the TPRκλ exhibits concentration dependent oligomers,

this would indicate that a domain toward the C-terminus, not encoded by the sequence, is required to regulate stable Sgt1 dimerisation e.g. CS or SGS.

Although CK2 inhibits Sgt1 dimerisation by phosphorylating the residue S361 located within the SGS domain, it has been suggested that there may exist a further inhibitory domain between amino acids 270 to 312 (CS to SGS domain linker) (Bansal *et al.* 2009a and Bansal *et al.* 2009b). Therefore, in addition to the above TPRκλ and ΔSGS Sgt1 mutants I suggest the production of a further Sgt1 deletion mutant that encodes amino acids 1-312, or a suitable nearby residue, for studies of concentration dependent/independent oligomerisation. Although lacking the SGS, this mutant would additionally incorporate the CS to SGS linker domain not encoded by the ΔSGS. Additionally phosphorylated wt Sgt1 may represent a suitable control for analysis, allowing wt dimerisation negative analysis. Combined, these investigations provide an interesting mass spectrometry led method to refine the regulation of Sgt1 dimerisation events.

In addition to the regulation of the oligomeric state by the C-terminal domains, of further interest is the influence of ionic strength on the stability of the dimerisation domain. Data presented using wt Sgt1 has suggested that destabilisation of the extended electrostatic interactions at the binding interface occurs between 150 and 500 mM AmAc. Therefore refining this range to study an extended number of ionic strengths, e.g. 80, 220, 290, 360, 430, and 600 mM AmAc, may indicate the minimum salt concentration suitable to destabilise binding.

Expanding these studies to incorporate the D61R substitution, which removes a single electrostatic interaction, may allow us to probe the effect of these conditions further. It is expected that initial losses of higher order oligomers would occur more rapidly than the wt, due to the reduced stability of the dimer interface. At higher ionic strengths, above 500 mM AmAc, oligomer AUC values are expected to plateau at similar values to the wt once all the electrostatic dimerisation interactions are negated. If the wt and D61R plateau at the similar oligomer AUC contributions, we may conclude that only the W58-H59 interaction is stabilising the binding interface. This would support the role of ascomycete specific loop for dimerisation.

Data presented here has shown that at the highest ionic strength, 1000 mM AmAc, TPRκ oligomer contributions have yet to reach a plateau. Therefore, it is proposed that

future studies incorporate AmAc concentrations above the 1000 mM limit studied here, e.g. 1250, 1500, 1750 and 2000 mM. Given a suitable range, we would expect to observe a plateau similar to oligomer contributions observed with the wt Sgt1. This hypothesis should only remain true if there are no added C-terminal contributions to oligomerisation.

4.6.2. Studying the Sgt1 Mediated Hsp90: Skp1 Binding Interactions Further

Hsp90: Sgt1 binding occurs in a 1:1 interaction between the Sgt1 CS, and Hsp90 ATPase binding domain (Fig. 4.13, Zhang *et al.* 2008). Probing this interaction may indicate why *in vivo S. cerevisiae* Hsp90: Skp1 interactions are mediated by Sgt1 dimers, whilst the human homologue requires only monomeric Sgt1 (Zhang *et al.* 2008).

It is still to be shown however if the stable $Hsp90_2Sgt1_2$ oligomer is mediated by one or two CS: ATPase domain binding pairs. Therefore, studying $Hsp90_2Sgt1_2$ using MS/MS would permit studies of these interactions. If a Sgt1 monomer can be successfully dissociated from the complex, with an observed $Hsp90_2$Sgt1 complex, these data would support that a single CS: ATPase pair mediates the Hsp90: Sgt1 interaction. In contrast, if a Sgt1 dimer can be dissociated, with an observed $Hsp90_2$ oligomer, this would indicate that stable binding requires two CS: ATPase interactions.

Although the H59A substitution inhibits dimerisation, pull down experiments have shown that this protein retains Skp1 binding activity (Willhöft 2013). As this mutation only modifies the homo-dimerisation domain, the Hsp90: Sgt1 interaction should remain unaffected. Therefore, H59A could be used to study if a monomer of Sgt1 is able to stably bind the Hsp90 dimer. If no Hsp90: H59A interactions are observed this interaction would support a role of the Sgt1 dimer to promote formation of this $Hsp90_2Sgt1_2$ complex. In contrast, if a single H59A monomer can bind Hsp90 this would allow progression of investigations to study the interactions that mediate Hsp90: Skp1 binding further. If this $Hsp90_2$H59A complex is unable to allow Skp1 binding, these data may indicate that Sgt1 dimerisation is a prerequisite to reduce steric clashes between Hsp90 and Skp1. This suggestion is supported by an observed 90˚ rotation between TPRκ monomers upon dimerisation (Fig. 4.9).

4.6.3. Further Studies of the Potential Secondary TPRκ Binding Interface

Although not directly relevant to the study of the Sgt1 mediated Hsp90: Skp1 interactions, a method to analyse the potential secondary binding domain is proposed here. These data would potentially explain the extended oligomerisation behaviour observed during gas phase TPRκ analysis.

Assuming continued concentration dependent oligomerisation, extending the TPRκ concentration range above 22.3 μM (at 150 mM AmAc) should promote increased levels of the tetramer and other higher order oligomers. It is proposed that increased intensity species would enable analysis of the structural conformations of these oligomers.

IM data would provide constraints for the computational modelling of TPRκ higher order oligomers. Combining these IM constraints, with the crystal structure of the TPRκ binding interface and data presented by Krachler, Sharma & Kleanthous (2010), computational Sgt1 TPR models of the end-on of dimerisation can be produced. Sequentially these data would confirm or refute the potential end-on stacking interface that has been proposed to account for the appearance of higher order TPRκ oligomers including tetramers and pentamers. If models refute the potential of this end-on oligomerisation behaviour, this would indicate that the proposed TPRκ C-terminal steric clashes (Willhöft 2013) are unable inhibit sequential stacking interactions.

4.7. References

Ali, M.M.U, *et al.* 2006 "Crystal Structure of an Hsp90-Nucleotide-p23/Sba1 Closed Chaperone Complex," Nature, Vol.440, pp.1013-1017

Austin, M.J., *et al.*, 2002 "Regulatory Role of SGT1 in Early R Gene-Mediated Plant Defences," Science, Vol.295 pp.2077-2080

Azevedo, C., 2002 "The RAR1 Interactor SGT1, an Essential Component of R Gene-Triggered Disease Resistance," Science, Vol.295 pp.2073-2076

Bansal, P.K., Abdulle, R., & Kitagawa, K., 2004 "Sgt1 Associates with Hsp90: an Initial Step of Assembly of the Core Kinetochore Complex," Molecular and Cellular Biology, Vol.24 pp.8069-8079

Bansal, P., *et al.*, 2009a "Sgt1 Dimerisation is Required for Yeast Kinetochore Assembly," Journal of Biological Chemistry, Vol.284 pp.3585-3592

Bansal, P., *et al.* 2009b "Sgt1 Dimerisation is Negatively Regulated by Protein Kinase CK2-Mediated Phosphorylation at Ser361," Journal of Biological Chemistry, Vol.284 pp.18692-18698

Borkovich, K., *et al.*, 1989 "Hp82 Is an Essential Protein That Is Required in Higher Concentrations for Growth of Cells at Higher Temperatures," Molecular & Cellular Biology, Vol.9 pp.3919-3930

Brueker, K., McLafferty, F., 2008 "Stepwise Evolution of Protein Native Structure with Electrospray into the Gas Phase, 10^{-12} to 10^{-2} s" Publications of the National Academy of Sciences, Vol.105 pp.18145-18152

Camahort, R., *et al.* 2007 "Scm3 is Essential to Recruit the Histone h3 Variant cse4 to Centromeres and to Maintain a Functional Kinetochore," Molecular Cell, Vol.26 pp.853-865

Catlett, M., & Kaplan, K., 2006 "Sgt1p is a Unique Co-Chaperone that Acts as a Client Adaptor to Link Hsp90 to Skp1p," Journal of Biological Chemistry, Vol.281 pp.33739-33748

Cho, U., *et al.* 2011 "Molecular Structures and Interactions in the Yeast Kinetochore," Cold Spring Harbour Symposia on Qualitative Biology, Vol.75 pp.395-401

Cunningham, C.N., Krukenberg, K.A., & Agard, D.A., 2008 "Intra- and Intermonomer Interactions are Required to Synergistically Facilitate ATP Hydrolysis in Hsp90," Journal of Biological Chemistry, Vol.283 pp.21170-21178

Curtis, R.A., & Lue, L., 2006 "A Molecular Approach to Bioseparations: Protein-Protein and Protein-Salt interactions," Chemical Engineering Science, Vol.61 pp.907-923

Dollins, D.E., *et al.*, 2007 "Structures of GRP94-Nucleotide Complexes Reveal Mechanistic Differences Between the Hsp90 Chaperones," Molecular Cell, Vol.28 pp.41-56.

Espelin, C.W., Kaplan, K.B., & Sorger, P.K., 1997 "Probing the Architecture of a Simple Kinetochore Using DNA-Protein Crosslinking," Journal of Cell Biology, Vol.139 pp.1383-1396

Flom, G., *et al.* 2006 "Effect of Mutation of the Tetratricopeptide Repeat and Asparatate-Proline 2 Domains of Sti1 on Hsp90 Signalling and Interaction in Saccharomyces cerevisiae," Genetics, Vol.172 pp.41-51

Gardner, R., *et al.* 2001 "The Spindle Checkpoint of the Yeast Saccharomyces cerevisiae Requires Kinetochore Function and Maps to the CBF3 Domain," Genetics, Vol.157 pp.1493-1502

Gasteiger, E., *et al.* 2005, "Protein Identification and Analysis Tools on the ExPASy Server; (In)" John M. Walker (Editor): The Proteomics Protocols Handbook, Humana Press pp.571-607

Hartl, F.U., Bracher, A., & Hayer-Hartl, M., 2011 "Molecular Chaperones in Protein Folding and Proteostasis," Nature, Vol.475 pp.324-332

Hirano, T., *et al.*, 1990 "A Repeating Amino Acid Motif in CDC23 Defines a Family of Proteins and a New Relationship Among Genes Required for Mitosis and RNA Synthesis," Cell Vol.60 pp.319-328

Jäschke, Y., *et al.* 2011 "Pleiotropic Corepressors Sin3 and Ssn6 Interact with Repressor Opi1 and Negatively Regulate Transcription of Genes Required for Phospholipid Biosynthesis in the Yeast Saccharomyces cerevisiae," Molecular Genetics & Genomics, Vol.285 pp.91-100

Jehn, B., Niedenthal, R., & Hegemann, J.H., 1991 "In Vivo Analysis of the Saccharomyces cerevisiae Centromere CDEIII Sequence: Requirements for Mitotic Chromosome Segregation," Molecular Cell Biology, Vol.11 pp.5212-5221

Jiang, W., Lechner, J., & Carbon, J., 1993 "Isolation and Characterization of a Gene (CBF2) Specifying a Protein Component of the Budding Yeast Kinetochore," Journal of Cell Biology, Vol.121 pp.513-519

Johnson, J., 2012 "Evolution and Function of Diverse Hsp90 Homologs and Cochaperone Proteins," Biochimica et Biophysica Acta, Vol.1823 pp.607-613

Jones, L.J., Haugl, and R.P., & Singer, V.L., 2003 "Proteomic Technologies Product Application Focus Development and Characterization of the NanoOrange® Protein Quantitation Assay: A Fluorescence-Based Assay of Proteins in Solution," BioTechniques, Vol.34 pp.850-861.

Kaplan, K.B., Hyman A.A., & Sorger, P.K., 1997 "Regulating the Yeast Kinetochore by Ubiquitin-Dependent Degradation and Skp1p-Mediated Phosphorylation," Cell, Vol.9 pp.491-500

Kebarle, P., & Verkerk, U., 2009 "Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know," Mass Spectrometry Reviews, Vol.28 pp.898-917

Kim, N., *et al.*, 2006 "A New Function of Skp1 in the Mitotic Exit of Budding Yeast Saccharomyces cerevisiae," The Journal of Microbiology, Vol.44 pp.641-648

Kitagawa, K., *et al.*, 1999 "SGT1 Encodes an Essential Component of the Yeast Kinetochore Assembly Pathway and a Novel Subunit of the SCF Ubiquitin Ligase Complex," Molecular Cell, Vol.4 pp.21-33

Kitagawa, K., & Hieter, P., 2001 "Evolutionary Conservation Between Budding Yeast and Humans Kinetochores," Nature Reviews; Molecular Cell Biology, Vol.2 pp.678-687

Kitagawa, K., *et al.* 2003 "Requirement of Skp1-Bub1 Interaction for Kinetochore-Mediated Activation of the Spindle Checkpoint," Molecular Cell, Vol.11 pp.1201-1213

Krachler, A., Sharma, A., & Kleanthous, C., 2010 "Self-Association of TPR Domains: Lessons Learned from a Designed, Consensus-Based TPR Oligomer," Proteins, Vol.78 pp.2131-2143

Krukenberg, K.A., *et al.*, 2011 "Conformational Dynamics of the Molecular Chaperone Hsp90," Quarterly Reviews of Biophysics, Vol.44 pp.229-225

Lingelbach, L.B., & Kaplan, K.B., 2004 "The Interaction Between Sgt1p and Skp1p is Regulated by HSP90 Chaperones and is Required for Proper CBF3," Molecular and Cellular Biology, Vol.24 pp.8938-8950

Lee, Y., *et al.*, 2004 "Humans Sgt1 Binds HSP90 through the CHORD-Sgt1 Domain and not the tetratricopeptide repeat domain, "The Journal of Biological Chemistry, Vol.279 pp.16511-16517

Makhnevych, T., & Houry, W.A., 2012 "The Role of Hsp90 in Protein Complex Assembly," Biochimica et Biophysica Acta, Vol.1823 pp.674-682

Main, E.R.G., Jackson, S., and Regan, L., 2003 "The Folding and Design of Repeat Proteins: Reaching a Consensus" Current Opinion in Structural Biology, Vol.13 pp.482-489

Main, E.R.G., *et al.* 2003 "Design of Stable α-Helical Arrays from an Idealized TPR Motif," Structure, Vol.11 pp.497-508

Mizuguchi, H., *et al.* 2007 "Scm3 is Essential to Recruit the Histone h3 Variant cse4 to centromeres and to Maintain a Functional Kinetochore," Cell, Vol.129 pp.1153-1164

Neckers, L., & Workman, P., 2012 "Hsp90 Molecular Chaperone Inhibitors: are we there yet?" Clinical Cancer Research, Vol.18 pp.64-76

Niikura, Y., & Kitagawa, K., 2003 "Identification of a Novel Splice Variant: Humans SGT1B (SUGT1B)," DNA Sequence, Vol.14 pp.436-441

Nyarko, A., *et al.* 2007 "TPR-Mediated self-association of plant SGT1," Biochemistry, Vol.46 pp.11331-11341

Ortiz, J., 1999 "A Putative Protein Complex Consisting of Ctf19, Mcm21, and Okp1 Represents a Missing Link in the Budding Yeast Kinetochore," Genes & Development, Vol.13 pp.1140-1155

Pearl, L., & Prodromou, C., 2006 "Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery," Annual Review of Biochemistry, Vol.75 pp.271-294

Peng, Y., *et al.* 2011 "Overlapping Kinetochore Targets of CK2 and Aurora B Kinases in Mitotic Regulation," Molecular Biology of the Cell, Vol.22 pp. 2680-2689

Perriches, T., & Singleton, M.R., 2012 "Structure of Yeast Kinetochore Ndc10 DNA-binding Domain Reveals Unexpected Evolutionary Relationship to Tyrosine Recombinases," Journal of Biological Chemistry, Vol.287 pp.5173-5179

Purvis, A., & Singleton, M.R., 2008 "Insights into Kinetochore-DNA Interactions from the Structure of Cep3Delta," EMBO Reports, Vol.9 pp.56-62

Richter, K., *et al.* 2001 "Coordinated ATP Hydrolysis by the Hsp90 Dimer," Journal of Biological Chemistry, Vol.276 pp.33689-33696

Rodrigo-Brenni, M.C., *et al.*, 2004 "Sgt1p and Skp1p Modulate the Assembly and Turnover of CBF3 Complexes Required for Proper Kinetochore Function," Molecular Biology of the Cell, Vol.15 pp.3366-3378

Russell, I.D., Russell, Grancell, A.S., & Sorger, P.K., 1999 "The Unstable F-box Protein p58-Ctf13 Forms the Structural Core of the CBF3 Kinetochore Complex," Journal of Cell Biology, Vol.145 pp.933-950

Salbo, R., *et al.* 2012 "Travelling-Wave Ion Mobility Mass Spectrometry of Protein Complexes: Accurate Calibrated Collision Cross-Sections of Human Insulin Oligomers," Rapid Communications in Mass Spectrometry, Vol.26 pp.1181-1193

Scarff, C., *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 Spectrometry," Rapid Communications in Mass Spectrometry, Vol.22 pp.3297-3304

Sikorski, R.S., *et al.* 1990 "A Repeating Amino Acid Motif in CDC23 Defines a Family of Proteins and a New Relationship Among Genes Required for Mitosis and RNA Synthesis," Cell, Vol.60 pp.307-317

Sivalingam, G.N., *et al.* 2013 "Amphitrite A Program for Processing Travelling Wave Ion Mobility Mass Spectrometry Data" International Journal of Mass Spectrometry Special Edition, Vol.345-347 pp.54-62

Spiechowicz, M., *et al.*, 2007 "Hsp70 is a New Target of Sgt1—An Interaction Modulated by S100A6," Biochemical and Biophysical Research Communications, Vol.357 pp.1148-1153

Stemmann, O., *et al.* 2002 "Hsp90 Enables Ctf13p/Skp1p to Nucleate the Budding Yeast Kinetochore," Proceedings of the National Academy of Sciences, Vol.99 pp.8585-8590

Sun, J., *et al.*, 2007 "Method for identifying nonspecific protein-protein interactions in nanoelectrospray ionization mass spectrometry," Analytical Chemistry, Vol.79 pp.8301-8311

Taipale, M., *et al.*, 2012 "Qualitative analysis of HSP90-client interactions reveals principles of substrate recognition," Cell, Vol.150 pp.987-1001

Vaughan, C., *et al.*, 2006 "Structure of an Hsp90-Cdc37-Cdk4 complex," Molecular Cell, Vol.23 pp.697-707

Vaughan, C., *et al.*, 2009 "A Common Conformationally Coupled ATPase Mechanism for Yeast and Humans Cytoplasmic HSP90s," The FEBS Journal, Vol.276 pp.199-209

Verghese, J., 2012 "Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (Saccharomyces cerevisiae) as a Model System," Microbiology and Molecular Biology Reviews, Vol.76 pp.115-158

Westermann, S., Drubin, D., & Barnes, G., 2007 "Structures and Functions of Yeast Kinetochore Complexes," Annual Review of Biochemistry, Vol.76 pp.563-591

Willhöft, O., 2012, Discussion Regarding the Oligomerisation of the Hsp90:Sgt1 wild type Oligomerisation Study, Including Theoretical Masses and Approximate K_d values, [Email] (Personal Communication 24th May 2012)

Willhöft, O., 2013 "Structural and Biophysical Characterisation of Sgt1, an Essential Component in the Assembly of the Yeast Kinetochore," Ph.D. Birkbeck College, University of London

Willhöft, O., *et al. In Preparation* "Linking Hsp90 to E3 Ubiquitin Ligases and Kinetochore Assembly: Structure of the Sgt1-Skp1 Complex"

Yoshida, Y., Murakami, A., Tanaka, K., 2011 "Skp1 Stabilizes the Conformation of F-Box Proteins," Biochemical and Biophysical Research Communications, Vol.410 pp.24-28

Zabka, M., *et al.*, 2008 "Sgt1 has Co-Chaperone Properties and is Up-Regulated by Heat Shock," Biochemical and Biophysical Research Communications, Vol.370 pp.179-183

Zeytuni, N., & Zarivach, R., 2012 "Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module," Structure, vol.20 pp.397-405

Zhang, M., *et al.* 2008 "Structural and Functional Coupling of Hsp90- and Sgt1-Centred Multi-Protein Complexes," EMBO Journal, Vol.27 pp.2789-2798

Zhang, M., *et al.*, 2010 "Structural Basis for Assembly of Hsp90-Sgt1-CHORD Protein Complexes: Implications for Chaperoning of NLR Innate Immunity Receptors," Molecular Cell, Vol.39 pp.269-281

5. Analysis of CbpA's Structure and DNA Binding Interactions

The *Escherichia coli* DNA binding protein CbpA promotes chromosome compaction during the stationary cell cycle phase, and cellular phosphate starvation events, forming aggregates exceeding 60 nm. The organisation of CbpA and DNA within these aggregates is currently unsolved. Native mass spectrometry has been applied here to study these CbpA-DNA interactions. Although unsuccessful in our attempts to observe these interactions using native mass spectrometry, these studies support that the buffers required for electrospray ionization are suitable for CbpA-DNA binding. Using IM-MS, we have additionally provided gas-phase constraints for computational modelling of the biologically relevant DNA binding CbpA dimer.

5.1.1. CbpA Structure and Expression

Figure 5.1. Domains of CbpA

Consistent with all type II J-domain proteins CbpA contains three domains: DnaK binding J-domain (red, amino acids 1-75), DNA binding CTD I (blue, amino acids 118-201) and CTD II (yellow, amino acids 202-306) that promotes dimerisation (Bird *et al.* 2006, and Cosgriff *et al.* 2010).

The *Escherichia coli* nucleoid associated DNA storage, and chaperone protein, CbpA (Curved binding protein A) (Uniprot ID; P36659, 34.5 kDa, 306 amino acids), is a type II J-domain protein (Fig. 5.1) observed in solution as a mixture of monomers and dimers (Bird *et al.* 2006, and Cosgriff *et al.* 2010). Currently, no complete structure of CbpA has been solved using high-resolution structural techniques (e.g. NMR or X-ray crystallography), with only a partial NMR structure of the J-domain available (PDBe file; 2KQX, amino acids 2-72) (Sarraf *et al.* 2010). Other investigations have applied sequence alignments to produce homology models, including Bird and co-workers (2006) (Fig. 5.2). Excluding a model of the CTD II dimerisation domain by Cosgriff and co-workers (2010)

(Fig. 5.3), a complete model of the CbpA dimer required for DNA binding has not yet been solved.

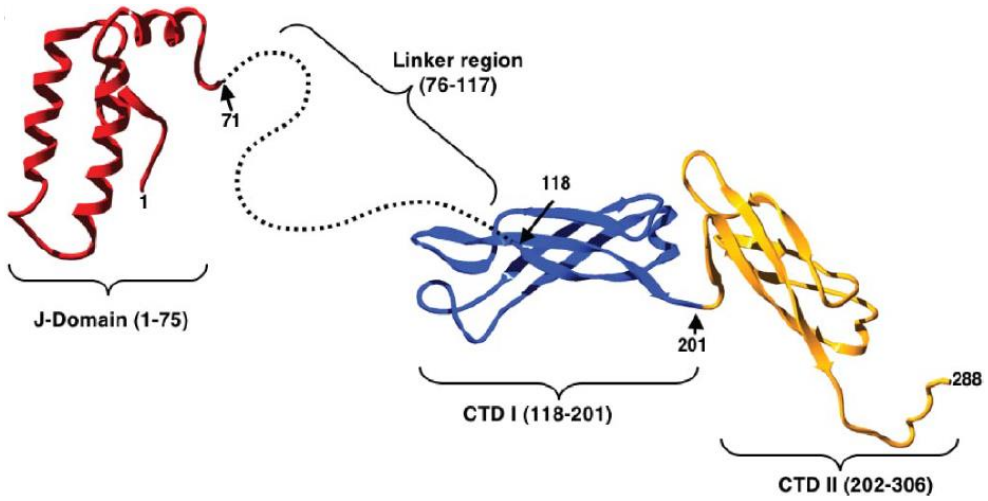

Figure 5.2. Sequence Alignment Based Model of CbpA

Model developed using Swiss Model (Schwede *et al.* 2003) alignment against several J-domain proteins for the N terminal region, and the *S. cerevisiae* Hsp40 protein, Sis1, for CTD I and II alignment. No structural homologues are known to exist for the proposed 'flexible' linker region (amino acids 76-117) (Bird *et al.* 2006). It is this linker region in addition with CTD I that represent the known DNA binding domain (Cosgriff *et al.* 2010).

Figure reproduced from Bird *et al.* 2006

The operon *cbpA* is under the control of σ^S (Yamashino *et al.* 1994) and Lrp (Chenoweth & Wickner 2008). Both σ^S and Lrp are controlled by a σ38-associated RNA polymerase. Expression of CbpA during the late growth to stationary phase occurs under the control of the σ^S operon (Yamashino *et al.* 1994). σ^S promoted expression results in the translation of an estimated 3 000 to 15 000 copies of CbpA, per cell (Azam *et al.* 1999, Chae *et al.* 2004, and Cosgriff *et al.* 2010), equating to ~3 % of the overall cellular protein population during the stationary phase (Gur, Katz & Ron 2005). Upon expression, CbpA rapidly (2 to 5 minutes) binds DNA, promoting chromosome compaction (Cosgriff *et al.* 2010).

Upregulated translation of *cbpA* by Lrp occurs during amino acid and phosphate starvation conditions, arresting the cell cycle through chromosome compaction events (Chenoweth & Wickner 2008, Wegrzyn *et al.* 1996, and Yamashino *et al.* 1994). Whilst

Lrp promotes upstream of σ^S, it remains dependent on the latter to enable transcription. This is in line with the broad induction behaviour of Lrp, which can promote the translation of ~10 % of all *E. coli* genes (Tani *et al.* 2002). In strict contrast to Lrp, the σ^S is able to promote transcription independently of Lrp (Chenoweth & Wickner 2008).

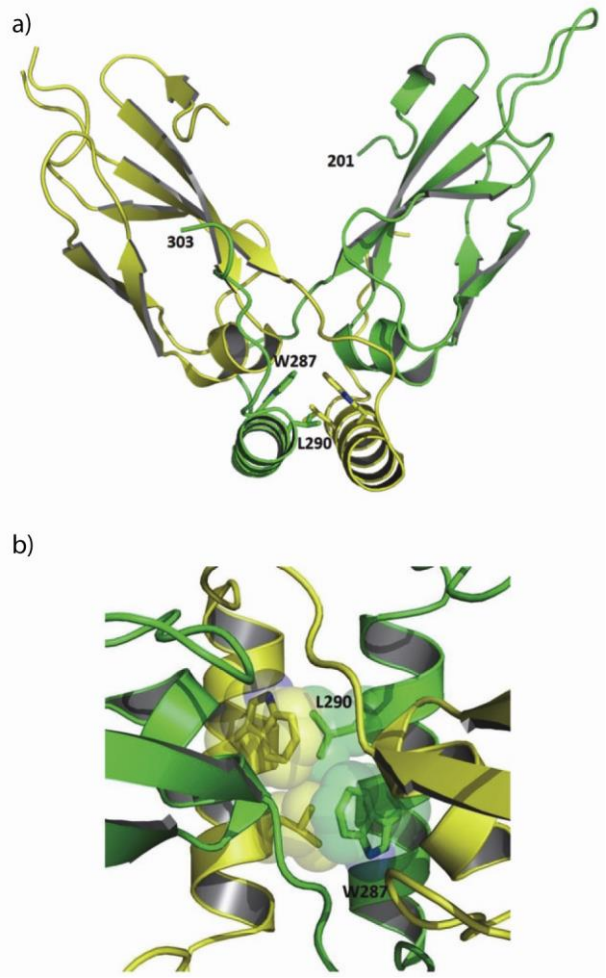

Figure 5.3. Model of the CbpA CTD II Dimerisation Interface

a) Structural model of the CbpA CTD II only dimer, with the component monomers shown in yellow and green respectively.

b) Dimerisation is stabilised by two sets of complementary L290-W287 interactions between neighbouring monomers (Cosgriff *et al.* 2010).

Figure reproduced and adapted from Cosgriff *et al.* 2010

5.1.2. Binding of DNA by CbpA

The first observations of CbpA-DNA binding interactions were made by Ueguchi and co-workers (1994), whereby CbpA was shown to bind preferentially to curved dsDNA (double stranded DNA). Subsequent work has shown that the DNA binding activity of CbpA is sequence non-specific, as super compaction of plasmid DNA by CbpA prevents

nuclease degradation by DNAase I. In contrast, sequence specific binding proteins do not regulate these degradation events (Cosgriff *et al.* 2010).

Upon formation of CbpA dimers, DNA binding is stabilised by the CTD I, with additional contributions from the linker domain (Bird *et al.* 2006, and Cosgriff *et al.* 2010). These binding events occur in a strongly co-operative manner, whereby once one CbpA dimer has bound DNA this promotes further proteins to bind the same length of DNA (Azam & Ishihama 1999). Due to this positive feedback mechanism of binding, CbpA is able to rapidly (2 to 5 minutes) produce aggregates of 60 nm in diameter (Fig. 5.4) (Cosgriff *et al.* 2010).

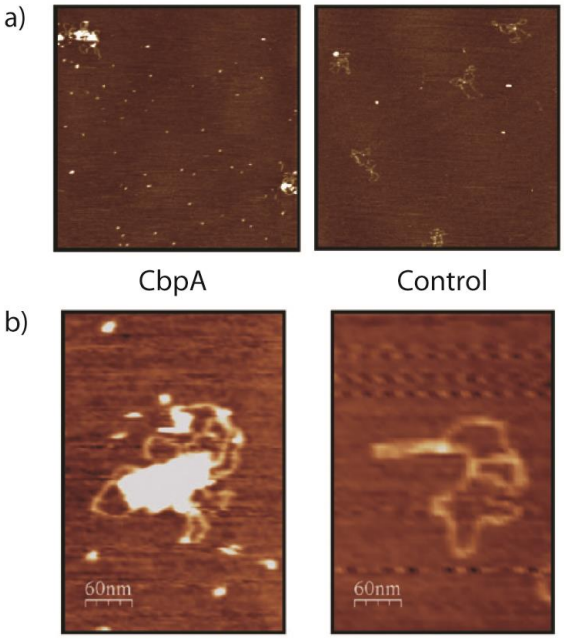

Figure 5.4. Atomic Force Microscopy Analysis of CbpA in the Presence of Plasmid DNA

a) AFM analysis of a 1.4 x 1.4 μm region of mica surface incubated in the presence of plasmid DNA, both with and without CbpA (Control = Plasmid only). In the presence of protein, an observed level of protein-DNA aggregation is observed. b) Zooming in on a single aggregate, allows direct comparison of the aggregate against the control. Data indicates that the aggregate may contain more than one plasmid. Component concentrations: CbpA 100 nM, Plasmid DNA 20 ng (Cosgriff *et al.* 2010).

Figure reproduced and adapted from Cosgriff *et al.* 2010

The characteristics of CbpA promoted DNA compaction are similar to the protein Dps (DNA protection during starvation). Dps bends the nucleotide backbone to bridge distant loci promoting genome super-compaction (Browning *et al.* 2010, and Dame *et al.* 2011). Both CbpA and Dps share similar roles in protecting DNA during the stationary phase. This is in addition to protective roles in acidic stress and phosphate/nucleotide starvation condition responses. Similarly, organisation of both CbpA and Dps within these protein-DNA aggregates remains unsolved.

5.1.2.1. Comparative Binding of *E. coli* DNA by Dps

Dps (Uniprot ID; P0ABT2, 18.7 kDa, 167 amino acids) expression is strongly upregulated during the stationary phase, exhibiting the highest copy number of any nucleoid protein during this time (up to 100 000 copies) (Cosgriff *et al.* 2010). This is consistent with observations outlining the pre-requisite for Dps in bacterial chromosome compaction during this time point of the cell cycle (Wolf *et al.* 1999).

Expression of the *dps* operon is controlled by a σ70-associated RNA polymerase (Grainger *et al.* 2008). Once translated, Dps binds DNA non-sequence specifically, with uniform aggregate distribution and nucleoid biocrystallization (Wolf *et al.* 1999). Down regulation of Dps by Fis and H-NS selectively mediates expression, either directly at the *dps* operon (Fis), or indirectly through inhibition of the promoter (H-NS) (Grainger *et al.* 2008).

Dps expression is further upregulated in response to a range of stress conditions: phosphate starvation (Almirón *et al.* 1992), high pressure, UV and gamma irradiation (Calhoun & Kwon 2011, and Zeth 2012). These expressions protect DNA through binding events, and are able to inhibit the deleterious effects of fenton-mediated oxidation[32] by trapping hydroxyl ions (Bellapadrona *et al.* 2010). This is in addition to protection of DNA during acid stress conditions, which if unchecked can promote strand breakage (Jeong *et al.* 2008).

Azam & Ishihama (1999) have previously compared the DNA binding activity of Dps and CbpA. Their results showed that Dps binds without preference to curved and linear DNA sequences (K_d = 172 nM and 178 nM respectively). In contrast, CbpA preferentially binds curved DNA at a K_d of 122 nM (linear DNA 175 nM) (Azam & Ishihama 1999). These data indicate that in the presence of curved DNA, CbpA is able to bind more rapidly than Dps, with no observable differences in the presence of linear DNA.

[32] H_2O_2 + Fe^{2+} source, that produces a pool of ·OH & ·OOH free radicals

5.1.3. The Chaperone Protein DnaK and its Interactions with J-Domain Proteins

In addition to promoting compaction of the cellular genome, CbpA is able to function as a DnaK co-chaperone. The chaperone protein DnaK (Uniprot ID; P0A6Y8, 69.1 kDa, 638 amino acids) is one of three *E. coli* Hsp70 (Heat shock protein 70 kDa) class proteins (Chae *et al.* 2004, and Ogata *et al.* 1996). DnaK contains two domains, a nucleotide binding domain with ATPase activity (amino acids 1-370), and a C-terminal substrate binding domain (amino acids 390-600) (Bertelsen *et al.* 2009). Hydrolysis of ATP promotes conformational changes on the C-terminal substrate-binding domain opening a 'lid' segment that enhances DnaK chaperone activity. Chaperone binding at the C-terminus strongly promotes further ATPase activity in the N-terminus (Bertelsen *et al.* 2009, and Srinivasan *et al.* 2012).

Within the heat shock response, DnaK binds to a variety of proteins in an ATP dependent manner (Bird *et al.* 2006). These interactions include binding of DnaK to DNA Gyrase to protect cellular DNA under thermal stress conditions. Ogata and co-workers (1996) proposed a model of protein expression to tackle the effects of heat shock through the regulation of DNA supercoiling. During heat shock, topoisomerase I and DNA Gyrase function together to relax super coiled DNA. This relaxation induces the expression of proteins that are required to respond to, and tackle the effects of heat shock. This relaxation promotes the induction of DnaK that interacts with DNA Gyrase to enhance the negative supercoiling of DNA preventing further transcription events (Ogata *et al.* 1996).

DnaK interacts with three J-domain containing proteins[33] to modulate stress responses: DnaJ, CbpA, and DjlA (Chae *et al.* 2004, and Sarraf *et al.* 2010). J-domain proteins are classified into one of three groups (I, II and III). Members of all three of these classes encode a ~75 amino acid J-domain that allows them to bind Hsp70 class proteins, stimulating their ATPase activity (Bird *et al.* 2006, and Sarraf *et al.* 2010). In addition to this J-domain, Type I proteins are further classified by their phenylalanine rich region, cysteine rich zinc finger domain, and two similarly folded CTDs. Unlike the type II and III, type I J-domain proteins are able to function independently of Hsp70 during heat shock. In contrast, Type II J-domain proteins, such as CbpA, lack the Zn finger domain indicative

[33] *E. coli* expresses a total of six J-domain containing proteins (Patury *et al.* 2009).

of the type I and are unable to prevent aggregation of non-native substrates in the absence of a co-chaperone (Bird *et al.* 2006, and Walsh *et al.* 2004). Type-III proteins encode only the J-domain (Bird *et al.* 2006).

5.1.3.1. DnaK: DnaJ Interactions

DnaJ (Uniprot ID; C4ZPU1, 41.1 kDa, 376 amino acids) is an Hsp40 class protein, required to regulate *E. coli* Hsp70 homologues, including DnaK (Walsh *et al.* 2004).

Peptide sequence analysis has shown that the binding site of 95 % of DnaK binding proteins overlap with DnaJ (Srinivasan *et al.* 2012). One such example of these interactions is the co-operative binding between ClpB, DnaJ and DnaK. Upon formation, this complex is able to disaggregate in the region of 250 proteins, sequentially promoting their correct refold. DnaJ complexes such as these promote the recognition of conformational unfolding characteristics, and are critical to responses requiring the identification of a diverse range of substrates (Ballet *et al.* 2012, and Mogk *et al.* 1999) accelerating their refolding pathway (Srinivasan *et al.* 2012).

5.1.3.2. DnaK: CbpA Interactions

CbpA binds to DnaK to act as a co-chaperone in response to stress conditions including low pH (Tucker, Tucker & Conway 2002). Expression under these stress conditions is controlled by an unidentified operon. CbpA subsequently promotes refolding of cytoplasmic proteins compatible with its DnaJ like refolding activity (Tucker, Tucker & Conway 2002).

Observations made using *cbpA* deletion mutants have shown no influence on *E. coli* growth. In contrast, *dnaJ* knockout mutants lead to strains that can be characterised by the inability to grow below 16 °C, or above 37 °C (Ueguchi *et al.* 1995). Combined, *dnaJ cbpA,* double deletion mutants are unable to undergo either cell growth or division at permissive temperatures, similar to strains with *dnaK* deletions. Insertion of a multicopy *cbpA* containing plasmid into these double deletion mutants effectively restores DnaK functionality enabling cell growth and division (Chenoweth, Trun & Wickner 2007, and

Ueguchi *et al.* 1995). This indicates that in a DnaJ knockout strain, CbpA is able to act as a functional homologue essential for cell growth and division (Chenoweth, Trun & Wickner 2007, and Ueguchi *et al.* 1995).

5.1.4. The CbpA Modulator Protein: CbpM

The *E. coli* modulator protein CbpM (Uniprot ID; P63264, 11.5 kDa, 101 amino acids) is a regulatory protein that inhibits with high specificity the co-chaperone and DNA binding activities of CbpA (Chae *et al.* 2004, and Chenoweth *et al.* 2007).

Both the σ^S and Lrp operons that contain *cbpA* additionally encode the CbpA regulatory protein CbpM through the downstream (3' end) *cbpM* (Synonym; *yccD*) (Chae *et al.* 2004, and Chenoweth & Wickner 2008). This combined *cbpAM* operon is strongly conserved across all γ-proteobacteria (Chintakayala & Grainger 2011). Co-expression results in similar transcription and translation levels, promoting complementary cellular concentrations (Chenoweth & Wickner 2008). Work by other individuals has additionally shown that CbpA is required to maintain the stability of CbpM; otherwise, it is rapidly degraded by the Lon and ClpP proteases expressed under the σ^{32} response factor[34] (Chenoweth & Wickner 2008, and Meyer & Baker 2011).

CbpM binds 1:1 to the CbpA J-domain (Bird *et al.* 2006, and Chae *et al.* 2004). Recent data presented by Chintakayala & Grainger (2011) has shown, using mutagenesis techniques, that this interaction is stabilised by a H33-E62 interaction between CbpA and CbpM respectively. Substitutions of CbpA R26E and R30E have been additionally shown to disrupt the CbpM binding activity (Chintakayala & Grainger 2011). As well as modifying CbpM binding, these R26E and R30E substitutions reduce CbpA dimerisation two fold. This observation is inconsistent with the known CbpA dimerisation domain (CTD II) that occurs at the direct C-terminal opposite to these two mutations. Therefore, the authors (Chintakayala & Grainger 2011) concluded that a level of structural regulation occurs between the two domains.

In addition to inhibiting CbpA-DNA interactions through the CTD I domain, binding by CbpM also prevents the binding of DnaK. This inhibition results from

[34] This σ^{32} response factor additionally regulates the expression of DnaK and DnaJ (Meyer & Baker 2011).

competitive interactions at the shared binding domain (Sarraf *et al.* 2010). Subsequently, CbpM is able to regulate all forms of CbpA-ligand/co-chaperone activity (Chae *et al.* 2004, and Chenoweth *et al.* 2007) (Fig. 5.5). How the CbpA binding activity of CbpM is regulated, modifying observed levels of DnaK/DNA binding has yet to be determined.

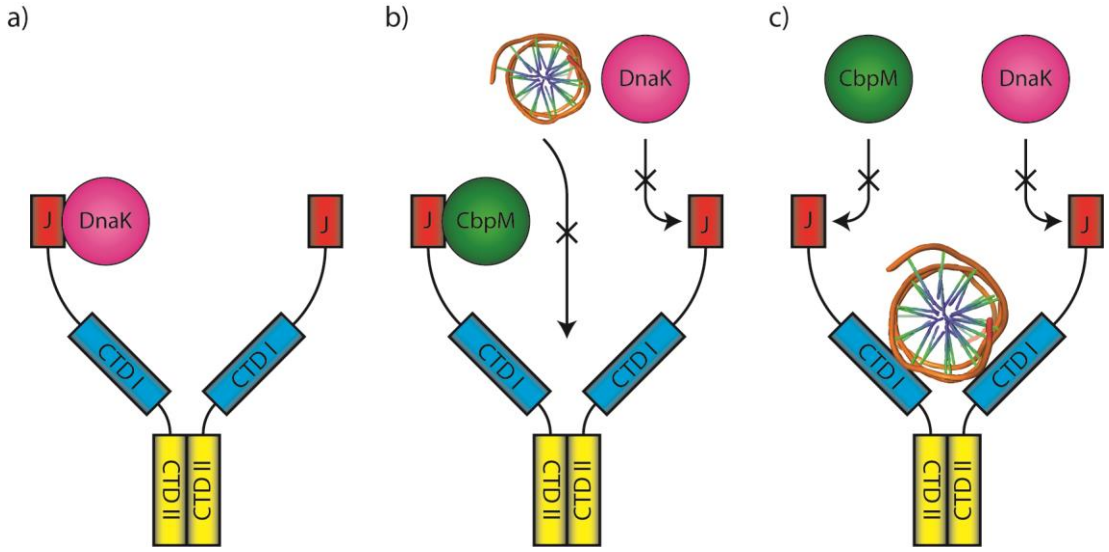

Figure 5.5. Regulation of CbpA Binding Activity by CbpM

a) CbpA: DnaK binding is mediated through the CbpA J-domain (J), stimulating the intrinsic ATPase activity of the DnaK. b) Binding of CbpM to the J domain however, inhibits DnaK and DNA binding interactions. c) Binding of DNA to the dimeric CbpA CTD I domain promotes the rapid formation of protein-DNA aggregation, further inhibiting co-chaperone-binding function (Chae *et al.* 2004, and Cosgriff *et al.* 2010). Image not to scale, or indicative of the size and structure of CbpM and DnaK. DNA PDB ID; 3BSE.

5.2. Experimental Aims and Objectives

The first gas phase observations of native dsDNA using ESI were performed by Light-Wahl and co-workers (1993). These observations were completed using a 20 bp length of dsDNA, and negative mode ionisation. Using negative mode micro-ESI Cheng and co-workers (1996) would expand on these observations presenting the first application of mass spectrometry to observe intact protein-DNA oligomers of the f1 bacteriophage Gene V protein and a 15 base complex of ssDNA (single strand DNA). Since these early experiments, mass spectrometry has been applied to the study of proteins that bind ssDNA

(Cheng *et al.* 1996), dsDNA (Arbely *et al.* 2011), and RNA (Basnak *et al.* 2010) constructs, all of varying length. Studies such as these have been completed using proteins that bind specifically to specific amino acid sequences, and structures (Basnak *et al.* 2010, and Xu *et al.* 1999). Both positive (Atmanene *et al.* 2010) and negative methods of ionisation (Cheng *et al.* 1996), as well as conformational studies using IM-MS (Atmanene *et al.* 2010) have been applied to complete these studies.

Despite studies of aggregating proteins and protein-DNA interactions now widespread,[35] research on systems combining both these characteristics, i.e. aggregating DNA binding proteins, using mass spectrometry based methods has not been currently cited within the literature. Successful development of a method to allow the investigator to probe these interactions would therefore present a promising new direction within the fields of mass spectrometry and IM-MS.

Investigations recently performed by Cosgriff and co-workers (2010) using atomic force microscopy (AFM) have shown that the *E. coli* nucleoid protein, CbpA, readily aggregates non-specifically on both linear, and plasmid DNA. Incubations of plasmid DNA and CbpA have been shown to produce protein-DNA aggregates that exceeded 60 nm in diameter (Fig. 5.4). The conformation of CbpA within these aggregates and the oligomerisation pathway that promotes their formation remains currently unidentified. Therefore, work presented here has targeted the study CbpA binding events using a number dsDNA constructs of varying lengths. Initial experiments will apply native mass spectrometry to optimise the conditions required to observe CbpA-DNA oligomers. These optimised conditions would sequentially allow the study of conformational changes exhibited by these oligomers using IM-MS.

Of further scientific interest are the interactions between CbpA and CbpM known to regulate the DNA and co-chaperone binding activity of the former. Although the binding interface supports a 1:1 interaction, the exact stoichiometry of this oligomer in addition to the influence of CbpM on the dimeric state of CbpA is not conclusively understood.

Recent work by Chintakayala and Grainger (2011) has indicated a potential mode of J-domain regulation of the oligomeric state. Substitutions of amino acids required for

[35] For a larger overarching review on the topic of aggregating protein studies by mass spectrometry the author directs the reader to the following review, Ashcroft (2010).

CbpM binding (R26E and R30E) additionally reduced the levels of observed CbpA dimers. Using mass spectrometry and MS/MS, in both the presence and absence of CbpM, the stability and organisation of CbpA dimers would be studied using a number of dimerisation domain mutants. These mutations include the hypothesised increased dimerisation affinity mutants, L224A and M275A, and reduced affinity mutants, P219A and L286A.

5.2.1. Average CCS Values for Computational Restraints

IM-MS wave speed and amplitude conditions have been clustered according to their calibration curve R^2 values.[36] These clustered calibrations have been subsequently applied to produce an average CCS for the biologically relevant CbpA dimer. This average CCS has been subsequently used by Harpal Sahota to apply constraints to computational models of the CbpA dimer. A brief over-view of the methods applied to complete this work are outlined below, and excluding any mention of the IM-MS calculated CCS values is a representation of work being performed by Harpal Sahota, under the primary supervision of Dr. Maya Topf (Sahota *Personal Communication* 2013).

The initial step in construction of the computational structure is the production of a homology alignment model. This model simulates the protein at the atomic level, using alignments of identified homologous structures. Identified PDB homologies are 2KQK, an NMR derived structure of the CbpA J-domain, and 3LZ8, which exhibits an 84 % sequence identity to the two CTDs.

For sequences lacking identified homologous structures, such as the linker region (amino acids 76-117), I-Tasser (Roy, Kucukural & Zhang 2010) is applied for *Ab-initio* modelling methods. Using I-Tasser, several secondary structure elements have been identified within the linker region and cross-correlated with other structure prediction methods, including, Psi-Pred (Jones 1999) and RaptorX (Peng & Xu 2011). This cross-comparison has indicated the presence of a potential helical structure toward the C-terminal end of the linker domain.

[36] The R^2 value provides a measure of how well calibrant ions fit the power fit curve used to calculate the CCS of the unknown analyte.

After cross-correlation of identified linker region structures, the programs Modeller (Eswar *et al.* 2006) and I-Tasser were applied to refine the known and predicted structural elements of CbpA. These methods have resulted in the production of 176 different models. Modeller and Chimera (Pettersen *et al.* 2004) were subsequently applied to refine these 176 models, removing structures with sub-optimised backbone arrangement and/or steric clashes. The remaining 114 models are currently being compared against the average experimental CCS value of the lowest charged dimer +15 ion, using the PA and EHSS methods, data forthcoming (Sahota *et al. In Preparation*).

5.3. Materials and Methods

5.3.1. Molecular Biology, Preparation, and Purification of CbpA

All CbpA protein samples were purified by the Grainger research group, prior to buffer exchange and mass spectrometric analysis protocols, using the methods outlined below.

cDNA fragments encoding CbpA were cloned into the vector pET21a and sequentially transfected into *E. coli* T7 express cells (New England Biolabs, MA, USA). Vector containing cultures were incubated at 37 ˚C to an OD_{650} of ~0.5 nm. At this OD_{650} value, over expression of CbpA was induced using IPTG, with incubation continuing for one hour prior to harvest.

Post cell fraction recovery and lysis, pooled samples were purified using two chromatography steps: NaCl gradient over a HiTrap Heparin Sepharose column (> 95 % purity), and HiTrap QFF column (> 99 % purity). All columns purchased from GE Healthcare Life Sciences (Buckinghamshire, UK).

Samples were supplied in the following final elution buffer: 20 mM Tris, 1 mM EDTA, 100 mM NaCl (pH 7.5). Please note that for long term sample storage at -80 ˚C, certain samples contained an additional 10 % glycerol content. Glycerol containing samples were dialysed using 6-8 kDa MWCO dialysis tubing (Spectrum Labs, Netherlands), immersed in 100 mM AmAc (pH 7.5) at 4 ˚C for 16 hours. To ensure rapid solution equilibrium, samples were gently agitated over the time course using a magnetic stirrer.

5.3.1.1. *De Novo* DNA Construct Design

De novo dsDNA constructs were optimised to minimise secondary structure elements and miss-aligned dimerisation between complementary forward and reverse ssDNA oligomers using OligoAnalyser v3.1 (Owczarzy *et al.* 2008). These designed *de novo* ssDNA oligomers additionally encoded 2 to 3 GC bases at both the 5' and 3' end of the sequence to stabilise the dsDNA structure.

Optimised dsDNA T_m (melting temperature) values were calculated using OligoCalc (Kibbe 2007). Design refinements ensured that the T_m of correctly aligned dsDNA forward/reverse complementary sequences exceeded, by at least 10 ˚C, those of any stable secondary structural elements.

5.3.2. Purification and Preparation of CbpA

Prior to analysis using mass spectrometric methods, samples were buffer exchanged using a single Micro Bio-Spin column (BioRad, Hertfordshire, UK), and three sequential washes using a single Amicon 1.5 ml 10 kDa MWCO centrifugal concentrator column (Millipore, Dundee, UK). If required, up to an additional two washes using a single Amicon centrifugal concentrator column were applied to improve mass spectra peak separation. All columns were prepared using the experimental concentration of AmAc at pH 7.5.

Buffer exchanged homogeneous samples were measured using the Qubit 2.0 fluorometer (Invitrogen, Paisley, UK), with a 1 in 100 dilution of sample to running buffer. In contrast, heterogeneous CbpA-DNA samples were measured using serial dilution Nanodrop 1000 (Thermo-Scientific, DE, USA) analysis. Extinction coefficients (table 5.1) were applied to correct Nanodrop 1000 mg/ml values. Using the corrected mg/ml values, sample molarity was calculated using the experimentally calculated mass (from denatured protein mass analysis) (table 5.1) of the CbpA dimer.

5.3.2.1. Purification and Preparation of DNA Constructs

Complementary forward and reverse ssDNA oligomers were purchased from Integrated DNA Technologies (IA, USA). Prior to buffer exchange and annealing, forward and reverse ssDNA complements were mixed in equal molar ratios to minimise unbound ssDNA. DNA buffer exchange utilised a modified version of the ethanol-sodium acetate precipitation method, replacing sodium cations with a source of ammonium ions (AmAc) more suitable for native ESI analysis. This method, outlined below, was repeated three times to improve sample purity:

1. DNA sample diluted with 0.1x volume of 7.5 M AmAc.
2. DNA sample diluted with 2x volume of 100 % ethanol, and gently mixed.
3. Sample centrifuged for 15 minutes at 12 000 g.
4. Residual buffer discarded using a pipette.
5. DNA pellet rinsed with 150 μL 70 % ethanol.
6. Residual ethanol removed by evaporation.
7. DNA pellet resuspended in a suitable volume of AmAc buffer.

If required, an additional Amicon 1.5 ml 10 kDa MWCO centrifugal concentrator column (Millipore, Dundee, UK) was used after annealing the complementary ssDNA constructs to improve mass spectra peak separation.

Diluted complementary ssDNA constructs were annealed using a Grant (Cambridge, UK) BTD dry block heating system, and the following incubation method:

1. Diluted DNA sample incubated for ten minutes, at 2-5 ˚C above the calculated dsDNA T_m to ensure dissociation of all secondary structure elements.
2. Incubation temperature dropped to 2-5 ˚C below the dsDNA T_m (ensuring this still exceeded undesirable secondary structure T_m).
3. Incubation continued for ten minutes allowing complementary strands to anneal.
4. Incubation temperature dropped by a further 5 ˚C.
5. Incubation continued for ten minutes.
6. Incubation returned to room temperature prior to measuring sample concentration.

T_m values of dsDNA constructs were calculated using OligoCalc (Kibbe 2007) adjusted to the ionic strength of the buffer. Post annealing of the forward and reverse

compliments, dsDNA sample concentration was measured using the Qubit 2.0 fluorometer (Invitrogen, Paisley, UK), and the theoretical mass values of the designed constructs (table 5.1).

5.3.3. Ion Mobility Analysis of CbpA

The following conditions are relevant to the work performed in section 5.4.1. IM data acquisitions were performed using a broad range of TriWave mobility separator wave amplitude and velocity conditions. Samples were buffer exchanged into 150 mM AmAc (pH 7.5) and diluted to 8 μM. Wave amplitude conditions were acquired over a range of 8 to 12 V in increments of 1 V. These conditions were compared over a range of wave velocity values between 200 to 400 m/s, in increments of 50 m/s. Combination of these values resulted in 25 analytical conditions.

Calibration of the experimental t_D values, to produce relevant CCS, was performed using the method outlined in section 2.2. Denatured Equine myoglobin, native BSA and *S. cerevisiae* ADH (Sigma-Aldrich Ltd, Dorset, UK) were used as the CCS calibrants. Clustering of power fit calibration curve R^2 values was applied to optimise the calculated average CCS values.

Additional instrumental values for mass spectrometric analysis were: capillary voltage; 0.9 to 1.5 kV, nanoflow pressure; 0.00 to 0.5 Bar, sampling cone; 80 V, extraction cone; 1 V, source temperature; 40 ˚C, Trap collision; 10 V, transfer collision; 10 V, Trap DC bias; 20 V, Trap gas; 5 ml/min, IM separator gas flow; 20 ml/min. Instrumental pressures: source backing; 6 mBar, Trap; 0.04 mBar, mobility separator; 0.52 mBar, ToF; $2x10^{-6}$ mBar. The *m/z* range utilised for sample analysis was 1 000 to 16 000, over two minutes using two second scans.

5.3.4. DNA Electrophoretic Mobility Shift Assay of CbpA-DNA Samples

DNA gel-electrophoretic mobility shift assays (EMSA) analysis was performed using pre-cast 6 % DNA retardation gels (Invitrogen, Paisley, UK), and 0.5 x Tris-Borate-EDTA buffer. Prior to loading of sample onto the gel, incubations of CbpA: DNA were diluted

using 10 % glycerol. Samples were measure against 100 bp DNA ladder (New England Biolabs, MA, USA), and control DNA samples. Empty lanes were loaded with a 10 % glycerol solution diluted with doubly deionised H_2O. All samples were loaded onto the gel to a final volume of 15 μL.

Electrophoretic separation was performed at 100 V 15 mA for 90 minutes. To resolve DNA bands, the gel was soaked in a dilute ethidium bromide solution (< 0.1 mg/ml) for two hours. Gel bands were resolved under UV light and photographed using the BioDoc-It imaging system (UVP, CA, USA).

5.4. Results

5.4.1. Control CbpA Analysis

In solution, CbpA exists as a mixture of monomers and dimers (Cosgriff *et al.* 2010). As a precursor to DNA and IM-MS analysis, control spectra were acquired to confirm the gas phase oligomeric state (Fig. 5.6). This is in addition to the denatured protein mass analysis used to determine the experimental CbpA mass for molarity calculations (data not shown). Comparison of the CbpA theoretical and experimental mass values has shown a good agreement between both, with a mass loss of 30 Da (table 5.1). The presence of phosphate bound cations is proposed to account for the DNA mass differences observed.

Native CbpA mass analysis agrees with previous data indicating that the predominant CbpA oligomer is the dimer with observed levels of monomeric species. Unexpectedly tetramer and pentamer oligomers, not previously reported in the literature, were observed at high and low experimental concentrations (8 and 20 μM) (Fig. 5.6 a and b respectively). Comparison of all control spectra acquired over this concentration range indicates no evidence of CbpA trimers.[37] Including the effect of non-specific gas phase oligomerisation to promote the observed higher order oligomers, the lack of observed gas-phase trimer is interesting and may highlight a potential oligomerisation pathway.

[37] Trimer (103.3 kDa) predicted to appear between +25 and +19 (4132 to 5436 *m/z*) using globular protein charge predictors (Kebarle & Verkerk 2009), and comparison with previously observed CSD for all other identified oligomers.

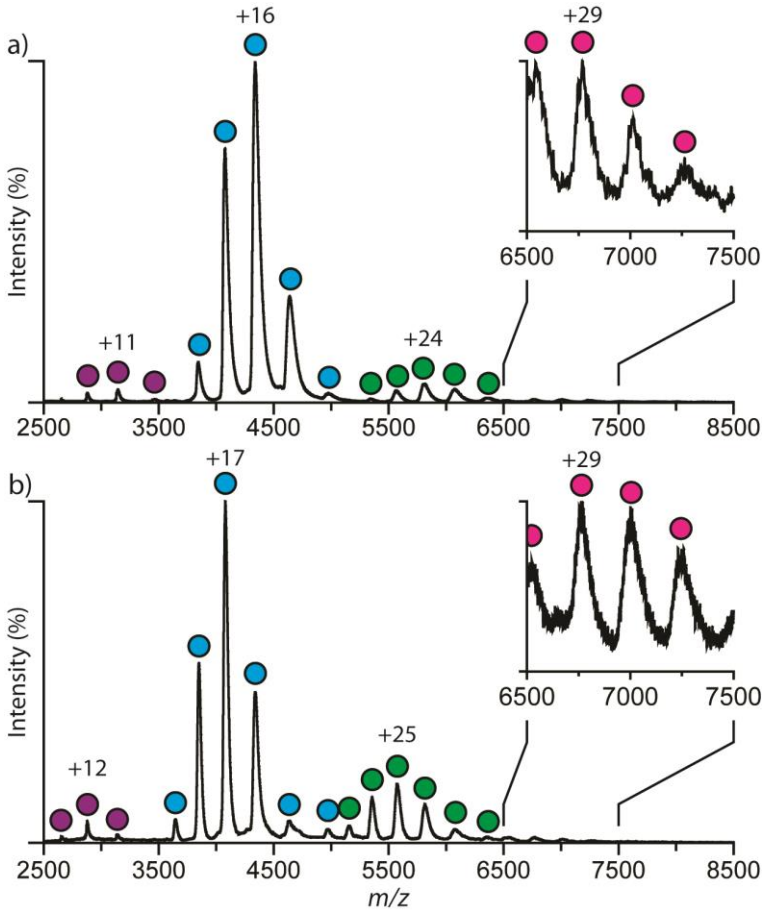

Figure 5.6. Control Mass Spectra of CbpA Showing Previously Unidentified Oligomers

Control spectra of CbpA at 150 mM AmAc pH 7.5, at (a) 8 μM and (b) 20 μM. CbpA oligomers observed: monomer (purple), dimer (blue), tetramer (green), and pentamer (pink).

Table 5.1. Theoretical and Experimentally Calculated Masses of CbpA, and DNA Constructs

Protein / DNA	Length (bp)	Purification Tag	Mol. E, Coeff. (M^{-1} cm^{-1})	Theoretical Mass (Da)	Experimental Mass (Da)	Mass Error ($^{+}/_{-}$ Da)	Mass Difference (Expt. - Theo.)
CbpA	N/A	None	44 920	34 455	34 426.16	0.92	-29.84
50[50]	50	N/A	N/A	30 781	30 838.29	20.39	+57.29
B100	100	N/A	N/A	61 824.02	63 111.32	33.7	+1 287.3
SOA	60	N/A	N/A	37 102.14	37 391.27	41.65	+289.13
SOB	64	N/A	N/A	39 579.62	39 831.64	48.41	+252.02

DNA masses represent values obtained from the mass spectra with the lowest Δm at FWHM, in the presence of AmAc.

5.4.1.1. R^2 Clustered Ion Mobility Analysis of CbpA

Currently no complete structures of either the CbpA monomer or DNA binding dimer have been produced. Applying IM derived CCS constraints, attempts were to be made by Harpal Sahota to produce a computational model of the biologically relevant DNA binding CbpA dimer using the methods previously detailed above (section 5.2.1). Additionally, the computed structure would target the incorporation of the alleged unstructured linker region (amino acids 76-117) lacking in current models.

The methods used here to calculate CCS have based adapted from recent observations by Salbo and co-workers (2012). This work described the requirement for multiple TriWave IM velocity separation parameters to improve the fitting between T-wave calibrants and their experimentally calculated drift tube Ω values. This suitability is defined as a high R^2 agreement between the power fit of the calibration curve and the calibrants required for calculating the CCS of the unknown t_D (Salbo *et al.* 2012).

Salbo and co-workers (2012) presented a good R^2 agreement between eleven native and non-native calibrants at T-wave wave velocities of 200 to 300 m/s, at constant wave amplitudes of 5.8 and 8.5 V respectively (R^2 0.99 and 0.93). Additional increases or decreases of these TriWave velocity and amplitude conditions reduced the agreement between the T-wave and drift tube values, with R^2 values of 0.83 and 0.76, at 100 (3.4 V) and 400 m/s (13 V) respectively. These optimised conditions were used to calibrate the CCS of human insulin oligomers.

Adapting these observations to the investigation presented here, our experimental design incorporated the observed suitable 200 and 300 m/s values, previously used, in addition to a mean value of 250 m/s. Due to the increased mass of the native CbpA dimer (68.9 kDa) compared to native human insulin used for the study above (5.8 kDa monomer, 11.6 kDa dimer), we additionally incorporated additional wave velocity conditions at 350 and 400 m/s. To compensate for the potential reduced R^2 values at the increasing wave velocities, each condition was studied at five different wave amplitude conditions: 7 to 12 V, in 1V increments.

Applying calibrants (section 5.3.3) that closely bracketed the t_D of the lowest charged CbpA monomer and dimer ions, the R^2 of the power fit curve for each of the IM separation conditions are presented in table 5.2. Wave amplitude and velocity conditions values were manually clustered to $R^2 \geq 0.975$, ≥ 0.9 and > 0 to calculate and compare experimental average CCS values.

Of the 25 mobility separation conditions used, 6 prevented the use of power fitted calibration curves (red, table 5.2) required for the CCS calibration of multiply charged species (Shvartsburg & Smith 2008). Conditions unable to produce the power fit calibration curves occur at combinations of low wave velocities and high amplitudes. These therefore prevent suitable mobility separation.

Table 5.2. Ion Mobility Calibration R^2 Values

		Wave Velocity (m/s)				
		200	250	300	350	400
Wave amplitude (V)	8	0.8625	0.975	0.9635	0.9746	0.9646
	9	0	0.9585	0.9805	0.9901	0.9636
	10	0	0.7484	0.9672	0.9736	0.967
	11	0	0	0.9936	0.9778	0.9619
	12	0	0	0.9341	0.9338	0.9659

Above data represents the calibration R^2 values for all IM wave amplitude and velocity conditions obtained. Green values represent mobility conditions that return R^2 values ≥ 0.975. Yellow values represent conditions returning values between $0.975 > 0.9$. Orange values represent calibrations resulting in deviations < 0.9. Conditions unable to produce calibration curves that permitted power fitting, required for multiply charged species, are shown in red.

IM analysis of the lowest charged CbpA dimer +15 ion (*m/z* 4591.2) supports the existence of two defined structural conformations (compact and extended) (Fig. 5.7c). In comparison, only a single conformation is observed for the lowest charged monomer +11 ion (*m/z* 3133.3) (Fig. 5.7b). Comparisons of the average CCS values calculated using these conformations, and clustered according to R^2 values are presented in table 5.3. Accounting for MSD calibrated CCS values indicates no change in conformation accuracy across the R^2 range studied.

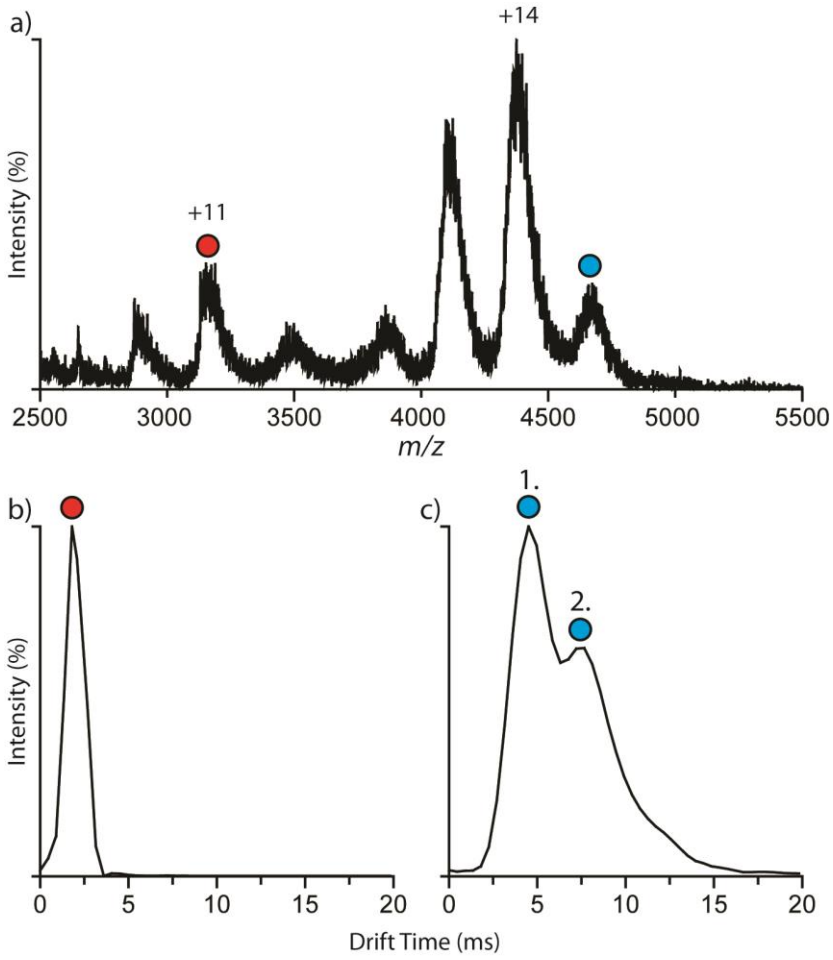

Figure 5.7. Ion Mobility Analysis of CbpA

a) Data represents extracted t_D values for lowest charged monomer (red, b), and dimer (blue, c) ions, at 250 m/s and 9 V. c) 1 and 2 represent the t_D of the compact and extended dimer conformations respectively. CCS values are determined from the FWHM of the lowest charged ions (monomer; +11, and dimer; +15), highlighted in the mass spectrum. AmAc concentration; 150 mM, pH 7.5. Samples previously stored in glycerol.

Table 5.3. Collision Cross Sections of Calibrated CbpA the Lowest charged Monomer and Dimer

Conformation	**$R^2 \geq 0.975$**	**$R^2 \geq 0.9$**	**$R^2 > 0$**
Monomer	2542.38 ($^{+}$/- 77.13) Å^2	2590.8 ($^{+}$/- 69.4) Å^2	2588.13 ($^{+}$/- 68.1) Å^2
Compact Dimer	4128.21 ($^{+}$/- 97.6) Å^2	4076.4 ($^{+}$/- 119.76) Å^2	4081.34 ($^{+}$/- 125.31) Å^2
Extended Dimer	4475.65 ($^{+}$/- 195.88) Å^2	4510.83 ($^{+}$/- 151.91) Å^2	4543.05 ($^{+}$/- 152.53) Å^2

5.4.2. CbpA DNA Binding

Unless induced under the control of the Lrp operon, CbpA expression is strongly upregulated by σ^S during the stationary phase (Yamashino *et al.* 2005). These expression events produce 3 000 to 15 000 copies of CbpA per cell, promoting compaction of the *E. coli* genome. CbpA has been shown to bind DNA promoting the rapid (2 to 5 minutes) formation of aggregates greater than 60 nm in size (Cosgriff *et al.* 2010). The organisation of CbpA and DNA within these aggregates is poorly understood. Therefore, the methods and results presented here outline the various solution conditions and DNA constructs that have been applied to study these events using native mass spectrometric methods.

5.4.2.1. Incubations of CbpA with a 50 Base Pair *De Novo* DNA Construct

CbpA binds DNA non-specifically (Cosgriff *et al.* 2010). In the absence of binding sequence preferences, initial investigations were performed using a designed *de novo* DNA construct. Parameters for sequence design were a 50 bp construct length, encoding a 50 % GC content. The 50 bp construct length exceeded the proposed 40 bp minimum length of DNA required for CbpA-DNA binding (Gur, Katz & Ron 2005). This construct was identified as 50[50], with forward and reverse complementary ssDNA sequences presented in table 5.4. A control mass spectrum of 50[50] is shown in figure 5.8.

Initial incubations of CbpA and 50[50] were analysed over a range of incubation ratios, from 4:1 to 1:8. Samples were buffer exchanged into 250 mM AmAc pH 7.5, and diluted to 13 μM CbpA concentration prior to addition of DNA and incubation. CbpA: 50[50] samples were incubated at room temperature and monitored at non-uniform time points over the course of three-hours. Mass spectra were acquired over a *m/z* range of 1 000 to 16 000 for two minutes using two second scans.

Across all incubation ratios studied, data indicated a reduction of CbpA relative to the DNA construct over the stated period of time (Fig. 5.9). At all incubation conditions and time points no CbpA: 50[50] oligomers were observed.

Table 5.4. Sequence Data of the 50[50] and B100 DNA Constructs

Construct	T_m (°C)	GC Content (%)	Forward or Reverse Compliment	ssDNA Sequence (5' to 3')
50[50]	94.6	50	Forward	CACGAGTGAATGAGACCTAAGTGACGACG AGAGTATGAGACCTGAGAGAG
			Reverse	CTCTCTCAGGTCTCATACTCTCGTCGTCACT TAGGTCTCATTCACTCGTG
B100	99.1	50	Forward	(GAAGAGAAGC)*10
			Reverse	(GCTTCTCTTC)*10

T_m values represent salt adjusted values; 50[50] at 250 mM AmAc, B100 at 150 mM.

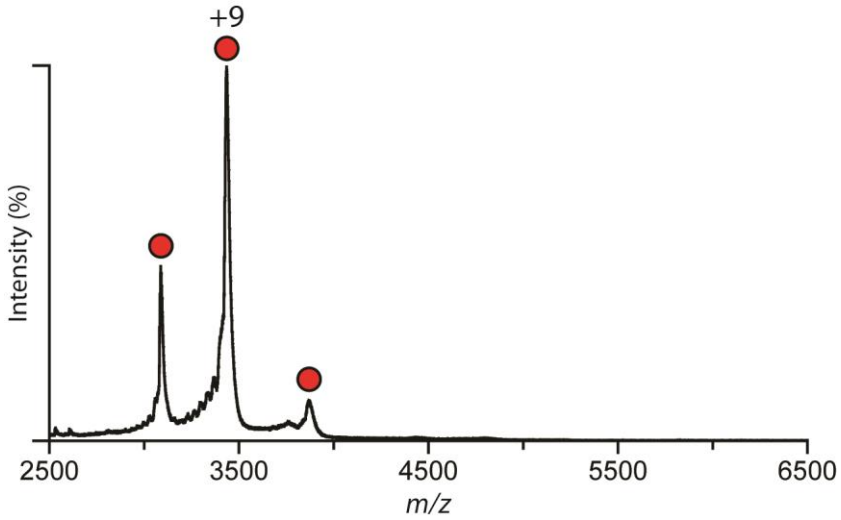

Figure 5.8. Control Mass Spectrum of the *De novo* 50[50] dsDNA Construct

Control analysis of the 50[50] dsDNA construct (red). AmAc concentration; 250 mM, pH 7.5.

Due to the failure of observing CbpA-DNA oligomers using these conditions the experimental limitations were examined. Two conditions were identified to be potentially limiting the promotion of observable protein-DNA oligomers. The first identified limitation was that the ratio of protein to DNA was too narrow. Therefore it was suggested that this range should be expanded to incorporate a broader incubation range (e.g. 20:1 to 1:20). The second identified limitation was the concentration of the AmAc buffer. It was

suggested that this might be too high (250 mM AmAc) and therefore inhibiting binding interactions. Studies progressed to study CbpA-DNA binding using a reduced AmAc buffer concentration of 150 mM (pH 7.5).

It is known that magnesium cations stabilise the phosphate backbone of dsDNA (Hartwig 2001). Therefore, in addition to the reduced concentration the AmAc, a source of magnesium cations was added to the buffer. The use of magnesium cations are additionally supported by recent work presented by Guéroult and co-workers (2012). This work showed that magnesium cations bound within the major groove stabilise protein-DNA complexes by "limiting the energetic and entropic costs of protein binding" (Guéroult *et al.* 2012).

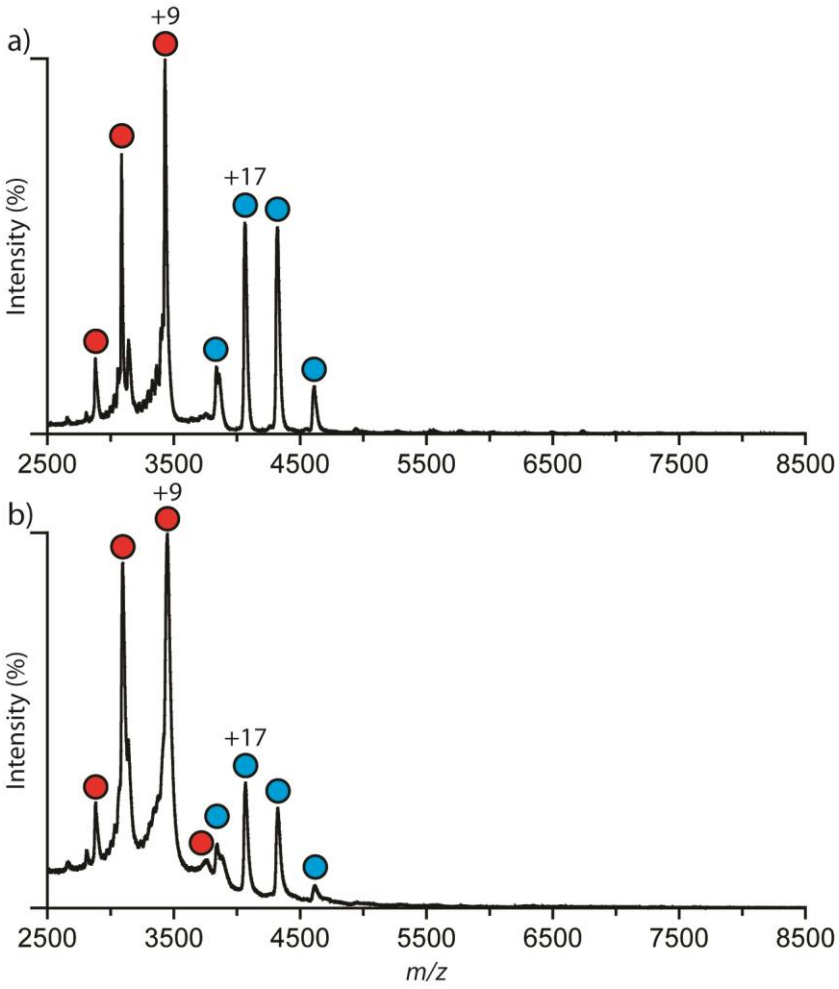

Figure 5.9. 4:1 Incubation of CbpA: 50[50]

CbpA dimer (blue) in the presence of 50[50] (red) (4:1 ratio) (a) before and (b) after incubation (one hour, room temperature). Although no CbpA: 50[50] oligomers are observed, a reduction in CbpA dimers relative to DNA between time zero and one hour are noted. AmAc concentration; 250 mM, pH 7.5.

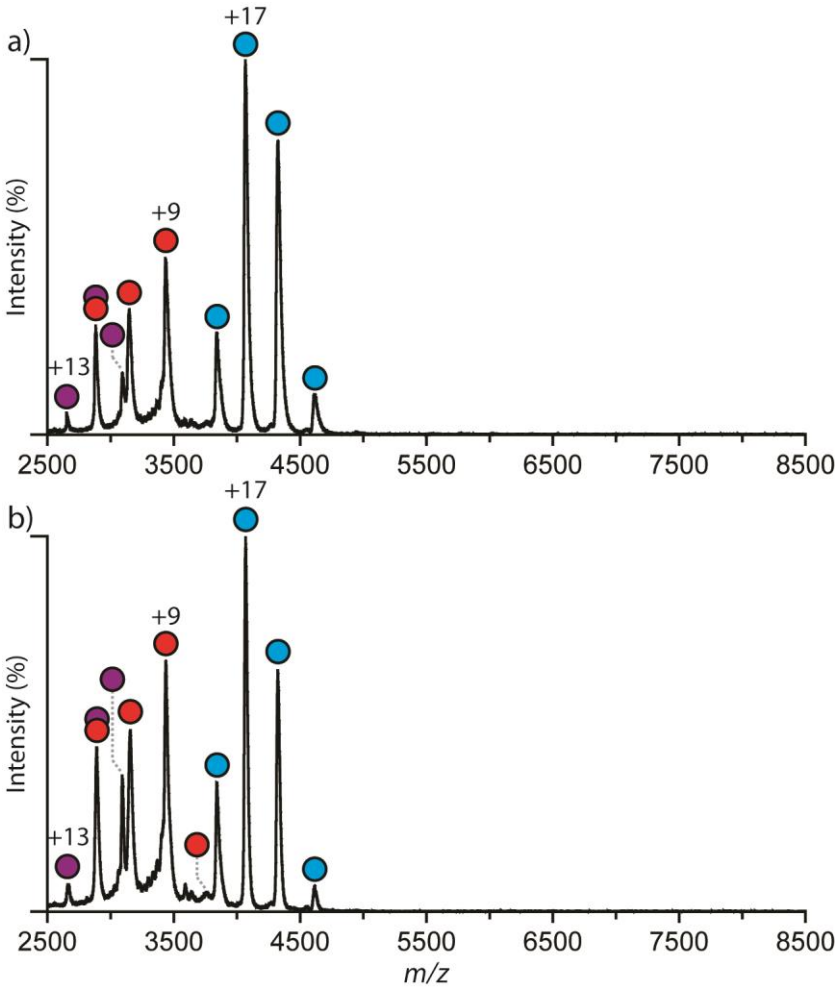

Figure 5.10. 1:1:4 Incubated CbpA: 50[50]: $MgCl_2$

1:1:4 incubation of CbpA (monomer = purple, dimer = blue), 50[50] (red) and $MgCl_2$, (a) before and (b) after incubation (one hour, room temperature). Although no CbpA: 50[50] oligomers are observed, a reduction in CbpA oligomers relative to DNA between time zero and one hour are noted. AmAc concentration; 150 mM, pH 7.5.

Lacking any indication of an optimal ratio of cations relative to DNA, initial work added a small content of $MgCl_2$, equal to four times the concentration of DNA, to samples prior to incubation at room temperature. Buffer exchanged (150 mM AmAc, pH 7.5) incubations of CbpA: 50[50]: $MgCl_2$ were studied using ratios of 1:1:2 and 1:2:4 (CbpA 11 μM). Mass spectra were acquired over a *m/z* range of 1 000 to 16 000, for two minutes using two second scans. Despite apparent reductions in CbpA relative to dsDNA, no CbpA: 50[50] oligomers were observed (Fig. 5.10).

5.4.2.2. Incubations of CbpA with a 100 Base Pair B-Form DNA Construct

Despite failing to study broader incubation ratios of CbpA: DNA, both with and without a source of magnesium cations, concerns were raised regarding the length of the dsDNA construct used. These concerns were the close relationship between the 50 bp length of the 50[50] construct used above, and the 40 bp minimum length CbpA-DNA binding proposed by Gur and co-workers (2005). It was hypothesised that a DNA construct of increased length may promote CbpA-DNA oligomers that can be resolved using native mass spectrometry methods.

Follow up CbpA-DNA binding experiments incorporated a readily available 100 bp dsDNA construct identified as B100 (B-form, 100 bp dsDNA). This 100 bp dsDNA construct was previously adapted from Gyi *et al.* (1998) for IM-MS studies by Jun Yan under the supervision of Dr Adam McKay (work unpublished). Structurally B100 encodes a 10x10 bp sequence repeat, which forms the biologically relevant B-form[38] dsDNA helix. Compatible with the 50[50] sequence previously studied this B100 construct encodes a 50 % GC content. Forward and reverse complementary sequences are presented in table 5.4. Control mass analysis of B100 is presented in figure 5.11.

Analysis of CbpA: B100 targeted the recreation of sample and incubation conditions that had been previously used to successfully promote CbpA-DNA oligomerisation. Work by Azam & Ishihama (1999) outlined suitable buffer and incubation conditions for analysis that had been previously applied to calculate CbpA-DNA K_d values using DNA EMSA.

Using the methods presented by Azam & Ishihama (1999), CbpA and B100 were incubated in 10:1 ratio and buffer exchanged into 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl and 1 mM DTT. This sample was incubated for 25 minutes at 25 ˚C prior to buffer exchange into 100 mM AmAc (pH 7.5), equal to the NaCl concentration above. Buffer exchange was performed by five sequential washes using a single Amicon 1.5 ml 10 kDa MWCO centrifugal concentrator column (Millipore, Dundee, UK). Mass spectra were acquired over *m/z* range of 1 000 to 32 000 for two minutes using two second scans.

[38] Watson - Crick model of DNA. Right hand rotation, 10.5 bp/turn that increases 3.32 Å in length per turn.

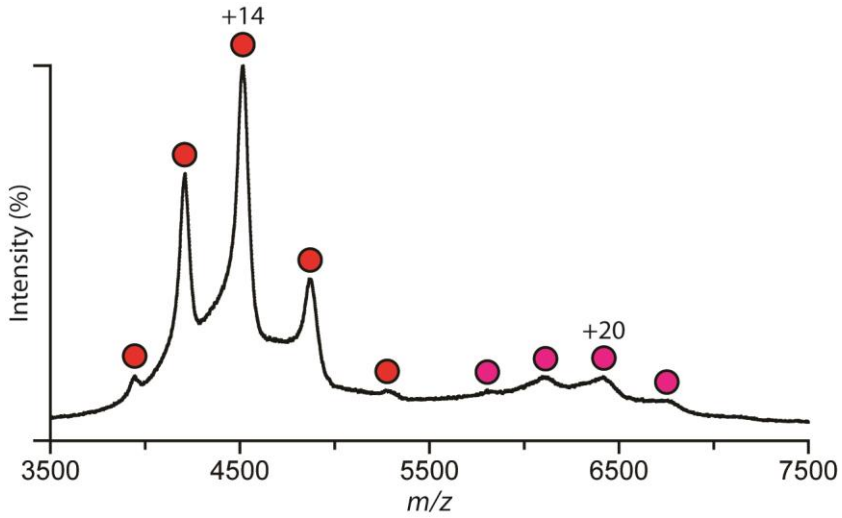

Figure 5.11. Control Mass Spectrum of the B100 dsDNA Construct

Control analysis of the B100 dsDNA construct adapted from Gyi *et al.* (1998). The monomeric dsDNA construct highlighted in red. The repetitive nature of the forward and reverse complement ssDNA allows misaligned base pair matching of these complementary sequences and is used to account for the unexpected mass corresponding to $B100_2$ (pink). These misalignments are proposed to account for the poor spectra baseline resolution by enabling increased salt adductation, and are supported by a large increase in the experimental mass values compared to the theoretical (table 5.1). AmAc concentration; 150 mM, pH 7.5.

Mass spectrometric analysis of the CbpA: B100 sample after incubation are presented in figure 5.12. Despite a nanodrop[39] calculated dimer concentration of 30 μM (50 μM expected), exhibiting an excess in protein to DNA (DNA = 6.5 μM, therefore final ratio = 9:2), no CbpA oligomers were observed.

Experimental observations by Azam & Ishihama (1999) noted a co-operative mechanism of CbpA-DNA binding. Applying these observations, DNA should remain observable in the absence of unbound protein, despite an excess of CbpA to B100. To confirm this suggestion, using the same conditions as above, B100 was incubated with a protein exhibiting no known DNA binding affinity for control analysis. It was hypothesised that if this protein was observed within mass spectra, post incubation and buffer exchange, these data would support that CbpA-DNA binding was occurring. Therefore, *S. cerevisiae* ADH (Sigma-Aldrich Ltd, Dorset, UK) was incubated with B100, at a ratio of 10:1.

[39] Nanodrop used in lieu of the Qubit method due to the presence of DNA.

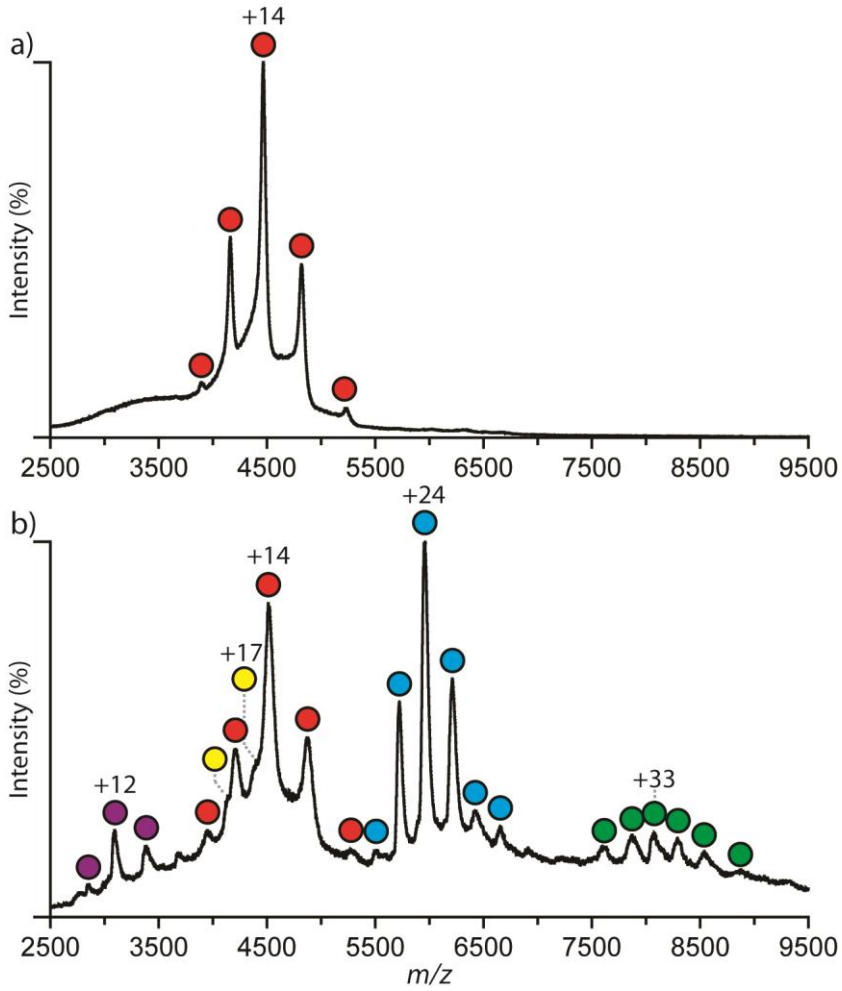

Figure 5.12. Incubation of CbpA: B100 Using the Azam & Ishihama Method

a) CbpA incubated in the presence of B100 (red) (experimentally calculated ratio of 9:2). Incubation of protein and DNA performed according to the method presented in Azam & Ishihama (1999) (see main body of text), prior to buffer exchange into 100 mM AmAc (pH 7.5). No protein or protein-DNA oligomers are observed. b) Repetition of the incubation using ADH (Monomer; purple, Dimer; Yellow, Tetramer; blue, Octamer; Green) in lieu of CbpA. As expected no ADH: B100 oligomers are observed.

Post incubation and buffer exchange, nanodrop analysis showed consistent experimental concentrations between incubations of B100 with CbpA or ADH. Mass spectrometric analysis (Fig. 5.12b) showed that both ADH and B100 are observed within the gas phase. As no ADH: B100 oligomers are observed, and the similar concentrations of ADH/CbpA between experiments, analysis supports that the solution phase conditions are suitable to stabilise CbpA-DNA binding. Concluding these observations with the positive feedback mechanism of CbpA-DNA binding (Azam & Ishihama 1999) a large CbpA

excess is therefore likely to be required to observe oligomers of unknown complexity. In contrast, if no CbpA-DNA binding events were observed we would expect comparable relative intensities of CbpA compared to the incubations containing ADH.

Considering the large-scale CbpA-DNA interactions, a collision voltage ramp was performed on the original CbpA: B100 sample. Using this ramp, attempts were made to induce collisional decomposition of any unobserved/low intensity gas phase complexes to study their resultant dissociation products.

At a TriWave Trap (see schematic in Fig. 1.13) collision voltage of 100 V (Transfer; 25 V) (Fig. 5.13), fragments with masses correlating to CbpA monomers and dimers were observed. At these high collision voltages however, no masses correlating to ssDNA or dsDNA are observed, despite observations of the latter prior to dissociation studies. These observations are proposed to occur due to the weaker covalent nature of ester bonds compared to the peptide bond. These conclusions are supported by previous CID observation by Kjeldsen & Zubarev (2011). Therefore, if these oligomers are dissociation products from the protein-DNA oligomers, these data support the prerequisite for CbpA dimers in the protein-DNA binding interaction.

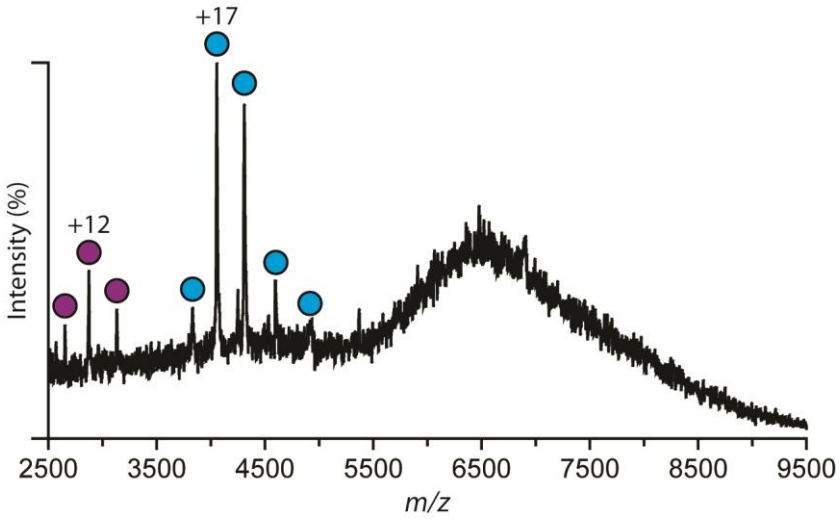

Figure 5.13. Collision Voltage Ramp of Incubated CbpA: B100

Collisional dissociation condition (Trap 100 V, Transfer 25 V, not quadrupole isolated) analysis of incubated CbpA: B100 sample (same sample as Fig. 5.11a). Spectrum confirms that an observable concentration of CbpA remains within the sample (solution concentration = 30 μM). Observed ions are hypothesised to be dissociation product from CbpA-B100 oligomers (monomer = purple, dimer = blue).

5.4.2.3. Incubations of CbpA with a 62 Base Pair Curved DNA Construct

Due to the repetitive nature of the B100 forward and reverse complements (10x10 bp) concerns were raised regarding the ability of these sequences to promote forward and reverse ssDNA sequence misalignments. This is supported by a sequential dilution of B100 studied using 2 % PAGE analysis (Fig. 5.14). PAGE analysis indicates that misalignments are able to promote a strong heterogeneous mixture of ssDNA and dsDNA elements along a single forward/reverse complement, in addition to a non-uniform construct length. Despite a level of support for CbpA: B100 binding events, in the absence of evidence supporting CbpA: ssDNA binding events work progressed to study binding using a third dsDNA construct.

Figure 5.14. 2 % Polyacrylamide Gel Electrophoresis Analysis of the B100 Construct
Sequential dilution of B100 from 100 % to 8 % of the original concentration (22.3 μM) analysed using 2 % PAGE and stained using dilute ethidium bromide solution (< 0.1 mg/ml). Comparison with a 100 bp ladder (New England Biolabs, MA, USA) indicates heavy level of smearing consistent with misaligned base pairing. Intense banding at 100 bp however supports that a relatively high concentration of correctly annealed forward reverse complements are formed, and is supported by mass spectrometric analysis presented here (Fig. 5.11).

As a consequence of the above conclusions a literature search was initiated to obtain sequence data for a suitable dsDNA construct that had been previously shown to allow

CbpA binding. This literature search identified three articles: Azam & Ishihama (1999), Cosgriff *et al.* (2010), and Gur, Katz & Ron (2005). The dsDNA constructs applied by these studies are outlined in brief below, with sequence information presented (when available) in table 5.5.

Table 5.5. Known CbpA Binding DNA Constructs

Construct	Length (bp)	T_m (°C)	GC Content (%)	Forward or Reverse Compliment	ssDNA Sequence (5' to 3')
SOA	62	89.6	40	Forward	GGCAAAAAACGGCAAAAAACGGCAAAAAACCGCAAAAAACCGCAAAAAAGCGCAAAAAAG
				Reverse	CTTTTTTGCGCTTTTTTGCGGTTTTTTGCGGTTTTTTGCCGTTTTTTGCCGTTTTTTGCC
SOB	64	94.5	50	Forward	GGGGATCGCACGATCTGTATACTTATTTGCCCGGGGATCGCACGATCTGTATACTTATTTGCCC
				Reverse	GGGCAAATAAGTATACAGATCGTGCGATCCCCGGGCAAATAAGTATACAGATCGTGCGATCCCC
EL50	50	90.1	48	Forward	GCCGAGGATCCAATTTGCTGTCACTCCCATTTAAACGCTAGATGCTGTCA*
				Reverse	TGACAGCATCTAGCGTTTAAATGGGAGTGACAGCAAATTGGATCCTCGGC

T_m values represent salt adjusted values at 150 mM AmAc. * represents γ-32P end labelling.

SOA and SOB sequences taken from Azam & Ishihama (1999), EL50 taken from Gur, Katz & Ron (2005)

Azam & Ishihama (1999), studied CbpA-DNA binding using two dsDNA constructs, with known curved and linear structures, identified as SOA and SOB (table 5.5).

These constructs were 62 and 64 bp in length, with calculated K_d values of 122 and 175 nM respectively.

Cosgriff *et al.* (2010) performed CbpA-DNA binding studies using an unspecified length of dsDNA, isolated from plasmids using restriction nucleases. Although biologically relevant, production of a suitable stock concentration for analysis was deemed unsuitable due to the calculated low yield and sample purity.

Gur, Katz & Ron (2005) observed CbpA-DNA binding using a single end labelled (γ-^{32}P-ATP) 50 bp dsDNA construct (EL50, table 5.5), of unspecified structural conformation.

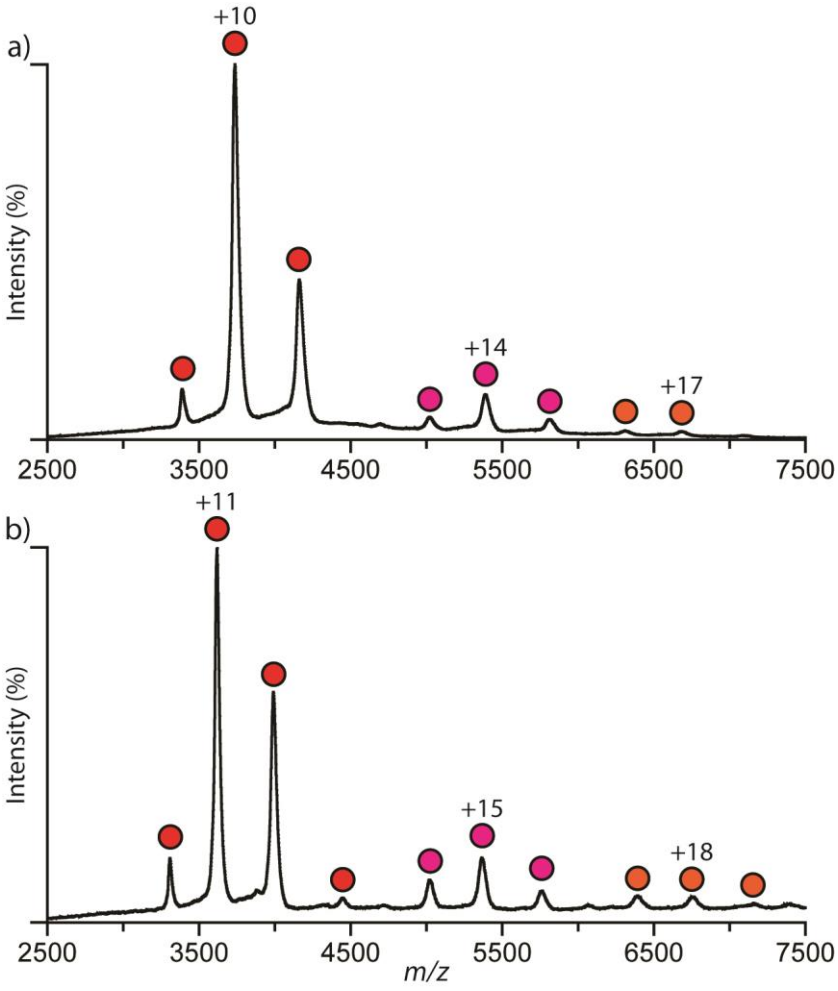

Figure 5.15. Control Mass Spectra Analysis of SOA and SOB dsDNA Constructs

Control analysis of the (a) SOA and (b) SOB dsDNA construct. Monomeric dsDNA are highlighted in red, along with unexpected masses corresponding to dimeric (pink) and trimeric (orange) SOA/SOB constructs. AmAc concentration; 150 mM, pH 7.5.

As the constructs studied by Azam & Ishihama (1999) would allow direct comparison between CbpA curved and non-curved DNA binding, both SOA and SOB were

purchased without sequence modifications. Control mass analysis of both SOA and SOB are presented in figure 5.15.

Combining all previous observations the refined experimental conditions supported incubation of a large excess of CbpA to DNA due to co-operative mode of oligomerisation. Additional information further supported the use of a magnesium cation source to stabilise the DNA backbone (Hartwig 2001) and protein-DNA oligomers (Guéroult *et al.* 2012). Subsequent literature searches expanded to support the requirement of one magnesium cation per two phosphate groups (Sander & Ts'o 1971).

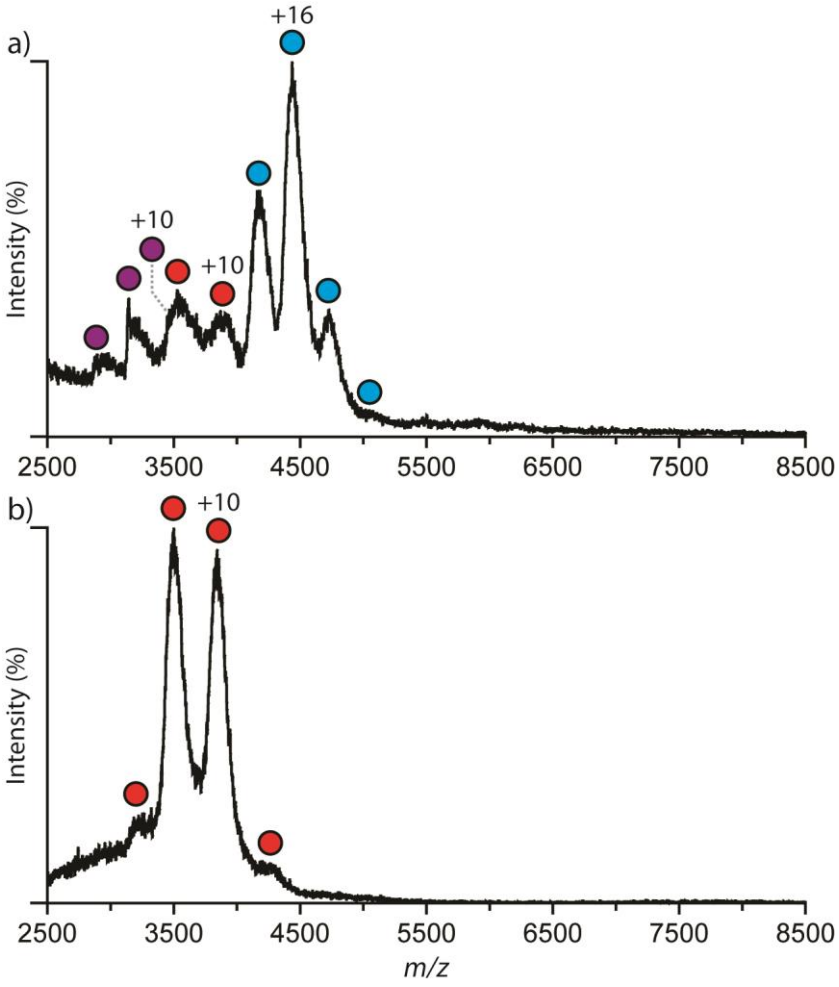

Figure 5.16. Incubation of CbpA: SOA at a Ratio of 10:1

CbpA: SOA: $MgCl_2$ (10:1:62) (a) before and (b) after one hour incubation at 37 ˚C. Data shows that one hour incubation is suitable to promote CbpA (monomer; purple, dimer; red) binding of DNA (SOA; Red), with the loss of all resolvable protein. AmAc concentration; 150 mM, pH 7.5. CbpA samples previously stored with glycerol.

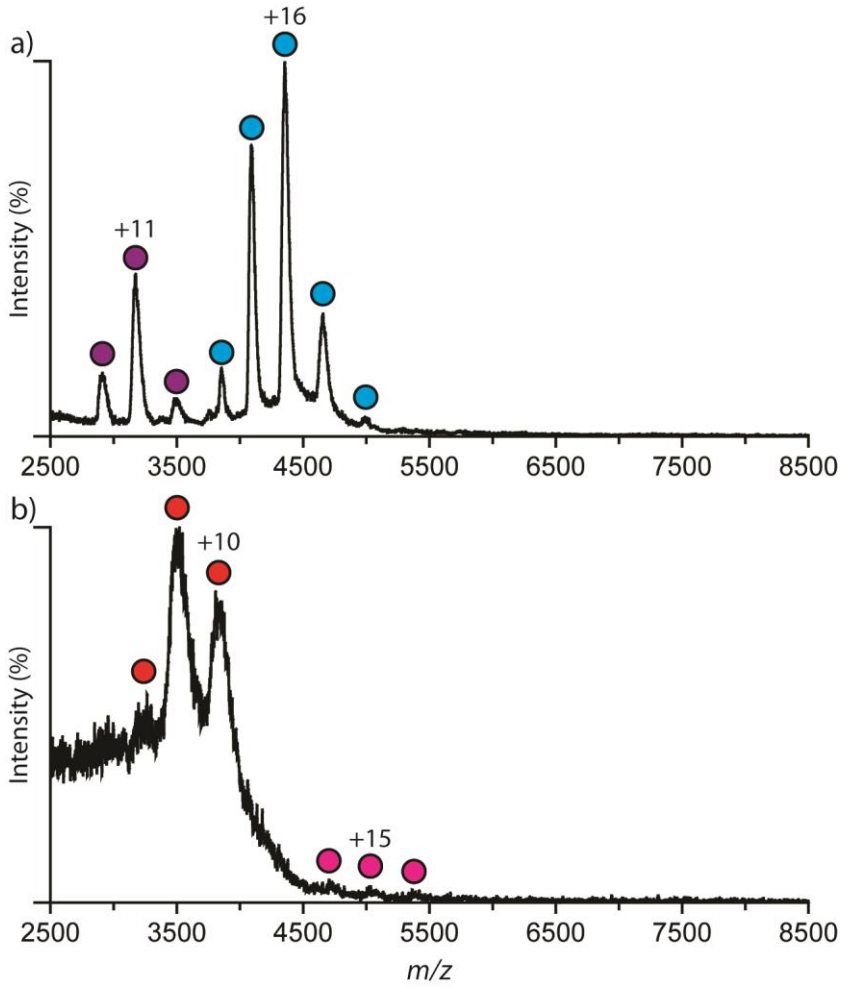

Figure 5.17. Incubation of CbpA: SOA at a Ratio of 20:1

CbpA: SOA: $MgCl_2$ (20:1:62) (a) before and (b) after seventeen hour incubation at 37 ˚C. Despite the large excess of CbpA (monomer = purple, dimer =red) and no resolvable SOA (red, and SOA_2; purple) at the start of incubation, suitable incubation time allows all CbpA oligomers to bind. This is supported by the observed excess of DNA within mass spectra after incubation. Acquisitions at one-hour further support the binding of all CbpA oligomers, these data are not shown due to low signal to noise ratio. AmAc concentration; 150 mM, pH 7.5. CbpA samples previously stored with glycerol.

CbpA and SOA were individually buffer exchanged into 150 mM AmAc (pH 7.5) prior to mixing at the required ratio with $Mg(CH_3COO)_2$.[40] CbpA: SOA: $Mg(CH_3COO)_2$ samples were incubated at a temperature of 37 ˚C, at two different sample ratios: 10:1:62 (Fig. 5.16), and 20:1:62 (Fig. 5.17). Mass spectra were acquired at zero, one and seventeen

[40] $Mg(CH3COO)_2$ was used in lieu of $MgCl_2$, used previously, due to the proposed increased suitability of the acetate anion for mass spectrometric analysis.

hours over a *m/z* range of 1 000 to 25 000, for two to ten minutes using two second scans. Despite the low signal to noise ratio of all spectra, no CbpA: SOA oligomers were observed, at any time point, in either sample incubation ratio.

As analysis using native mass spectrometric methods had been unable to resolve any protein-DNA oligomers using a construct known to bind CbpA, DNA-EMSA was applied to study if the conditions used were suitable to promote oligomerisation. Two 10:1 and 20:1 sample sets frozen at 1 and 17 hours (stored at -20 ˚C) previously studied using mass spectrometry were analysed. These samples were compared against DNA control samples at concentrations equal to the 10:1 and 20:1 protein-DNA samples, used to observe changes in the levels of unbound DNA.

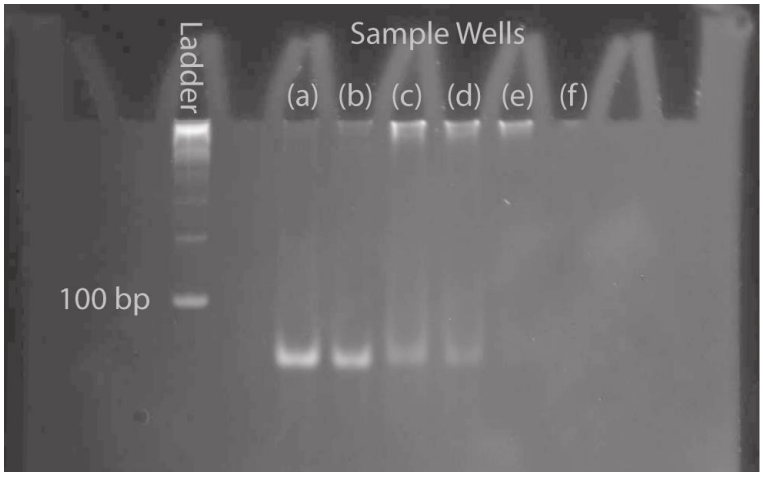

Figure 5.18. Protein-DNA Electrophoretic Mobility Shift Assay Using 6 % DNA Retardation Gel

DNA EMSA of incubated CbpA: SOA, measured against a 100 bp ladder. SOA controls representative of concentration equal to (a) 10:1 and (b) 20:1 dilutions with CbpA. e and c) 1 and 17 hour incubations, respectively, of 10:1:62 CbpA: SOA: $Mg(CH_3COO)_2$. f and d) 1 and 17 hour incubations, respectively, of 10:1:62 CbpA: SOA: $Mg(CH_3COO)_2$. Starting CbpA concentration in all wells 7.65 $^{+}/_{-}$ 0.68 μM. Samples incubated for seventeen hours (c and d) show higher levels of unbound DNA (increased intensity banding), indicating decomposition of oligomers between this time point and 1 hour.

Analysis of the stained DNA-EMSA gel (Fig. 5.18) supports that the AmAc buffer required for electrospray ionisation is suitable for CbpA: SOA binding. This is supported by the majority of observable DNA withheld at the top of the gel lane indicative of protein-DNA binding. Further analysis of solutions recovered from the gel well supports the existence of potential large protein complexes unable to migrate into the gel

(0.5 to 0.85 μM) despite clear evidence of unbound DNA. These replicate similar observation made by both Azam & Ishihama (1999) and Cosgriff *et al.* (2010).

Resulting from the combined limitations of native mass spectrometric methods to study intact CbpA: DNA interactions and the potential mass of CbpA: DNA oligomer, the experimental work was suspended whilst experimental redesign was performed. Due to time constraints however, no further experimental work was performed.

5.5. Discussion

Although failing to observe CbpA-DNA binding events, data presented here has provided potential insights into the mode of protein-protein and protein-DNA oligomerisation mechanisms. This is in addition to providing constraints for computational modelling approaches.

5.5.1. Validity of R^2 Refined CbpA CCS Values

Work recently published by Salbo and co-workers (2012) has presented a logical step in the optimisation of protocols for producing IM calibration curves. By applying T-wave calibration conditions optimised by their R^2 values allows for increased accuracy CCS calculations of the analyte by reducing deviations that modify the curve distribution. Compared with the work presented by Salbo and co-workers the limiting factor of our investigation may be stated to be the number of standards applied to calculate the calibration curve.

Across the R^2 clustering range studied (table 5.3), accounting for MSD we observe no differences in CCS. This absence of CCS changes is argued here to occur as a function of calibration methodology, by removing calibrant ions that deviate from published Ω values by $^{+}/_{-}$ 5 %. Data presented here indicates that removal of these outlying points improves CCS replication by reducing deviations of the calibration curve that would otherwise affect this value. This may represent a more suitable method for calculating CCS allowing analysis of a heterogeneous mixture of ions. The use of optimised mobility

conditions to reduce the removal of these outlying points, whilst improving R^2, should represent increased CCS accuracy.

IM analysis has indicated the presence of two stable structural conformations for the CbpA lowest charged dimer ion. If these two conformations are biologically relevant their role in the activity of CbpA remains unresolved. Considering the nature of the alleged unstructured domain (amino acids 76-117) that links the J-domain to the two CTD domains however, it is hypothesised here that these two conformations represent open and compact states. This conclusion is additionally supported by the single structural conformation of the CbpA monomer (Fig. 5.7b). The function of these conformational species however has yet to be concluded but may indicate potential internal regulation of protein-DNA and protein-protein binding events.

5.5.2. Studies of CbpA-DNA Binding

Despite failing to observe CbpA-DNA oligomers using a range of DNA constructs and conditions, certain conclusions can be made regarding binding events. One of the largest considerations is that per single 62 bp length of DNA (SOA) greater than twenty CbpA dimers can hypothetically bind. This is supported by the absence of observable CbpA in mass spectra post incubation at both 10:1 and 20:1 solution ratios, whilst retaining resolvable DNA charge states (Fig. 5.17). This conclusion is additionally supported by the proposed method of co-operative binding first presented by Azam & Ishihama (1999).

These data however, are unable to resolve how CbpA dimers stack on the DNA, nor if they are able to stack sequentially on top of each other. AFM studies of linear (non-plasmid) DNA have shown that CbpA: DNA aggregates can contain multiple DNA strands (Cosgriff *et al.* 2010), and suggests that cross-linking between CbpA dimers is possible.

How CbpA cross-links DNA is again unknown, but may indicate the function of the observed tetrameric and pentameric oligomers in control analysis (Fig. 5.6). If these oligomers are native conformations, this would potentially support a second mode of dimerisation to link neighbouring DNA bound dimers. Orientation of these CbpA dimers

to form a plausible secondary dimerisation interface, along with the amino acids that mediate these interactions still to be cited within the literature.

5.6. Future Directions

5.6.1. Studying CbpA-DNA Interactions within the Gas Phase

Considering the challenging nature of observing intact CbpA-DNA oligomers using native mass spectrometric methods, it is suggested that future work progresses to use proteomics based structural tools more typical of aggregating protein studies.

Hydrogen/deuterium exchange is a mass spectrometric method currently applied to the study of aggregating proteins such as the structurally ordered amyloid fibril. In studies such as these, the sample is incubated in a deuterated solvent for a specific period of time to allow hydrogen-deuterium exchange. Samples are subsequently buffer exchanged to remove excess deuterium, and analysed to compare the difference in deuteration between a control (e.g. monomer) and the aggregate using proteolytic methods. Comparison of hydrogen/deuterium mass shift indicates the level of solvent accessible area at the residue level, indicative of bound and unbound subunits (Ashcroft 2010).

Despite providing information regarding the solvent accessible/inaccessible areas, hydrogen/deuterium exchange fails to indicate the local environment of the binding interface, including orientation of neighbour oligomers and their respective amino acids in the absence of known structures. Using these structural data hydrogen/deuterium exchanges can be mapped onto X-ray crystal, NMR or computational structural models to calculate the environment (Jaswal 2013). This therefore currently represents a limiting step for structural analysis of CbpA-DNA oligomers until computational modelling of the CbpA dimer is complete.

In comparison, cross-linking connects two functional amino acid groups within a linker-derived distance allowing the study of the local environment relative to the initial binding group. Applying a bottom-up proteomics approach[41] for liquid chromatography mass spectrometry and MS/MS analysis, cross-linking enables the construction of a three

[41] In contrast to other mass spectrometry instrumentation, Fourier transform-ion cyclotron resonance can be also applied to top-down cross-linking studies (Sinz 2006).

dimensional model from peptide sequence data (Sinz 2006). This is in addition to contributions from IM-MS methods to separate linker bound peptides using differences in CCS (Preston *et al.* 2012).

Protein-DNA cross-linking, although not as regularly discussed in the literature as studies of protein-protein interactions, has been previously performed. Known protein-DNA chemical cross-linkers include, 1,2,3,4-Diepoxybutane. 1,2,3,4-Diepoxybutane promotes the alkylation of adenine and guanine bases, producing 2-hydroxy-3,4-epoxybut-1-yl monoadducts. These monoadducts are subsequently able to react with nucleophilic amino acid side chains, such as those present in serine, lysine and tyrosine producing protein-DNA cross-links. Additionally 2-hydroxy-3,4-epoxybut-1-yl monoadducts are able to cross-link DNA through further hydrolysis and alkylation reactions (Gherezghiher *et al.* 2013).

In addition to the above, 5-iodouracil labelled ssDNA has been previously used to cross-link DNA to proteins through UV photo catalysed reactions with aromatic amino acid side chains. Under UV catalytic conditions 5-iodouracil is cleaved at the iodine-carbon bond producing a reactive vinyl radical. This vinyl radical binds preferentially to aromatic rings such as those found in tryptophan residues, producing a carbon-carbon bond that cross-links the DNA to protein (Steen *et al.* 2001).

The above represent two of only a handful of protein-DNA cross linking methods currently available which are stable under both the MS/MS, and proteolytic cleavage conditions required for analysis. In contrast, studies of CbpA-CbpA oligomerisation in the presence of DNA, using protein-protein cross-linking methods, will allow for a more in depth structural analysis due to the range and specificities of protein cross-linkers available.[42]

5.6.2. Studying CbpM Regulated CbpA Activity

Although CbpA activity is regulated by CbpM, through its interactions with the J-domain of the former, the native stoichiometry CbpA: CbpM oligomer has not been fully resolved.

[42] For a larger overarching review on the topic of protein-protein cross linking studies the author directs the reader to the following review, Sinz (2006)

Forming part of the original project aims (section 5.2) these studies would probe CbpA: CbpM binding interactions, stability, and stoichiometry.

Additional interest is the potential J-domain regulated CbpA oligomer stability, highlighted previously (section 5.1.4) (Chintakayala & Grainger 2011). If CbpM binding to this domain promotes oligomer dissociation, this will support a model that inhibits DNA interactions through removal of the dimer dependent binding interface. In contrast, if CbpA dimerisation remains stable, this would infer that CbpM inhibits dsDNA binding by restricting access to dimeric CTD I domain.

5.7. References

Almirón, M., *et al.*, 1992 "A Novel DNA-Binding Protein with Regulatory and Protective Roles in Starved Escherichia coli," Genes and Development, Vol.6 pp.2646-2654

Arbely, E., *et al.*, 2011 "Acetylation of Lysine 120 of p53 Endows DNA-Binding Specificity at Effective Physiological Salt Concentration," Proceedings of the National Academy of Sciences, Vol.108 pp.8251-8256

Ashcroft, A., 2010 "Mass Spectrometry and the Amyloid Problem-How Far can we go in the Gas Phase?" Journal of the American Society for Mass Spectrometry, Vol.21 pp.1087-1097

Atmanene, C., *et al.*, 2010 "Combination of Noncovalent Mass Spectrometry and Travelling Wave Ion Mobility Spectrometry Reveals Sugar-Induced Conformational Changes of Central Glycolytic Genes Repressor/DNA Complex," Analytical Chemistry, Vol.82 pp.3597-3605

Azam, T.A., & Ishihama, A., 1999 "Twelve Species of the Nucleoid-Associated Protein from Escherichia coli, Sequence Recognition Specificity and DNA Binding Affinity," Journal of Biological Chemistry, Vol.274 pp.33105-33113

Azam, T.A., *et al.*, 1999 "Growth Phase-Dependent Variation in Protein Composition of the Escherichia coli Nucleoid Growth Phase-Dependent Variation in Protein Composition of the Escherichia coli Nucleoid" Journal of Bacteriology, Vol.181 pp.6361-6370

Ballet, T., *et al.*, 2012 "DnaK Prevents Human Insulin Amyloid Fiber Formation on Hydrophobic Surfaces," Biochemistry, Vol.51 pp.2172-2180

Basnak, G., *et al.*, 2010 "Viral Genomic Single-Stranded RNA Directs the Pathway Toward a T=3 Capsid," Journal of Molecular Biology, Vol.395 pp.924-936

Bellapadrona, G., *et al.*, 2010 "Dps Proteins Prevent Fenton-Mediated Oxidative Damage by Trapping Hydroxyl Radicals within the Protein Shell," Free Radical Biology and Medicine, Vol.48 pp.292-297

Bertelsen, E.B., *et al.*, 2009 "Solution Conformation of Wild-Type E. coli Hsp70 (DnaK) Chaperone Complexed with ADP and Substrate," Proceedings of the National Academy of Sciences, Vol.106 pp.8471-8476

Bird, J.G., *et al.* 2006 "Functional Analysis of CbpA, a DnaJ Homolog and Nucleoid-Associated DNA-Binding Protein," Journal of Biological Chemistry, Vol.281 pp.34349-34356

Browning, D., Grainger, D., & Busby, S., 2010 "Effects of Nucleoid-Associated Proteins on Bacterial Chromosome Structure and Gene Expression," Current Opinion in Microbiology, Vol.13 pp.773-780

Calhoun, L., & Kwon, Y., 2011 "Structure, Function and Regulation of the DNA-Binding Protein Dps and its Role in Acid and Oxidative Stress Resistance in Escherichia coli: A Review," Journal of Applied Microbiology, Vol.110 pp.375-386

Chae, C., *et al.*, 2004 "CbpA, a DnaJ Homolog, is a DnaK Co-Chaperone, and its Activity is Modulated by CbpM" Journal of Biological Chemistry, Vol.279 pp.33147-33153

Cheng, X., *et al.* 1996 "Direct Measurement of Oligonucleotide Binding Stoichiometry of Gene V Protein by Mass Spectrometry," Proceedings of the National Academy of Sciences, Vol.93 pp.7022-7027

Chenoweth, M., Trun, N., Wickner, S., 2007 "*In vivo* Modulation of a DnaJ Homolog, CbpA, by CbpM," Journal of Bacteriology, Vol.189 pp.3635-3638

Chenoweth, M., & Wickner, S., 2008 "Complex Regulation of the DnaJ Homolog CbpA by the Global Regulators Sigma S and Lrp, by the Specific Inhibitor CbpM, and by the Proteolytic Degradation of CbpM," Journal of Bacteriology, Vol.190 pp.5153-5163

Chintakayala, K., & Grainger, D., 2011 "A Conserved Acidic Amino Acid Mediates the Interaction between Modulators and Co-Chaperones in Enterobacteria" Journal of Molecular Biology, Vol.411 pp.313-320

Cosgriff, S., *et al.* 2010 "Dimerisation and DNA-dependent aggregation of the Escherichia coli nucleoid protein and chaperone CbpA," Molecular Microbiology, Vol.77 pp.1289-1300

Dame, R., Kalmykowa, O., & Grainger, D., 2011 "Chromosomal Macrodomains and Associated Proteins: Implications for DNA Organization and Replication in Gram Negative Bacteria," PloS genetics, Vol.7 e1002123

Eswar, N., 2006 "Comparative Protein Structure Modelling Using Modeller," Current Protocols in Bioinformatics, DOI: 10.1002/0471140864.ps0209s50

Grainger, D., *et al.* 2008, "Selective Repression by Fis and H-NS at the Escherichia coli Dps promoter" Molecular Microbiology, Vol.68 pp.1366-1377

Gherezghiher, T., 2013 "1,2,3,4-Diepoxybutane-Induced DNA-Protein Cross-Linking in Human Fibrosarcoma (HT1080) Cells," Journal of Proteome Research, Vol.12 pp.2151-2164

Guéroult, M., *et al.*, 2012 "Mg (2+) in the Major Groove Modulates B-DNA Structure and Dynamics," PloS One, Vol.7 e41704

Gur, E., Katz, C., & Ron, E., 2005 "All Three J-Domain Proteins of the Escherichia coli DnaK Chaperone Machinery are DNA Binding Proteins." FEBS letters, Vol.579 pp.1935-1939

Gyi, J., *et al.* 1998 "Solution Structures of DNA.RNA Hybrids with Purine-Rich and Pyrimidine-Rich Strands: Comparison with the Homologous DNA and RNA Duplexes," Biochemistry, Vol.37 pp.73-80

Hartwig, A., 2001 "Role of Magnesium in Genomic Stability," Mutation Research, Vol.475 pp.113-121

Jaswal, S., 2013 "Biological Insights from Hydrogen Exchange Mass Spectrometry," Biochimica et Biophysica Acta, Vol.1834 pp.1188-1201

Jeong, K.C., *et al.*, 2008 "Acid Stress Damage of DNA is Prevented by Dps Binding in Escherichia coli O157:H7," BMC Microbiology, Vol.8 pp.181

Jones, D., 1999 "Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices," Journal of Molecular Biology, Vol.292 pp.195-202

Kebarle, P., & Verkerk, U., 2009 "Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know," Mass Spectrometry Reviews, Vol.28 pp.898-917

Kibbe, W.A., 2007 "OligoCalc: An Online Oligonucleotide Properties Calculator," Nucleic Acids Research, Vol.35 (web server issue) pp.W43-W46

Kjeldsen, F., & Zubarev, R., 2011 "Effects of Peptide Backbone Amide-to-Ester Bond Substitution on the Cleavage Frequency in Electron Capture Dissociation and Collision-Activated Dissociation," Journal of the American Society for Mass Spectrometry, Vol.22 pp.1441-1452

Li, Y.F., Hayashi, T., & Terawiki, Y., 1998 "Functional Domains of Rts1 and P1 RepA Proteins for Initiation of Replication," Plasmid, Vol.40 pp.140-149

Li, J, Qian, X., & Sha, B., 2003 "The Crystal Structure of the Yeast Hsp40 Ydj1 Complexed with its Peptide Substrate," Structure, Vol.11 pp.1475-1483

Light-Wahl, K., *et al.*, 1993 "Observation of a Small Oligonucleotide Duplex by Electrospray Ionisation Mass Spectrometry," Journal of the American Chemical Society, Vol.115 pp.803-804.

Meyer, A., & Baker, T., 2011 "Proteolysis in the Escherichia coli Heat Shock Response: A Player at Many Levels," Current Opinion in Microbiology, Vol.14 pp.194-199

Mogk, A., 1999 "Identification of Thermolabile Escherichia coli Proteins: Prevention and Reversion of Aggregation by DnaK and ClpB." EMBO Journal, Vol.18 pp.6934-6949

Ogata, Y., *et al.* 1996 "DnaK Heat Shock Protein of Escherichia coli Maintains the Negative Supercoiling of DNA Against Thermal Stress," Journal of Biological Chemistry, Vol.271 pp.29407-29414

Owczarzy, R., *et al.*, 2008 "IDT SciTools: A Suite for Analysis and Design of Nucleic Acid Oligomers," Nucleic Acids Research, Vol.36 (web server issue) pp.W163-W169

Patury, S., Miyata, Y., & Gestwicki, J., 2009 "Pharmacological Targeting of the Hsp70 Chaperone," Current Topics in Medical Chemistry, Vol.9 pp.1337-1351

Peng, J., & Xu, J., 2011 "RaptorX: Exploiting Structure Information for Protein Alignment by Statistical Inference," Proteins, Vol.79 pp.161-171

Pettersen, E., *et al.*, 2004 "UCSF Chimera--A Visualization System for Exploratory Research and Analysis," Journal of Computational Chemistry, Vol.25 pp.1605-1612

Preston, G., *et al.*, 2012 "Covalent Cross-Linking within Supramolecular Peptide Structures," Analytical Chemistry, Vol.84 pp.6790-6797

Roy, A., Kucukural, A., & Zhang, Y., 2010 "I-TASSER: a Unified Platform for Automated Protein Structure and Function Prediction," Nature Protocols, Vol.5 pp.725-738

Sahota, H., 2013, Discussion Regarding the Methods Being Applied to Derive a CCS Refined Computational Model of the Biologically Relevant CbpA; [Email] (Personal Communication 10^{h} May 2013)

Sahota, H., *et al. In Preparation*

Salbo, R., *et al.* 2012 "Travelling-Wave Ion Mobility Mass Spectrometry of Protein Complexes: Accurate Calibrated Collision Cross-Sections of Human Insulin Oligomers," Rapid Communications in Mass Spectrometry, Vol.26 pp.1181-1193

Sarraf, N., *et al.*, 2010 "Structural Basis of the Regulation of the CbpA Co-Chaperone by its Specific Modulator CbpM" Journal of Molecular Biology, Vol.398 pp.111-121

Sander, C., & Ts'o, P., 1971 "Interaction of Nucleic Acids. VIII. Binding of Magnesium Ions by Nucleic Acids," Journal of Molecular Biology, Vol.55 pp.1-21

Schwede, T., *et al.* 2003 "SWISS-MODEL: An Automated Protein Homology-Modelling Server," Nucleic Acids Research, Vol.31 pp.3381-3385

Shvartsburg, A.A., & Smith, R.D., 2008 "Fundamentals of Travelling Wave Ion Mobility Spectrometry," Analytical Chemistry, Vol.24 pp.9689-9699

Srinivasan, S., *et al.* 2012 "Molecular Chaperones DnaK and DnaJ Share Predicted Binding Sites on Most Proteins in the E. coli Proteome," Molecular Biosystems, Vol.8 pp.2323-2333

Steen, H., *et al.*, 2001 "Mass Spectrometric Analysis of a UV-Cross-Linked Protein-DNA Complex: Tryptophans 54 and 88 of *E . coli* SSB Cross-Link to DNA," Protein Science, Vol.10 pp.1989-2001

Tani, T., *et al.* 2002 "Adaptation to Famine: A Family of Stationary-Phase Genes Revealed by Microarray Analysis," Proceedings of the National Academy of Sciences, Vol.99 pp.13471-13476

Tucker, D., Tucker, N., & Conway, T., 2002 "Gene Expression Profiling of the pH Response in Escherichia coli" Journal of Bacteriology, Vol.184 pp.6551-6558

Ueguchi, C., Yamada, H., & Mizuno, T., 1994 "An Analogue of the DnaJ Molecular Chaperone in Escherichia coli," Biochemistry, Vol.91 pp.1054-1058

Ueguchi, C., *et al.*, 1995 "A Study of the Double Mutant of dnaJ and cbpA, Whose Gene Products Function as Molecular Chaperones in *Escherichia coli*," Journal of Bacteriology, Vol.177 pp.3894-3896

Walsh, P., *et al.*, 2004 "The J-protein family: Modulating Protein Assembly, Disassembly and Translocation," EMBO reports, Vol.5 pp.567-571

Wegrzyn, A., Taylor, K., & Wegrzyn, G., 1996 "The cbpA Chaperone Gene Function Compensates for dnaJ in Lambda Plasmid Replication During Amino Acid Starvation of Escherichia coli," Journal of Bacteriology, Vol.178 pp.5847-5849

Wolf, S., *et al.*, 1999 "DNA Protection by Stress-Induced Biocrystallization," Nature Letters, Vol.400 pp.83-85

Xu, N., *et al.*, 1999 "Electrospray ionization-mass spectrometry study of the interaction of cisplatin-adducted oligonucleotides with human XPA minimal binding domain protein," Analytical Biochemistry, Vol.272 pp.26-33

Yamashino, T., *et al.* 1994 An Analogue of the DnaJ Molecular Chaperone Whose Expression is Controlled by Sigma s During the Stationary Phase and Phosphate Starvation in Escherichia coli," Molecular Microbiology, Vol.13 pp.475-483

Zeth, K., 2012 "Dps Biomineralizing Proteins: Multifunctional Architects of Nature," The Biochemical Journal, Vol.445 pp.297-311

6. Project Conclusions

6.1. Summary of Project

The use of mass spectrometry and IM-MS to study protein-ligand interactions is becoming increasingly widespread, both in fundamental research, and as a clinical tool. The work presented here has used a combination of different mass spectrometric and IM-MS tools, including MS/MS and CIU, in addition to solution phase modifications. These mass spectrometric methods, when compared to other structural methods, represent increases in mass sensitivity as well as dynamic range. This is in addition to applications for determining the small changes in structural conformation, and the ability to study the organisation of complex multimeric sub complexes, with neither being readily achievable using other structural methods. These methods have made it possible to investigate here the gas phase oligomers and structural conformations of three proteins that are required, within their respective organisms, to ensure cellular viability.

Exhibiting a native metastable fold, mutations of α_1-Antitrypsin can promote misfolding events increasing the susceptibility to aggregation events at sites of synthesis through toxic gain of function. Promotions of increased stability AAT polymerogenic intermediates required for polymerisation are known to propagate under biologically relevant conditions. Using IM-MS we have characterised the structure and stability of the slow polymerisation mutant K154N. Our results have shown that this mutant populates a conformation intermediate under biologically relevant temperatures, corresponding to a significant 8.3 % gain in CCS. Using CID we have shown that this K154N substitution promotes increases in AAT intermediate structural stability, common with other more severe polymerogenic mutations.

Saccharomyces cerevisiae Sgt1 dimers mediate binding between Hsp90 and Skp1, to initiate chromosome separation. Mutations of the Sgt1 TPR domain that prevent dimerisation further inhibit Sgt1 mediated Hsp90: Skp1 interactions, arresting the cell cycle at the G2/M interface. Work here has shown that Sgt1 dimerisation is promoted by an Ascomycota specific structural loop within the TPR domain. This observation is supported

by comparisons of the wt Sgt1 oligomeric state with H59A and D61R structural loop substitutions, studied under native conditions. Our data has additionally shown that this loop is orientated in a manner that supports the formation of higher order oligomers under increasing sample concentrations, including previously unobserved tetramers and pentamers. Analysis using MS/MS, combined with a recently resolved crystal structure however, has shown that these Sgt1 dimers do not represent the structural minimum for stable Skp1 binding.

In addition to observations regarding the role of Ascomycota specific structural loop in Sgt1 dimerisation, work presented here has indicated, for the first time, that Sgt1 dimers are stabilised by contributions from the C-terminus. Deletion mutants lacking this C-terminus have been shown to undergo concentration-dependent oligomerisation. The results here have further indicated that in the absence of this C-terminal CS and SGS domains, the Sgt1 TPR domain may enable higher-order binding through a potential second mode of dimerisation.

The *Escherichia coli* DNA binding protein CbpA is required to promote chromosome compaction during the stationary cell cycle phase, and cellular phosphate starvation events. How CbpA oligomerizes on the DNA backbone, promoting genome compaction, is not currently understood. It had been proposed that native mass spectrometry methods would enable the study of these interactions. Although unsuccessful in observing protein-DNA interactions using native mass spectrometry, the studies presented have shown that the buffers required for nESI are suitable for CbpA-DNA binding. IM-MS has however proved to be successful in providing gas-phase constraints for computational modelling of the biologically relevant DNA binding CbpA dimer by Harpal Sahota, results forthcoming.

6.2. Of Mountains and Molehills: Towards Ever Higher Peaks

As discussed in chapter 1, improvements in instrumentation and experimental methods made over the last century have allowed mass spectrometry, with contributions from IM, to evolve from a tool used to study the fundamental nature of the atom, to one that now allows us to study the basis of life itself. The question then remains, what of the next century?

Where does a field that has impacted on our fundamental understanding of the universe go next? Three topics of increasing interest within the wider biological mass spectrometry community that the author believes to be driving both the experimental and technological development of the field to ever greater 'heights,' are explored in brief below.

6.2.1. Towards Ever Higher Peaks: Solution Phase Modified Protein Stabilisation

A recent trend in native mass spectrometry and IM-MS analysis is the addition of salt additives to the electrospray buffer, in order to promote the solution modified gas phase stability of proteins and protein complexes. Recent investigations using CID and CIU methods have shown that additions of tuned salts, at concentrations less than 2 mM, can be used to drastically modify the conformational stability of multimeric proteins in the gas phase (Fig. 6.1). Interestingly, CIU studies of avidin in the presence of magnesium cations, has been reported to stabilise unique, additional co-populated structural conformations, not previously observed in the absence or presence of other salts (Han & Ruotolo 2013).

Both anions and cations have been recently classified according to their influence on protein gas-phase stability. Both classes of ions modulate stabilisation in a manner that mirrors the typical solution phase Hofmeister series. The stability series of cations is currently denoted using the following order: NH_4^+ ~TMA^+ < Rb^+ < K^+ < Na^+ < Tris H^+ < Ba^{2+} ~Li^+ < Ca_2^+ < Mg^{2+} (Han, Hyung & Ruotolo 2013).[43] The order of this ion series currently correlates positively to the charge per unit area; increased charge density ions promote improved stabilising effect compared to those with low charge density (Han, Hyung & Ruotolo 2012). The hypothesised mechanisms of cation-mediated protein stability occurs through either multidentate cross linking interactions, or by replacing proton charge carriers with less mobile cation charge carriers restricting charge mobility and coulombic unfolding under dissociation conditions (Han, Hyung & Ruotolo 2013).

In contrast to the direct high to low series of cations, anions are clustered into one of three groups based on their stabilising effects: high ($Tartrate_2^-$, Cl^-, $citrate_2^-$, NO_3^-), medium (SO_4^{2-}, HPO_4^{2-}, SCN^-, F^-), and 'little to no' (HCO_3^-, I^-, ClO_4^-). In further contrast to the

[43] Hofmeister cation series stabilizing affects: $TMA^+ > NH_4^+ > Rb^+ > K^+ > Na^+ > Li^+ > Ca_2^+ > Mg_2^+ > Ba_2^+$ (Han, Hyung & Ruotolo 2013)

cation series, increases in charge density do not represent increased protein stability (Han, Hyung & Ruotolo 2013). Mechanistically, these negatively charged ions exhibit a dissociative cooling mechanism of protein stabilisation, dissociating to sequester excess protein rotational and vibrational energy. Those failing to bind, or dissociate, contribute to the 'little to no affect' stability group above (Han, Hyung & Ruotolo 2013).

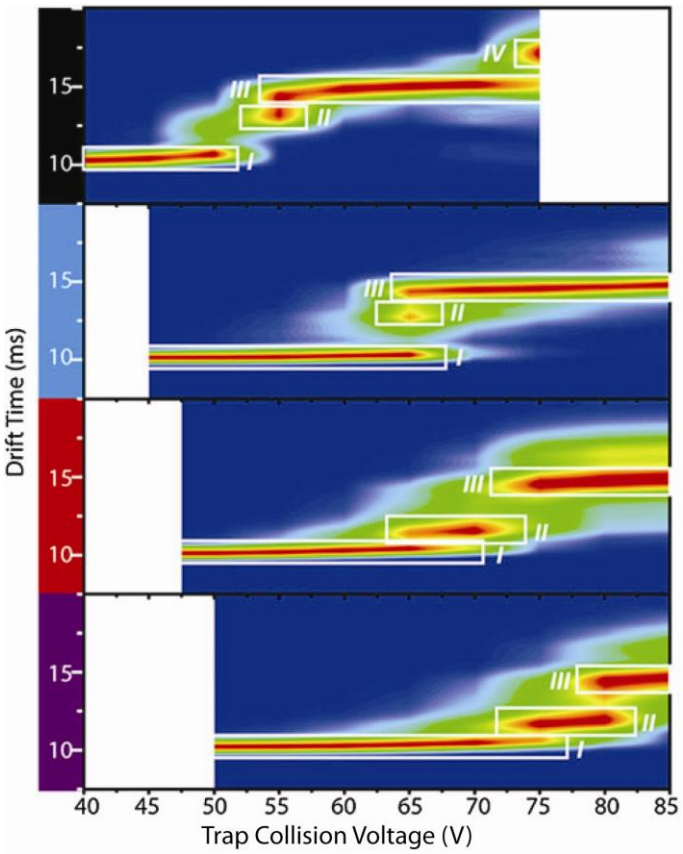

Figure 6.1. Conformational Stability of Avidin in the Presence of Salt Additives CIU of avidin (+16, most intense peak across all solution conditions) with no salt (black), 4 mM NH_4Cl (blue), 2 mM $MgCH_3CO_2$ (red), and 2 mM $MgCl_2$ (purple). Structure I is the native conformation; whilst II, III and IV represent co-populated unfolding intermediates. Data shows that individually Mg^{2+} and Cl^- ions can stabilise protein folding (+10 % against control). Simultaneous binding promotes co-operative stabilising effects (+50 % against control). Interestingly in the presence of Mg^{2+} ions, avidin conformation II represents a reduced, unique, t_D not previously accessed. This may be additionally supported by a potential additional intermediate between II and III in the presence of $MgCl_2$.

Figure reproduced and adapted from Han & Ruotolo 2013

Progressing towards the future, studies such as these support a role for stabilising the native-like structure of proteins and protein-protein interactions within the gas phase. This area of work therefore has direct applications in studies including stabilising weak protein-ligand interactions, and membrane proteins. Membrane proteins, due to their location require 'sticky' detergent clusters to enable sample recovery and ionisation of the

native like oligomers. To resolve these proteins within mass spectra, high collision voltages are required to remove the detergent cluster and may destabilise their native like structure (Borysik & Robinson 2012). Use of these tuned salt additives therefore may represent a method to stabilise the native like structure of these macro molecules under the harsh conditions required to remove detergent clusters.

Of future interest with regards to work performed in chapter 5, is the potential adaptation of these tuned salt adducts to stabilise protein-DNA interactions. Considering that divalent cations such as Mg^{2+} are required to stabilise the dsDNA phosphate backbone (Guéroult *et al.* 2012), this presents an interesting alternative application to the studies previously performed.

6.2.2. Towards Ever Higher Peaks: Spatially Resolved Mass Analysis

Although not directly connected with the work performed here, mass spectrometry imaging (MSI) has slowly progressed over the last twenty years into a method that enables *ex vivo* sample study, with spatial and chemical identification (Fig. 6.2). This technique therefore has applications in a range of biological studies, including proteomic and metabolomic analysis, whilst additionally driving developments in mass spectrometric experimental methods, software and hardware (Chughtai & Heeren 2010).

Combining suitable spatial resolution and dynamic mass quantitation range, MSI provides advantages over other commonly used experimental imaging methods such as AFM (Kiss *et al.* 2013). MSI can therefore be readily adapted to studies which probe changes in molecular organisation and cellular biochemistry, including studies of neurodegenerative disorders and the progression of cancerous tissues (Chughtai & Heeren 2010).

Recent advances within the field of MSI have seen the adaptation of the Timepix, a hybrid active pixel sensor originally used for photon detection systems, for mass spectrometry applications. Structurally, the basic Timepix assembly comprises a 256x256 pixel semiconductor that, when activated, produces electron pairs to induce charge on a CMOS read out chip (Campbell 2011, and Jungmann & Heeren 2013).

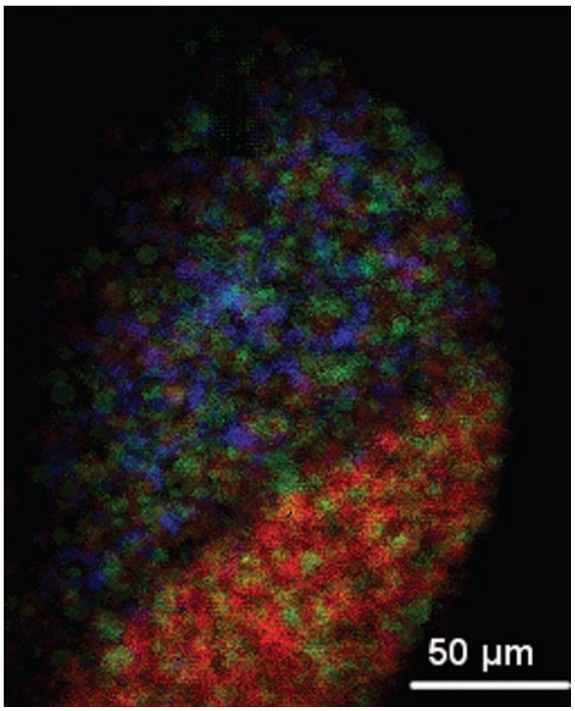

Figure 6.2. Mass Spectrometry Imaging: Spatially Resolved Mass Values

Overlay of Secondary ion ToF mass spectrometry resolved image of mouse testis cross section. Diagonal line represents the border between tissue and the indium containing glass slide. Colours denote masses correlating to: indium (*m/z* 115); red, unstated organic ion (*m/z* 358); blue, and cholesterol-OH (*m/z* 369); green.

Figure reproduced from Kiss *et al.* 2013

Used within mass spectrometric instrumentation, this assembly is combined with an MCP floated in front of the pixel sensor that upon ion impact discharges an electron ion cloud that can be detected by the Timepix (Jungmann & Heeren 2013).[44] As each pixel functions as an independent detector, the MCP-Timepix assembly allows combination of three different ion measurement modes: ion counting, ToF, and time-over-threshold (measure of pixel activity above detection threshold) (Jungmann *et al.* 2013).

Subsequently array detectors such as Timepix allow multiple ion measurements at any instant in time, with increases in both instrumental detection limits and the duty cycle repetition rate (Jungmann & Heeren 2013). Combined with additional software developments, MSI has the potential to accelerate diagnosis and prescription of therapeutics in the future, with the high-throughput required by clinical laboratories.

6.2.3. Towards Ever Higher Peaks: Life in the Gas Phase

By the time John B. Fenn presented his Nobel lecture on the eve of receiving his award ("Electrospray Wings for Molecular Elephants," Fenn 2002), ESI had already been used to

[44] This single pixel activation requires ~650 electrons (Jungmann *et al.* 2013).

ionise the tobacco mosaic virus. Collections of the virus within the gas phase were sequentially used to successfully infect tobacco plants (Siuzdak *et al.* 1996). Subsequent work presented ToF mass resolved values of the tobacco mosaic (Fig. 6.3), and rice yellow mottle, viruses. The masses of these viruses were calculated using a combination of *m/z* peak centring and charge values, and equal to 39 to 42 MDa and 6 to 7 MDa respectively. These masses however were calculated with a mass error value of $^{+}/_{-}$ 15 %, and represents limited accuracy (Fuerstenau *et al.* 2001).

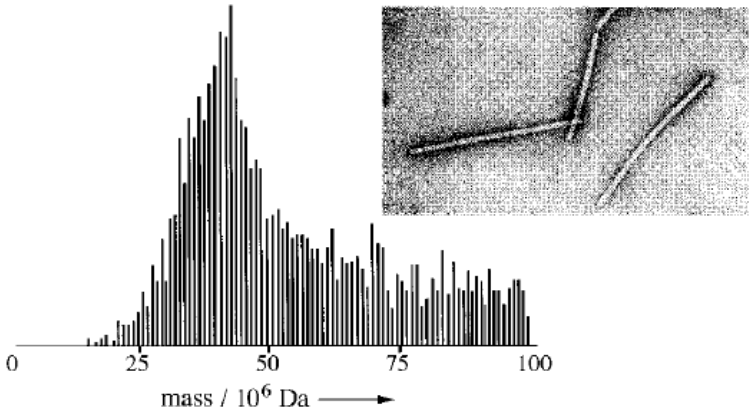

Figure 6.3. Mass Analysis of the Intact Tobacco Mosaic Virus

Earliest result of a tobacco mosaic virus, observed intact, using mass spectrometric methods (electron micrograph, top right). Resolved mass values were produced using a combination of *m/z* peak centring (above) and charge values. The molecular weight of the virus occurs between 39-42 MDa with a mass deviation error of $^{+}/_{-}$ 15 %, existing in line with the mass limitations of the instrument used.

Figure reproduced and adapted Siuzdak *et al.* 1996

Using a commercially available, quadrupole-ToF instrument (Q-ToF II, Waters, Manchester, UK) modified for high mass studies, scientists within the Heck research group have recently calculated the mass of the intact HK97 bacteriophage capsid within suitable error; 17 942 kDa $^{+}/_{-}$ 4 kDa (Fig. 6.4) (Snijder *et al.* 2013). The low error of this mass ($^{+}/_{-}$ 4 kDa) is attributed to combining the use of $Xe_{(g)}$ for enhanced collisional desolvation (Snijder *et al.* 2012), and software algorithms that minimise mass standard deviation within the analyte CSD (LeastMass, Tseng *et al.* 2011). These results are currently the largest mass resolved with suitable accuracy by any mass spectrometer within the literature.

Studies such as these can be applied to the analysis of viral capsid assembly, and virus-host interactions models (Uetrecht & Heck 2011). This is in addition to applications

within medicine and nanotechnology, in which these capsids are used for nano-reactors/containers (Snijder *et al.* 2013). In tandem, with interest regarding the studies of these interactions, are the instrumental developments to improve the mass analysis range of mass spectrometry instrumentation. Indeed Snijder and co-workers (2013) hypothesised that with the *R* of the instrument used (*R* = 7 000) for their above study, given a suitable desolvation period, analysis of ions of up to 40 Gigadaltons would theoretically be resolvable. Above this mass limit, it is hypothesised that the peak width and separation would start to overlap so that they become irresolvable (Snijder *et al.* 2013). These desolvation events therefore represent a key limiting step in the high mass resolution of ToF instruments.

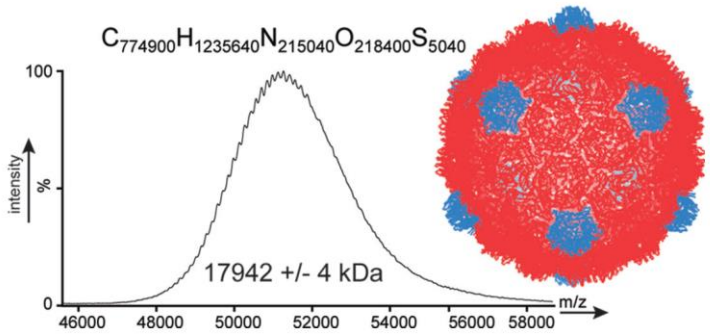

Figure 6.4. Mass Analysis of the Intact HK97 Virus Capsid

Recent data presented by Snijder and co-workers (2013) has resolved the mass of the HK97 bacteriophage viral capsid, to be 17 942 $^{+}/_{-}$ 4 kDa, with a mass deviation of 1.3 % compared to the theoretical (17 742 kDa). To date this represent the largest identifiable mass clearly resolved using mass spectrometry based methods.

Figure reproduced and adapted from Snijder *et al.* 2013

Developments to extend the mass range of Orbitrap mass analysers, which have greatly increased *R* values compared to ToF, have been recently applied to clearly resolve masses exceeding 800 kDa (*E. coli*, GroEL) to baseline. This improved base line resolution (*R* value of 25 000 at *m/z* 5 000, 16 000 at *m/z* 10 000) reportedly occurs as a function of the increased desolvation time (Rose *et al.* 2012). Therefore developments of Orbitrap instruments that allow high *R*, high *m/z* studies, may represent an alternative future route to clearly resolved high mass analytes.

Combination of these developments may therefore make it possible to one day study life within the gas phase.

6.3. Final Remarks

Although only using a fraction of biologically relevant mass spectrometric methods, the work presented here represents a broad application of several such techniques to studies of protein structure, conformation and stability. These results additionally highlight the application of biological MS to study polymerogenic and deletion mutants, as well as their effect on inhibited/unregulated oligomerisation. This is in addition to the potential to study protein-DNA gas phase interactions, using solution phase modulation techniques.

Combination of the studies presented here, in tandem with literature highlighting both past progression and future applications, has shown that growth of both the biological and wider field of MS is far from reaching a plateau. Indeed, although the comparison of 'mountains and molehills' may be relevant to Thomson's potential comparison of current MS techniques to his work more than a century ago, we may in future be increasingly able to paraphrase these words as 'like comparing moons to molehills.'

6.4 References

Borysik, A., & Robinson, C., 2012 “The 'Sticky Business' of Cleaning Gas-Phase Membrane Proteins: a Detergent Oriented Perspective,” Physical Chemistry Chemical Physics, Vol.14 pp.14439-14449

Campbell, M., 2011 “10 Years of the Medipix2 Collaboration,” Nuclear Instruments and Methods in Physics Research Section A, Vol.644 pp.S1-S10

Chughtai, K., & Heeren, R., 2010 “Mass Spectrometric Imaging for Biomedical Tissue Analysis,” Chemical Reviews, Vol.110 pp.3237-3277

Fenn, J.B., 2002 “Wings for Molecular Elephants [Nobel Lecture],” Angewandte Chemie-International Edition, Vol.42 pp.3871-3894

Fuerstenau, S., *et al.* 2001 “Mass Spectrometry of an Intact Virus,” Angewandte Chemie-International Edition, Vol.2001 pp.541-544

Guéroult, M., *et al.*, 2012 “Mg (2+) in the Major Groove Modulates B-DNA Structure and Dynamics,” PloS One, Vol.7 e41704

Han, L., Hyung, S, & Ruotolo, B., 2012 “Bound Cations Significantly Stabilize the Structure of Multiprotein Complexes in the Gas Phase,” Angewandte Chemie-International Edition, Vol.51 pp.5692-5695

Han, L., & Ruotolo, B., 2013 “Travelling-Wave Ion Mobility-Mass Spectrometry Reveals Additional Mechanistic Details in the Stabilization of Protein Complex Ions through Tuned Salt Additives,” International Journal of Ion Mobility Spectrometry, Vol.16 pp.41-50

Han, L., Hyung, S, & Ruotolo, B., 2013 “Dramatically Stabilizing Multiprotein Complex Structure in the Absence of Bulk Water Using Tuned Hofmeister Salts,” Faraday Discussions, Vol.160 pp.371-388

Jungmann, J.H., *et al.*, 2012 “Biological Tissue Imaging with a Position and Time Sensitive Pixelated Detector,” Journal of the American Chemical Society, Vol.23 pp.1679-1688

Jungmann, J.H., & Heeren, R.M.A., 2013 “Detection Systems for Mass Spectrometry Imaging: A Perspective on Novel Developments with a Focus on Active Pixel Detectors,” Rapid Communications In Mass Spectrometry, Vol.27 pp.1-23

Jungmann, J.H., *et al.*, 2013 “An In-Vacuum, Pixelated Detection System for Mass Spectrometric Analysis and Imaging of Macromolecules,” International Journal of Mass Spectrometry, Vol.341-342 pp.34-44

Kiss, A., *et al.* 2013 “Microscope Mode Secondary Ion Mass Spectrometry Imaging with a Timepix Detector,” The Review of Scientific Instruments, Vol.84 pp.013704 1-7

Rose, R., *et al.* 2012 “High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies,” Nature Methods, Vol.9 pp.1084-1088

Siuzdak, G., *et al.*, 1996 “Mass Spectrometry and Viral Analysis.” Chemistry and Biology, Vol.3 pp.45-48

Snijder, J., *et al.* 2013 “Studying 18 MDa Virus Assemblies with Native Mass Spectrometry,” Angewandte Chemie-International Edition, Vol.52 pp.4020-4023

Tseng, Y-H. *et al.*, 2011 "Interpreting the Charge State Assignment in Electrospray Mass Spectra of Bioparticles," Analytical Chemistry, Vol.83 pp.1960-1968

Uetrecht, C., & Heck, A., 2011 "Modern Biomolecular Mass Spectrometry and its Role in Studying Virus Structure, Dynamics, and Assembly," Angewandte Chemie-International Edition, Vol.50 pp.8248-8262

Appendix I

Structure
Article

Structural Dynamics Associated with Intermediate Formation in an Archetypal Conformational Disease

Mun Peak Nyon,[1,5] Lakshmi Segu,[1,5] Lisa D. Cabrita,[1,2,5] Géraldine R. Lévy,[1,2,5] John Kirkpatrick,[2] Benoit D. Roussel,[3] Anathe O.M. Patschull,[1,2] Tracey E. Barrett,[1] Ugo I. Ekeowa,[3] Richard Kerr,[2] Christopher A. Waudby,[1,2] Noor Kalsheker,[4] Marian Hill,[4] Konstantinos Thalassinos,[1,2] David A. Lomas,[3,6,*] John Christodoulou,[1,2,6,*] and Bibek Gooptu[1,6,*]
[1]Institute of Structural and Molecular Biology (ISMB), Department of Biological Sciences, Birkbeck College, London, WC1E 7HX, UK
[2]ISMB, Division of Biosciences, University College London, London, WC1E 6BT, UK
[3]Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Cambridge, CB2 0XY, UK
[4]Division of Clinical Chemistry, Queen's Medical Centre, Nottingham, NG7 2UH, UK
[5]These authors contributed equally
[6]These authors contributed equally
*Correspondence: dal16@cam.ac.uk (D.A.L.), j.christodoulou@ucl.ac.uk (J.C.), b.gooptu@mail.cryst.bbk.ac.uk (B.G.)
DOI 10.1016/j.str.2012.01.012

SUMMARY

In conformational diseases, native protein conformers convert to pathological intermediates that polymerize. Structural characterization of these key intermediates is challenging. They are unstable and minimally populated in dynamic equilibria that may be perturbed by many analytical techniques. We have characterized a *forme fruste* deficiency variant of α_1-antitrypsin (Lys154Asn) that forms polymers recapitulating the conformer-specific neo-epitope observed in polymers that form in vivo. Lys154Asn α_1-antitrypsin populates an intermediate ensemble along the polymerization pathway at physiological temperatures. Nuclear magnetic resonance spectroscopy was used to report the structural and dynamic changes associated with this. Our data highlight an interaction network likely to regulate conformational change and do not support the recent contention that the disease-relevant intermediate is substantially unfolded. Conformational disease intermediates may best be defined using powerful but minimally perturbing techniques, mild disease mutants, and physiological conditions.

INTRODUCTION

The conformational diseases are characterized by conformational transitions in mutant proteins that allow multimerization and aberrant protein deposition. They include Alzheimer's disease and Lewy body dementia, the prion encephalopathies, systemic amyloidosis, and the serpinopathies (Lomas and Carrell, 2002). The serpinopathies result from the polymerization of mutants of members of the serine protease inhibitor or serpin superfamily of proteins (Gooptu and Lomas, 2009). The archetype of the serpinopathies is α_1-antitrypsin deficiency. α_1-antitrypsin is synthesized by hepatocytes (Laurell and Jeppsson, 1975) and released into the circulation, where it protects the lung from the actions of neutrophil elastase (Gooptu et al., 2009a). Severe α_1-antitrypsin deficiency is found in 1:2000 individuals of North European descent (Blanco et al., 2001) and typically results from homozygosity for the Z allele (Glu342Lys). The Z mutation causes α_1-antitrypsin to adopt an intermediate conformation and rapidly polymerize (Figure 1A; Lomas et al., 1992). Polymers accumulate within the endoplasmic reticulum of hepatocytes in association with neonatal hepatitis, cirrhosis, and hepatocellular carcinoma. Concomitant deficiency of circulating α_1-antitrypsin predisposes to severe, early onset emphysema (Gooptu and Lomas, 2008).

Previous data suggest that serpin polymers are formed by intermolecular linkage of the reactive center loop of one molecule with β sheet A of another (Dunstone et al., 2000; Gooptu et al., 2000; Huntington et al., 1999; Lomas et al., 1992; Schulze et al., 1990). However, a recent model suggests that the polymer is linked by a more extensive domain swap involving both the reactive center loop and strand 5 of β sheet A (Krishnan and Gierasch, 2011; Yamasaki et al., 2008). The controversy may be partly explained by our recent finding that structurally distinct polymers are favored by different conditions used to induce polymerization in vitro (Ekeowa et al., 2010). Resolving these issues at a residue-specific level is important for targeting drug design against the pathological polymerization of α_1-antitrypsin deficiency. We have therefore used nuclear magnetic resonance (NMR) spectroscopy and a deficiency-associated and polymerogenic α_1-antitrypsin variant to characterize structural and dynamic changes associated with population of the pathological intermediate state.

RESULTS

Detection of α_1-Antitrypsin$_{Queen's}$

A mutation that, to our knowledge, has not previously been described was detected in an individual whose circulating α_1-antitrypsin levels (median value 0.6 mg/ml; normal range 1.5–3.5 mg/ml) were between those expected of a Z homozygote and an MZ heterozygote. Genotyping revealed a compound heterozygote for the Z allele and a Lys154Asn variant that we have named α_1-antitrypsin$_{Queen's}$.

Population of the Pathological Intermediate

Recombinant Lys154Asn α_1-antitrypsin was characterized biochemically and biophysically (Figure S1, available online;

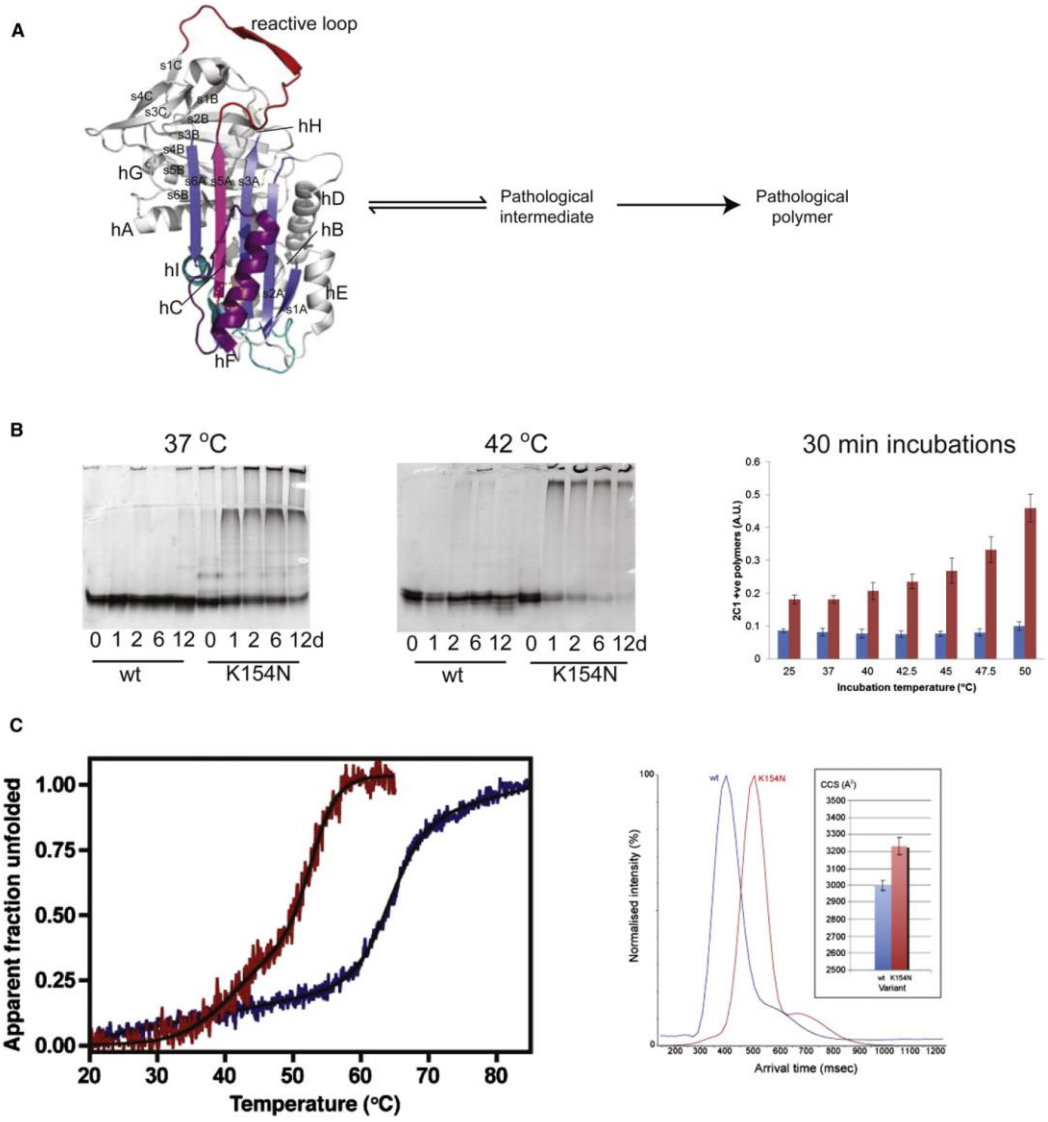

Figure 1. Pathological Polymerization of Lys154Asn α_1-Antitrypsin

(A) Polymerisation pathway from native conformer (mutation site circled).

(B) 7.5% (w/v) native PAGE; polymerization of wild-type and Lys154Asn α_1-antitrypsin (0.5 mg/ml protein [pH 7.4]) in vitro over 12 days at 37°C and 42°C. Polymerization is reported by loss of the monomeric band and the appearance of aggregated protein. (Right) Polymers formed by 30 min incubation between 40°C and 50°C of nonglycosylated wild-type (blue) and Lys154Asn (red) α_1-antitrypsin are detected by 2C1 mAb ELISA (Z α_1-antitrypsin polymer calibrated; Miranda et al., 2010). Differences between wild-type and mutant are significant ($p < 0.05$ to $p < 0.005$) for all temperatures. The increase in signal relative to starting material reaches significance for Lys154Asn α_1-antitrypsin for incubations at 47.5 ($p < 0.05$) and 50 ($p < 0.005$)°C. Data are mean ± standard deviation (error bars) of three experiments.

(C) (Left) Thermal denaturation CD spectroscopy (left; mean ellipticity at 222 nm, n = 10) for wild-type (blue) and Lys154Asn (red) α_1-antitrypsin. (Right) Arrival times and collision cross-section (CCS) values calculated by IM-MS for wild-type (blue) and Lys154Asn α_1-antitrypsin indicated a 7.8% increase in CCS in Lys154Asn relative to wild-type α_1-antitrypsin at 39°C (no difference at 20°C and 34°C). Data are mean ± standard deviation (error bars) of three experiments. See also Figure S1.

Figures 1 B and 1C). The variant had 64% of the functional activity of wild-type a_1-antitrypsin (i.e., 1.6-fold greater stoichiometry ofi nhibition [SI]) with an apparent association rate constant ($k_{app'}$) with bovine a-chymotrypsin of 5.5 (+/ −1.1) 3 10^4 $M^{-1}s^{-1}$ (n = 3). When corrected for the increase in SI, the association rate constant (k_{app}) for Lys154Asn a_1-antitrypsin was 1.3 3 10^5 $M^{-1}s^{-1}$, an order of magnitude lower than the wild-type control (1.16 3 10^6 $M^{-1}s^{-1}$). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) confirmed increased substrate-like behavior in Lys154Asn a_1-antitrypsin relative to the wild-type (Figure S1 C).

The Queen's variant formed polymers under physiological conditions (Figure 1 B). This was slow at 37 °C, and more rapid at 42 °C and higher temperatures. The general pathological relevance of these polymers was demonstrated by the presence of the 2C1 neo-epitope that is specific for polymers associated with disease (Miranda et al., 2010 ; Figure 1 B, right).

Polymer formation occurs via population of an intermediate state (Dafforn et al., 1999). We therefore studied intermediate formation by Lys154Asn a_1-antitrypsin. The variant populated intermediate states more readily than wild-type a_1-antitrypsin in urea-induced and thermal denaturing conditions (Figure S1 A; Figure 1 C). Since we had established the disease relevance of the polymerogenic intermediate at temperatures in the physiological range, we focused on structural changes induced by incubation in these conditions. CD spectroscopy studies (Figures 1 C and S1 D) indicated that the mutant fold underwent a subtle structural transition between 35 °C and 45 °C before more dramatic transition between 50 °C and 60 °C. Increasing concentration did not affect the first transition but caused the second transition to commence at a lower temperature. Wild-type a_1-antitrypsin remained stable to 60 °C and underwent a sharp unfolding transition at higher temperatures, consistent with previous studies (Gooptu et al., 2009b; Parfrey et al., 2003). Increasing concentration cold-shifted the second half of the transition but did not affect its initial phase (Figure S1 D). Increasing protein concentration 2.5-fold (0.4 to 1.0 mg/ml) reduced the midpoint of the overall transition for wild-type a_1-antitrypsin by ~1.5 °C and that of Lys154Asn a_1-antitrypsin by ~1.0 °C. The profiles of the first part of the transition observed for wild-type a_1-antitrypsin and the first transition observed for Lys154Asn a_1-antitrypsin are both concentration-independent. Therefore, this likely reports upon structural change within the molecule. The subsequent transition is concentration-dependent. It therefore likely reports, to some degree, upon polymerization of a species populated during the first phase.

Polymerogenic intermediate formation in a_1-antitrypsin has been previously studied by ANS fluorescence. ANS fluoresces strongly upon binding to regions in which both polar and hydrophobic motifs are exposed to solvent and so binds to intermediate states, rather than fully folded or unfolded polypeptides. The ability of the Queen's variant of a_1-antitrypsin to bind ANS was therefore assessed (Figure S1 E). Lys154Asn a_1-antitrypsin showed hyperfluorescence compared with the wild-type protein at 25 °C. This was unexpected, since assays directly reporting structural information (CD spectroscopy, intrinsic fluorescence, and NMR spectroscopy) strongly indicated that both proteins were similarly well-folded and stable at this temperature. These observations of Lys154Asn a_1-antitrypsin behavior are similar to those described for Z a_1-antitrypsin (Knaupp et al., 2010). Moreover, incubating Lys154Asn a_1-antitrypsin with ANS for longer resulted in even greater hyperfluorescence, whereas increasing temperature to 37 °C did not enhance this further (Figure S1 E). Wild-type a_1-antitrypsin showed no change in ANS fluorescence at either temperature or upon prolonged incubation. These data indicate that for a_1-antitrypsin variants in which conversion to the polymerogenic intermediate in solution is facile, ANS binding stabilizes the intermediate state sufficiently to skew the position of the native-intermediate equilibrium to the right. For Lys154Asn a_1-antitrypsin, this effect is so significant that it outweighs any effect of temperature change between 25 °C and 37 °C. This supports our general view that the degree to which assays perturb solution equilibria must be considered when studying intermediate formation.

We hypothesized that ion-mobility mass spectrometry (IM-MS) might have sufficient sensitivity to detect population of the intermediate while minimally perturbing the solution equilibrium. At 20 °C and 34 °C, monomeric Lys154Asn a_1-antitrypsin was indistinguishable from the wild-type protein by IM-MS. However, at 39 °C, there was a 7.8% increase in the collision cross-section (CCS) of monomeric Lys154Asn a_1-antitrypsin relative to the wild-type protein in keeping with population of an intermediate state (Figure 1 C). The CCS of the intermediate state has been calculated as ~18% greater than the native state, so a 7.8% increase in CCS is consistent with substantial population (~40%) of the intermediate state in equilibrium at physiological temperatures.

NMR Spectroscopic Characterization of a_1-Antitrypsin Solution Behavior and Mutation Effects

We have observed and assigned backbone resonances for almost all residues in wild-type a_1-antitrypsin in a series of NMR experiments (Biological Magnetic Resonance Bank entry 17804; Nyon et al., 2011), allowing us to analyze its solution behavior (Figures 2 and S2). Chemical shift index analysis (Berjanskii and Wishart, 2006) was performed for C a backbone bond signals alone or in conjunction with side-chain signals (Figure S2 A). This method did not use any crystallographically derived predictions, so it permitted residue-specific comparison of the secondary structure behavior of the protein directly observed in solution with that observed crystallographically. Although most of the secondary structure behavior ofi ndividual residues was conserved between solution and the crystal lattice conditions, b sheet C showed considerable lability in solution. The solution behavior of helices C and I was reported as a-helical by C a backbone bond signals. However, when side-chain signals were also considered, these reported random coil behavior. Such differences may indicate that these motifs also displayed some lability in solution (Figure S2 A).

Two-dimensional (2D) NMR spectroscopy of ^{15}N-labeled Lys154Asn a_1-antitrypsin was then performed (Figure 2). At 25 °C the well-dispersed regions of the spectrum contained a similar number (148) of cross-peaks to the number seen for the wild-type protein (152), and mean cross-peak intensities were similar for both proteins. A subset of 123 residues could be confidently assigned in the well-dispersed region of the ^{1}H-^{15}N TROSY-HSQC spectrum of Lys154Asn a_1-antitrypsin. Overall peak integrals in the central region were also similar between the spectra for wild-type and Lys154Asn a_1-antitrypsin

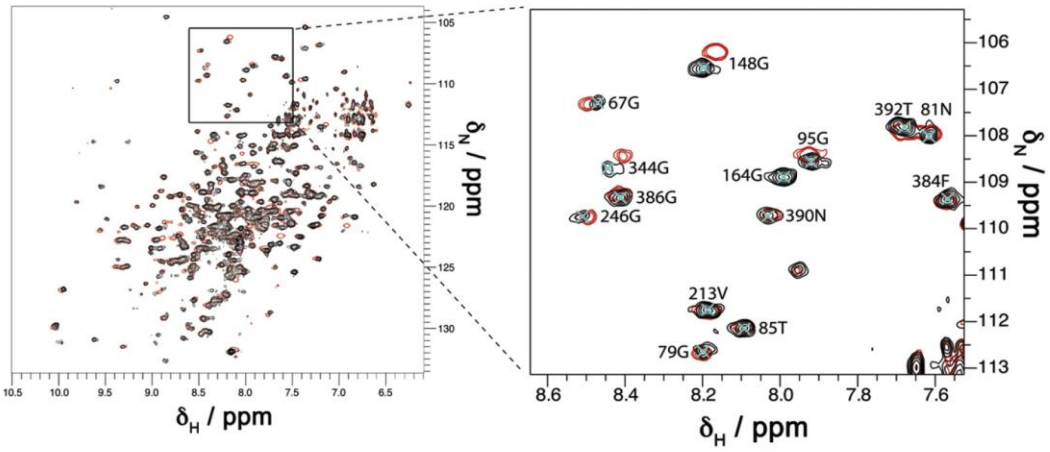

Figure 2. ^{1}H-^{15}N TROSY-HSQC Spectrum of Wild-Type and Lys154Asn α_1-Antitrypsin
Zoom (right) illustrates examples of reporter residues (wild-type, black; Lys154Asn, red).
See also Figure S2.

(Figure S2B). Nevertheless intensity and chemical shift differences were induced by the presence of the mutation in the native state of the protein (Figure S2C). The degree of variation reported by intensity and chemical shift in the observed residues correlated closely. The magnitude of changes are shown by heatmap coloring of spheres mapped to the crystal structure and demonstrate the mutation most affects residues around the F-helix, strands 1–3 of β sheet A, and the reactive loop. In particular, the spectral effects were most dramatic on residues around hF and s3A, where the chemical shift changes were too great to allow their confident assignment in the mutant protein (Figure S2C, black spheres).

NMR Spectral Changes and Polymerogenicity

Our principal aim was to assess changes in α_1-antitrypsin associated with population of the disease-relevant intermediate state. Native state behavior was defined as that observed for wild-type α_1-antitrypsin at the temperatures studied in which no polymerization was observed. At 37°C, slow, irreversible, and linear loss of the observable, monomeric sample due to polymerization was observed in Lys154Asn but not the wild-type α_1-antitrypsin. The linearity indicated a constant rate of polymer formation and hence that the population of the polymerogenic intermediate was also constant, that is, the solution was in pseudo-equilibrium. ^{1}H-^{15}N TROSY-HSQC spectra reported residue-specific structural and dynamic changes in Lys154Asn α_1-antitrypsin by nonuniform intensity change in assigned cross-peaks relative to wild-type α_1-antitrypsin (Figure 3A). Values are scaled for monomer concentration; they take into account changes observed between wild-type and mutant protein at 25°C in the absence of polymerization and changes observed in wild-type α_1-antitrypsin between 25°C and 37°C. They therefore report differential changes in Lys154Asn α_1-antitrypsin associated with population of the polymerogenic state.

Most reporter cross-peaks signaled minimal change in Lys154Asn α_1-antitrypsin at 37°C (Figure 3B). Those in strand 5 in β sheet A, helix I, and the connecting linker strongly preserved intensity (Figure 3C). NMR intensity data are highly sensitive reporters of changes in structure and dynamics, so relatively large changes may be induced by quite small changes in solution behavior. Nevertheless, residues reporting the most dramatic intensity change in the polymerogenic ensemble notably occur in regions previously linked with conformational change (upper s3A, upper hF, underlying the upper s4A site, and hA; Cabrita et al., 2004; Mahadeva et al., 1999; Zhou et al., 2003).

A further 12 distinct cross-peaks (signal:noise 11.5) were observed in the well-dispersed region of the spectrum in Lys154Asn but not the wild-type α_1-antitrypsin at 37°C (Figure S3A). The new cross-peaks disappeared on cooling and so may represent residues within a nonnative monomeric state, rather than polymers, since these did not dissociate (and would predictably give peaks of very low intensity). If so, these data support a process of native-non-native monomeric conformational exchange that is slow (≥ms) by NMR timescales, consistent with previous data (Dafforn et al., 1999).

The mutation will mildly perturb local electrostatic interactions involving Lys154 (Figure S3B). Its effects on resonances of nearby residues were too great to dissect by cross-peak analysis. We therefore studied a eukaryotic cell model of disease (Gooptu et al., 2009b) by mutagenesis. These clarified that the hydrogen bond of Lys154 (F-helix) with the main-chain carbonyl oxygen of Lys174 (neighboring linker) regulated against polymer formation (Figure S3B).

The residue-specific changes caused by the Lys154Asn mutation differed markedly from those observed in urea conditions causing a similar degree of polymerization (1.5 M; Figure S3C). Both patterns differ from that indicated by studies in guanidine or low pH conditions (Krishnan and Gierasch, 2011; Yamasaki

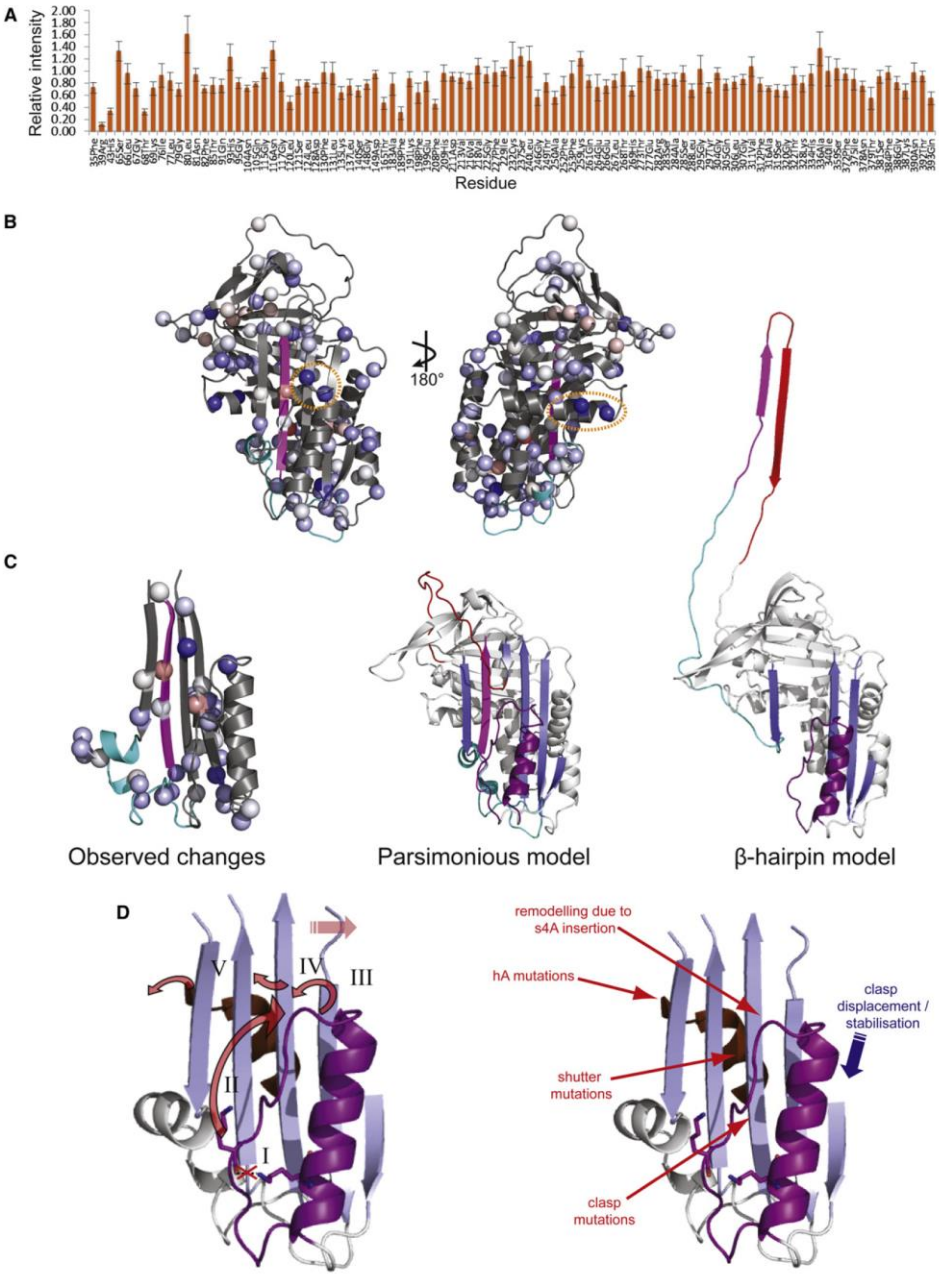

Figure 3. Population of the Intermediate by Lys154Asn α_1-Antitrypsin under Physiological Conditions

(A) Differential change in intensity for reporter cross-peaks in the ^{1}H-^{15}N TROSY-HSQC of Lys154Asn α_1-antitrypsin at 37°C. Intensities are scaled relative to the wild-type protein at the same temperature and the intensities observed for Lys154Asn α_1-antitrypsin relative to the wild-type at 25°C $[(I_{K154N,37}/I_{wt,37})/(I_{K154N,25}/I_{wt,25})]$.

et al., 2008). Thus, the polymerogenic monomer ensemble to which this mild deficiency mutation reduces the energetic barrier, differs structurally and/or dynamically from polymerogenic ensembles induced by other conditions. The resultant polymers are also structurally distinct (Ekeowa et al., 2010).

DISCUSSION

Studying key native to intermediate transitions can elucidate mechanisms of the conformational diseases and aid targeting of novel therapeutic strategies. However, the intrinsic tendency of intermediates to adopt more stable, polydisperse, and multimeric conformations renders this inherently challenging. Moreover, structurally distinct serpin polymers are induced by chemical denaturants compared to pathological mutations or heating (Ekeowa et al., 2010). Techniques that can perturb equilibria by stabilizing particular structural endpoints (e.g., selective labeling, proteolysis, and/or peptide binding) may also skew characterization of intermediate ensembles in conformational diseases. We have therefore undertaken the residue-specific study of a *forme fruste* disease mutant in solution, under physiological conditions, using minimally perturbing techniques. This is arguably the best strategy for studying intermediate states relevant to conformational diseases.

Lys154Asn α_1-antitrypsin populates a polymerogenic intermediate state at 37°C. Polymers formed from Lys154 mutants recapitulate the pathognomic 2C1 neo-epitope. Polymerization of Lys154Asn α_1-antitrypsin is therefore a good structural model for the pathological polymerization in vivo of more common and severe variants, such as Z α_1-antitrypsin. Serendipitously, polymerization occurs slowly enough to allow the study of the solution behavior of α_1-antitrypsin when the pathological intermediate state is populated. We have overcome obstacles posed by the molecule's large size (45 kDa) to use NMR spectroscopy and characterize residues that remain native-like or alter their solution behavior in association with formation of the pathological intermediate. Our data are consistent with sufficient population of the intermediate state in Lys154Asn α_1-antitrypsin at physiological temperatures to contribute to our observations. Most observed residues behave in a highly native-like way under these conditions. Relative intensity values indicate preservation of native-like signal after correcting for changes due to the mutation at 25°C and changes in protein tumbling and polymerization at 37°C. A value of 100% would therefore indicate entirely native-like structural and dynamic behavior in a particular residue. The mean relative intensity across the entire subset of residues that could be analyzed in this way was 82%. It is therefore unlikely that the intermediate ensemble is substantially unfolded as required for extensively domain-swapped models of polymerization. Specifically, analysis of cross-peak intensities supports high stability in the hI-s5A region proposed to unfold in the β-hairpin model (Figure 3C). This is indicated by relative intensities of 90% for the seven residues within this region compared with the equivalent cross-peaks in the native state. Similar preservation of native-like signal is indicated for the core strands 4 and 5 of β sheet B (residues 370–388).

Larger changes (Figures 3A and 3B) may be due to a number of differences in structural and/or dynamic behavior, so their interpretation is complex. Nevertheless, the localization of the greatest changes from wild-type native behavior is striking. Such changes are reported by residues in regions associated with remodeling during formation of the intermediate state in parsimonious models of polymerization (Figure 3B, left panel, circled: Thr165 at the top of hF, relative intensity 0.46, Phe189, upper s4a, 0.31). The effects on these residues cannot be simply explained by proximity to the mutation site, since they lie further from it than many residues that report highly native-like behavior. Residues reporting a major loss of relative intensities are also observed at the C-terminal end of hA (Figure 3B, right panel ellipse: Arg39, relative intensity 0.11 and His 43, 0.33), even further from the mutation site. This corresponds to the site of another polymerogenic α_1-antitrypsin deficiency mutation (Ile39Cys; Mahadeva et al., 1999). Mutations in nearby serpin residues (Tyr38Ser, antithrombin; Leu41Pro, α_1-antitrypsin) are also associated with deficiency and disease (Dafforn, 1999). Our data suggest that structural and/or dynamic changes here are associated with formation of the pathological intermediate. The effects would be transmitted most directly to the

All intensities were scaled for concentration and 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) control intensity. Error bars are defined according to variability observed in the datasets from which the values are derived, according to the formula

$$\left[\left(\frac{I_{K154N,37}}{I_{WT,37}}\right)\Big/\left(\frac{I_{K154N,25}}{I_{WT,25}}\right)\right]\times\sqrt{\left(\frac{d_{K154N,37}}{I_{K154N,37}}\right)^2+\left(\frac{d_{WT,37}}{I_{WT,37}}\right)^2+\left(\frac{d_{K154N,25}}{I_{K154N,25}}\right)^2+\left(\frac{d_{WT,25}}{I_{WT,25}}\right)^2},$$

where I is the intensity for data obtained from wild-type (WT) or Lys154Asn (K154N) α_1-antitrypsin at 25° or 37°C as indicated by the subscripts and d terms represent the observed standard deviation of peak intensities within each spectra.

(B) Relative intensities quantified in (A) mapped onto the subset of reporter residues (spheres) in red-white-blue heatmap coloring (increases, red; unchanged, white; reductions, blue). Greatest change (maximally blue by RGB, residue 39) corresponds to a relative intensity of 0.11. Increasing redness indicates increasing intensity on the same scale. Ellipse: hA region reporting major change from native-like intensity.

(C) (Left) Intensity change and stability observed in β sheet A, hF, hI, and connecting linkers. Current structural models of polymerogenic intermediates in the parsimonious (center) and β-hairpin (right) linkage models of polymerisation for comparison (β strand 5A magenta; hI and connecting linker cyan).

(D) (Left) Unlocking the clasp. Proposed scheme of polymerogenic intermediate formation through propagation of conformational change from the mutation site through β sheet A (blue) and more distant motifs. The Lys154Asn mutation abolishes a hydrogen bond between hF and the neighboring linker (clasp, purple) region (I). Under physiological conditions this destabilization causes remodeling of the upper part of the clasp (II). Consequent structural changes in the upper part of s3A (III) facilitate lateral movement required to open the s4A site. Changes in s3A are also transmitted via (IV) the shutter region and s6B (brown) to the C-terminal residues of hA (V). Most of the residues within β sheet A remain within a native-like environment during formation of the intermediate. This is seen not only for residues in s5A but also for residues in hI and the connecting linker. (Right) Factors stabilizing (blue, arrow denotes stabilizing displacement of clasp region induced by the Gly117Phe mutation; Gooptu et al., 2009b) and destabilizing (red) the network of regulatory interactions.

See also Figure S3.

A-helix from s3A via the intervening shutter and s6B motif (Figure 3D), known hotspots for pathological, polymerogenic mutations.

Taking together the data from biochemical, biophysical, enzyme-linked immunosorbent assay (ELISA) and cell studies, and previous observations (Bruce et al., 1994; Sharp et al., 1999), we propose that the helix F and linker region constitutes a ''clasp'' motif (Figure 3D, purple). This may seal a network of interactions that regulate conformational change. Compromise of the network is associated with formation of the pathological intermediate.

Our data indicate that the methods used to induce intermediate formation when studying conformational diseases can be major determinants of the structural distribution of intermediate states within the solution ensemble. A recent study undertook residue-specific labeling of single Cys mutants during early guanidine denaturation. It supported extensive unfolding of strands 5 and 6, as well as helices F and I (Krishnan and Gierasch, 2011). That pattern differs from our observations, both in the presence of the Lys154Asn mutation at 37°C and in low concentrations of urea. The modulation of monomeric ensembles to produce analogous but structurally distinct polymer states is relevant to studies of other conformational diseases, where chemical denaturants are commonly used in vitro to induce multimerization. It may also explain the phenomenon of strain specificity observed in conformational disorders, such as prion diseases and amyloidoses.

We have demonstrated that NMR spectroscopy is optimal for studying pathological intermediate formation in the serpinopathies, providing residue-specific detail without perturbing dynamic solution equilibria. Comparison of wild-type α_1-antitrypsin characteristics with those of mild deficiency variants also powerfully dissects disease-relevant behavior. Such studies appear timely, given the range of proposed polymerisation models based upon data obtained using a range of more perturbing techniques. Further studies using different variants and NMR modalities, for example, detailed dynamics studies, will further characterize this contentious mechanism of pathological conformational change to optimize drug design and so treat the serpinopathies.

EXPERIMENTAL PROCEDURES

Molecular Biology, Protein Purification, and Cell-Free Characterization

Mutations were introduced into α_1-antitrypsin cDNA within the pQE31 and/or the pcDNA3.1 plasmids by polymerase chain reaction mutagenesis (Zhou et al., 2001; Gooptu et al., 2009b). pQE31 plasmids containing cDNA encoding hexahistidine-tagged, recombinant α_1-antitrypsin were transfected into XL1 Blue *Escherichia coli* (Stratagene, Santa Clara, California, USA). The proteins were expressed and purified as previously described (Parfrey et al., 2003). They were characterized using SDS-, nondenaturing and transverse urea gradient (TUG)-PAGE, circular dichroism (CD), intrinsic and 8-anilinonaphthalene-1-sulfonate (ANS) fluorescence spectrometry and by enzyme inhibitory activity and kinetics assays (Daffornet al., 1999, 2004; James et al., 1999; Stone and Hofsteenge, 1986).

Ion Mobility Mass Spectrometry

Samples were buffer exchanged into 100 mM ammonium acetate (pH 7.0), desalted, and concentrated to a final concentration of 20 μM. Proteins were introduced to the mass spectrometer by nano electrospray ionization. All IM-MS experiments were performed in a hybrid quadrupole, orthogonal acceleration time-of-flight (oa-TOF) mass spectrometer equipped with a traveling wave (T-Wave) ion mobility separation device (Synapt HDMS, Waters, Manchester, UK; Pringle et al., 2007). The source temperature was 40°C, capillary voltage optimized 1.0–1.2 kV, and cone voltage was 20 V. The pressure in the T-Wave ion mobility cell was 0.55 mbar. The mobility gas was nitrogen. T-Wave height and velocity were set at 10 V and 300 m/s, respectively. Arrival time distributions were converted to collision cross-sections (CCS) by power fit (Hilton et al., 2010; Thalassinos et al., 2009). The calibrant protein was equine myoglobin (Sigma-Aldrich, Gillingham, UK). Data acquisition and processing were carried out using MassLynx 4.1 software (Waters Corp., Milford, MA, USA). Data are the mean (+/− SD) of three repeats.

Sample Preparation for NMR Spectroscopy

cDNA coding for hexahistidine-tagged α_1-antitrypsin within a pQE31 plasmid was transformed into strain BL21-Gold (DE3) *E. coli* (Stratagene). Cells were cultured in M9 minimal media (in H_2O or D_2O) at 37°C with 1 g/l of $^{15}NH_4Cl$ (Spectra Stable Isotopes, Columbia, MD, USA) and 2 g/l of either glucose or ^{13}C-glucose (Sigma-Aldrich) as the sole nitrogen and carbon sources, respectively. Following induction of protein expression cells were grown for 6 hr, rather than the 4 hr, period used with rich media. In all other respects protein expression and purification was carried out as described above. Samples were stored in 25 mM Na_2HPO_4 (pH 8.0), 50 mM NaCl, and 1 mM ethylenediaminetetraacetic acid. Sample homogeneity was confirmed by SDS-, nondenaturing and transverse urea gradient (TUG)-PAGE and by assessment of inhibitory activity.

NMR Spectroscopy Conditions

NMR spectra used to compare wild-type and Lys154Asn α_1-antitrypsin were collected on uniformly ^{15}N-labeled samples at 298 K, 175 μM on a Varian UnityInova 600 MHz spectrometer with an HCN cryoprobe. Ten percent D_2O/1% 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was added to concentrated α_1-antitrypsin samples prior to NMR spectroscopy.

Backbone assignments were obtained from TROSY (Pervushin et al., 1997) versions of HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and HN(CO)CACB spectra (Eletsky et al., 2001; Salzmann et al., 1998), together with a 3D NOESY-TROSY-^{15}N HSQC spectrum (Zhu et al., 1999). ^2H decoupling (1 kHz WALTZ-16; Shaka et al., 1983) was applied in triple-resonance experiments. Magnetization was transverse on CA or CB. Two-dimensional TROSY-^{15}N HSQC spectra for assignment. Wild-type and Lys154Asn α_1-antitrypsin comparison spectra were recorded using single-transition-to-single-transition polarization transfer and phase cycling for coherence order selection (Rance et al., 1999; Zhu et al., 1999). Spectra were processed and analyzed using nmrPipe (Delaglio et al., 1995) and CCPN (Fogh et al., 2002). All spectra were referenced to DSS at 0.0 ppm, manually phased and baseline corrected. Lys154Asn α_1-antitrypsin spectra were analyzed by changes in cross-peak intensity relative to the data on wild-type α_1-antitrypsin ($I_{K154N}/I_{wild\text{-}type}$) and magnitude of chemical shift change ($\Delta\delta$). This was calculated as

$$\sqrt{\left(\Delta\delta^1H\right)^2 + \left(\Delta\delta^{15}N/5\right)^2},$$

where Δ^1H was the change in the chemical shift along the ^1H axis and $\Delta^{15}N$ the change in chemical shift along the ^{15}N axis (parts per million [ppm]).

Cell Biological Characterization

The pcDNA3.1 plasmids containing wild-type and mutant α_1-antitrypsins were transiently transfected into COS-7 cells. The resulting protein expression was characterized by western blot analyses of SDS- and nondenaturing PAGE of intracellular material as previously described (Miranda et al., 2008, 2004). Luciferase controls were used to confirm equivalent loading between samples. The polymer load was quantified by ELISA using the polymer-specific 2C1 monoclonal antibody as the primary antibody (Miranda et al., 2010).

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and can be found with this article online at doi:10.1016/j.str.2012.01.012.

ACKNOWLEDGMENTS

The authors wish to thank Dr. James Irving (Department of Medicine, Cambridge Institute for Medical Research, University of Cambridge, UK) for helpful discussions during the preparation of this paper. M.P.N. received a Wellcome Trust Value in People Award; G.L. and A.O.M.P. hold Wellcome Trust programme PhD Studentships; U.I.E. is an MRC Clinical Training Fellow; and B.G. is a Wellcome Trust Intermediate Clinical Fellow. This work was supported by funding from the Wellcome Trust, the Medical Research Council (UK), the Human Frontier Science Programme, the Biotechnology and Biological Sciences Research Council, Papworth Hospital NHS Trust, and the Birkbeck College Faculty Research Fund.

Received: August 23, 2011
Revised: December 2, 2011
Accepted: January 3, 2012
Published: March 6, 2012

REFERENCES

Berjanskii, M., and Wishart, D.S. (2006). NMR: prediction of protein flexibility. Nat. Protoc. *1*, 683–688.

Blanco, I., Fernández, E., and Bustillo, E.F. (2001). Alpha-1-antitrypsin PI phenotypes S and Z in Europe: an analysis of the published surveys. Clin. Genet. *60*, 31–41.

Bruce, D., Perry, D.J., Borg, J.-Y., Carrell, R.W., and Wardell, M.R. (1994). Thromboembolic disease due to thermolabile conformational changes of antithrombin Rouen-VI (187 Asn—>Asp). J. Clin. Invest. *94*, 2265–2274.

Cabrita, L.D., Dai, W., and Bottomley, S.P. (2004). Different conformational changes within the F-helix occur during serpin folding, polymerization, and proteinase inhibition. Biochemistry *43*, 9834–9839.

Dafforn, T.R. (1999). The serpin database. Department of Haematology (Cambridge Institute for Medical Research).

Dafforn, T.R., Mahadeva, R., Elliott, P.R., Sivasothy, P., and Lomas, D.A. (1999). A kinetic mechanism for the polymerization of alpha1-antitrypsin. J. Biol. Chem. *274*, 9548–9555.

Dafforn, T.R., Pike, R.N., and Bottomley, S.P. (2004). Physical characterization of serpin conformations. Methods *32*, 150–158.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Dunstone, M.A., Dai, W., Whisstock, J.C., Rossjohn, J., Pike, R.N., Feil, S.C., Le Bonniec, B.F., Parker, M.W., and Bottomley, S.P. (2000). Cleaved antitrypsin polymers at atomic resolution. Protein Sci. *9*, 417–420.

Ekeowa, U.I., Freeke, J., Miranda, E., Gooptu, B., Bush, M.F., Pérez, J., Teckman, J.H., Robinson, C.V., and Lomas, D.A. (2010). Defining the mechanism of polymerization in the serpinopathies. Proc. Natl. Acad. Sci. USA *107*, 17146–17151.

Eletsky, A., Kienhöfer, A., and Pervushin, K. (2001). TROSY NMR with partially deuterated proteins. J. Biomol. NMR *20*, 177–180.

Fogh, R., Ionides, J., Ulrich, E., Boucher, W., Vranken, W., Linge, J.P., Habeck, M., Rieping, W., Bhat, T.N., Westbrook, J., et al. (2002). The CCPN project: an interim report on a data model for the NMR community. Nat. Struct. Biol. *9*, 416–418.

Gooptu, B., and Lomas, D.A. (2008). Polymers and inflammation: disease mechanisms of the serpinopathies. J. Exp. Med. *205*, 1529–1534.

Gooptu, B., and Lomas, D.A. (2009). Conformational pathology of the serpins: themes, variations, and therapeutic strategies. Annu. Rev. Biochem. *78*, 147–176.

Gooptu, B., Hazes, B., Chang, W.-S.W., Dafforn, T.R., Carrell, R.W., Read, R.J., and Lomas, D.A. (2000). Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. Proc. Natl. Acad. Sci. USA *97*, 67–72.

Gooptu, B., Ekeowa, U.I., and Lomas, D.A. (2009a). Mechanisms of emphysema in alpha1-antitrypsin deficiency: molecular and cellular insights. Eur. Respir. J. *34*, 475–488.

Gooptu, B., Miranda, E., Nobeli, I., Mallya, M., Purkiss, A., Brown, S.C., Summers, C., Phillips, R.L., Lomas, D.A., and Barrett, T.E. (2009b). Crystallographic and cellular characterisation of two mechanisms stabilising the native fold of alpha1-antitrypsin: implications for disease and drug design. J. Mol. Biol. *387*, 857–868.

Hilton, G.R., Thalassinos, K., Grabenauer, M., Sanghera, N., Slade, S.E., Wyttenbach, T., Robinson, P.J., Pinheiro, T.J.T., Bowers, M.T., and Scrivens, J.H. (2010). Structural analysis of prion proteins by means of drift cell and traveling wave ion mobility mass spectrometry. J. Am. Soc. Mass Spectrom. *21*, 845–854.

Huntington, J.A., Pannu, N.S., Hazes, B., Read, R.J., Lomas, D.A., and Carrell, R.W. (1999). A 2.6 A structure of a serpin polymer and implications for conformational disease. J. Mol. Biol. *293*, 449–455.

James, E.L., Whisstock, J.C., Gore, M.G., and Bottomley, S.P. (1999). Probing the unfolding pathway of alpha1-antitrypsin. J. Biol. Chem. *274*, 9482–9488.

Knaupp, A.S., Levina, V., Robertson, A.L., Pearce, M.C., and Bottomley, S.P. (2010). Kinetic instability of the serpin Z α_1-antitrypsin promotes aggregation. J. Mol. Biol. *396*, 375–383.

Krishnan, B., and Gierasch, L.M. (2011). Dynamic local unfolding in the serpin α-1 antitrypsin provides a mechanism for loop insertion and polymerization. Nat. Struct. Mol. Biol. *18*, 222–226.

Laurell, C.-B., and Jeppsson, J.-O. (1975). Proteinase inhibitors in plasma. In The Plasma Proteins, P. Fw, ed. (New York: Academic Press), pp. 229–265.

Lomas, D.A., and Carrell, R.W. (2002). Serpinopathies and the conformational dementias. Nat. Rev. Genet. *3*, 759–768.

Lomas, D.A., Evans, D.L., Finch, J.T., and Carrell, R.W. (1992). The mechanism of Z alpha 1-antitrypsin accumulation in the liver. Nature *357*, 605–607.

Mahadeva, R., Chang, W.-S.W., Dafforn, T.R., Oakley, D.J., Foreman, R.C., Calvin, J., Wight, D.G., and Lomas, D.A. (1999). Heteropolymerization of S, I, and Z alpha1-antitrypsin and liver cirrhosis. J. Clin. Invest. *103*, 999–1006.

Miranda, E., Römisch, K., and Lomas, D.A. (2004). Mutants of neuroserpin that cause dementia accumulate as polymers within the endoplasmic reticulum. J. Biol. Chem. *279*, 28283–28291.

Miranda, E., MacLeod, I., Davies, M.J., Pérez, J., Römisch, K., Crowther, D.C., and Lomas, D.A. (2008). The intracellular accumulation of polymeric neuroserpin explains the severity of the dementia FENIB. Hum. Mol. Genet. *17*, 1527–1539.

Miranda, E., Pérez, J., Ekeowa, U.I., Hadzic, N., Kalsheker, N., Gooptu, B., Portmann, B., Belorgey, D., Hill, M., Chambers, S., et al. (2010). A novel monoclonal antibody to characterize pathogenic polymers in liver disease associated with alpha1-antitrypsin deficiency. Hepatology *52*, 1078–1088.

Nyon, M.P., Kirkpatrick, J., Cabrita, L.D., Christodoulou, J., and Gooptu, B. (2011). $^{(1)}$H, $^{(15)}$N and $^{(13)}$C backbone resonance assignments of the archetypal serpin $\alpha_{(1)}$-antitrypsin. Biomol. NMR Assign., in press.

Parfrey, H., Mahadeva, R., Ravenhill, N.A., Zhou, A., Dafforn, T.R., Foreman, R.C., and Lomas, D.A. (2003). Targeting a surface cavity of alpha 1-antitrypsin to prevent conformational disease. J. Biol. Chem. *278*, 33060–33066.

Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc. Natl. Acad. Sci. USA *94*, 12366–12371.

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

Rance, M., Loria, J.P., and Palmer, A.G. (1999). Sensitivity improvement of transverse relaxation-optimized spectroscopy. J. Magn. Reson. *136*, 92–101.

Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wüthrich, K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc. Natl. Acad. Sci. USA *95*, 13585–13590.

Schulze, A.J., Baumann, U., Knof, S., Jaeger, E., Huber, R., and Laurell, C.-B. (1990). Structural transition of alpha 1-antitrypsin by a peptide sequentially similar to beta-strand s4A. Eur. J. Biochem. *194*, 51–56.

Shaka, A.J., Keeler, J., Frenkiel, T., and Freeman, R. (1983). An improved sequence for broadband decoupling: WALTZ-16. J. Magn. Reson. Im. *52*, 335–338.

Sharp, A.M., Stein, P.E., Pannu, N.S., Carrell, R.W., Berkenpas, M.B., Ginsburg, D., Lawrence, D.A., and Read, R.J. (1999). The active conformation of plasminogen activator inhibitor 1, a target for drugs to control fibrinolysis and cell adhesion. Structure *7*, 111–118.

Stone, S.R., and Hofsteenge, J. (1986). Kinetics of the inhibition of thrombin by hirudin. Biochemistry *25*, 4622–4628.

Thalassinos, K., Grabenauer, M., Slade, S.E., Hilton, G.R., Bowers, M.T., and Scrivens, J.H. (2009). Characterization of phosphorylated peptides using traveling wave-based and drift cell ion mobility mass spectrometry. Anal. Chem. *81*, 248–254.

Yamasaki, M., Li, W., Johnson, D.J., and Huntington, J.A. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. Nature *455*, 1255–1258.

Zhou, A., Carrell, R.W., and Huntington, J.A. (2001). The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop. J. Biol. Chem. *276*, 27541–27547.

Zhou, A., Stein, P.E., Huntington, J.A., and Carrell, R.W. (2003). Serpin polymerization is prevented by a hydrogen bond network that is centered on his-334 and stabilized by glycerol. J. Biol. Chem. *278*, 15116–15122.

Zhu, G., Kong, X.M., and Sze, K.H. (1999). Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. J. Biomol. NMR *13*, 77–81.

